



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Comunicação

Introdução à Preservação Digital

**Conceitos estratégias e actuais
consensos**



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

20 de Abril de 2007
S. Martinho de Bougado, Trofa



Conteúdo

- ▶ **Contextualização**
- ▶ **Exemplos** de obsolescência tecnológica
- ▶ **Conceitos**/definições
- ▶ **Estratégias** de preservação
 - Preservação de tecnologia
 - Preservação do objecto conceptual
- ▶ **Notas finais**



Contextualização

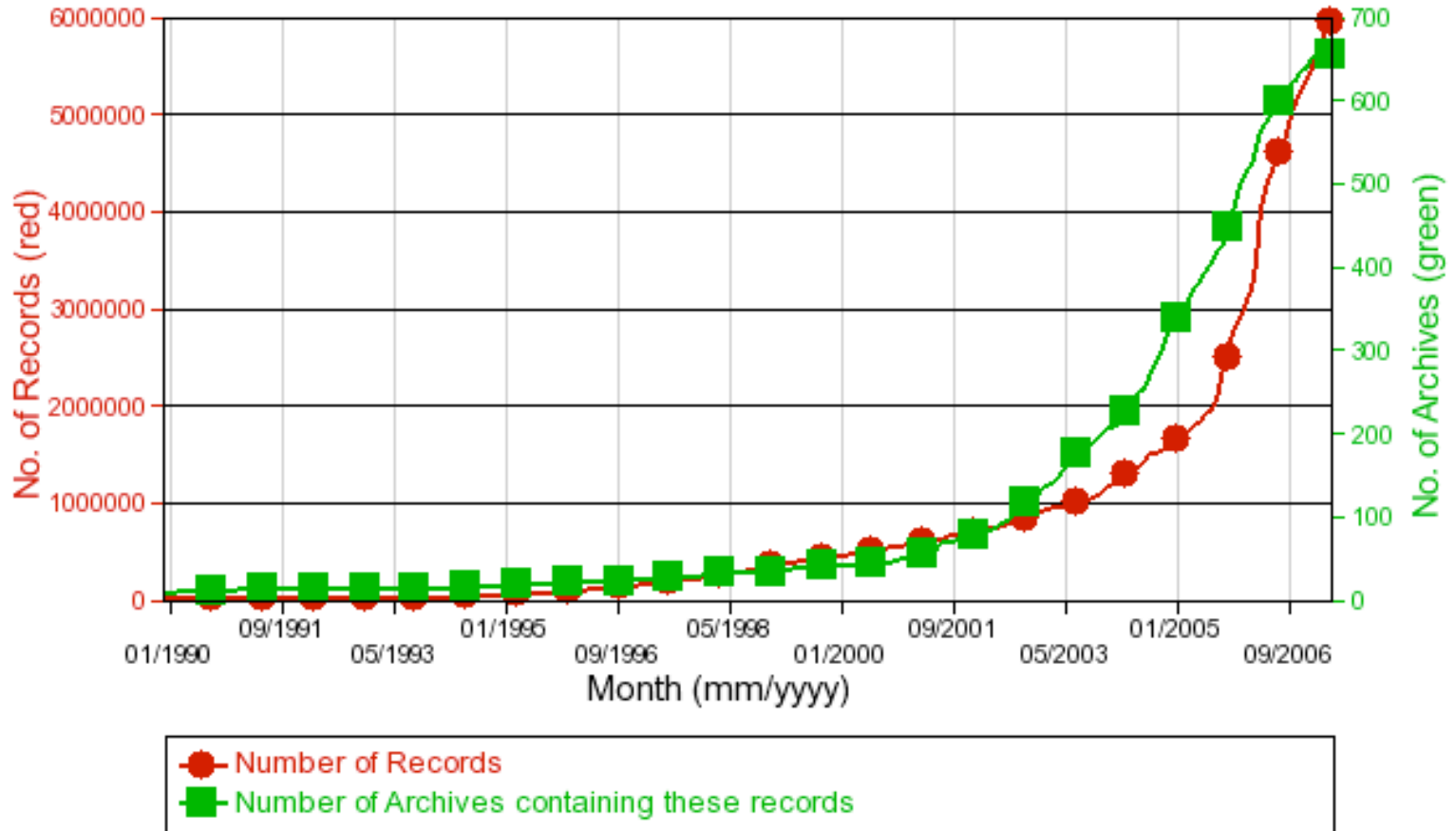
- ▶ Vivemos numa **era digital**
 - Grande parte da **produção intelectual** é criada em **formatos digitais**
 - **Facilidade** na sua **criação** e **disseminação**
 - **Qualidade** dos resultados obtidos
- ▶ Rápida **obsolescência** da tecnologia
 - Embora um objecto digital possa ser **copiado infinitas vezes sem perder qualidade**
 - Depende de um **contexto tecnológico** para que possa ser interpretado
 - Os objectos acabam **prisoneiros** dos seus próprios **suportes** e/ou **formatos**



Nº registos em repositórios institucionais

Registry of Open Access Repositories (ROAR)

Generated by <http://archives.eprints.org/>





UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

5

Evolução dos suportes audio [1877-2007]

http://en.wikipedia.org/wiki/Audio_format

Tempo...



Evolução dos suportes audio [1877-2007]

http://en.wikipedia.org/wiki/Audio_format



1947

Reel-to-Reel,
Magnetic Tape



1930s

Wire recording



1895

Gramophone record



1877

Cilindro do fonógrafo de Edison

Tempo...

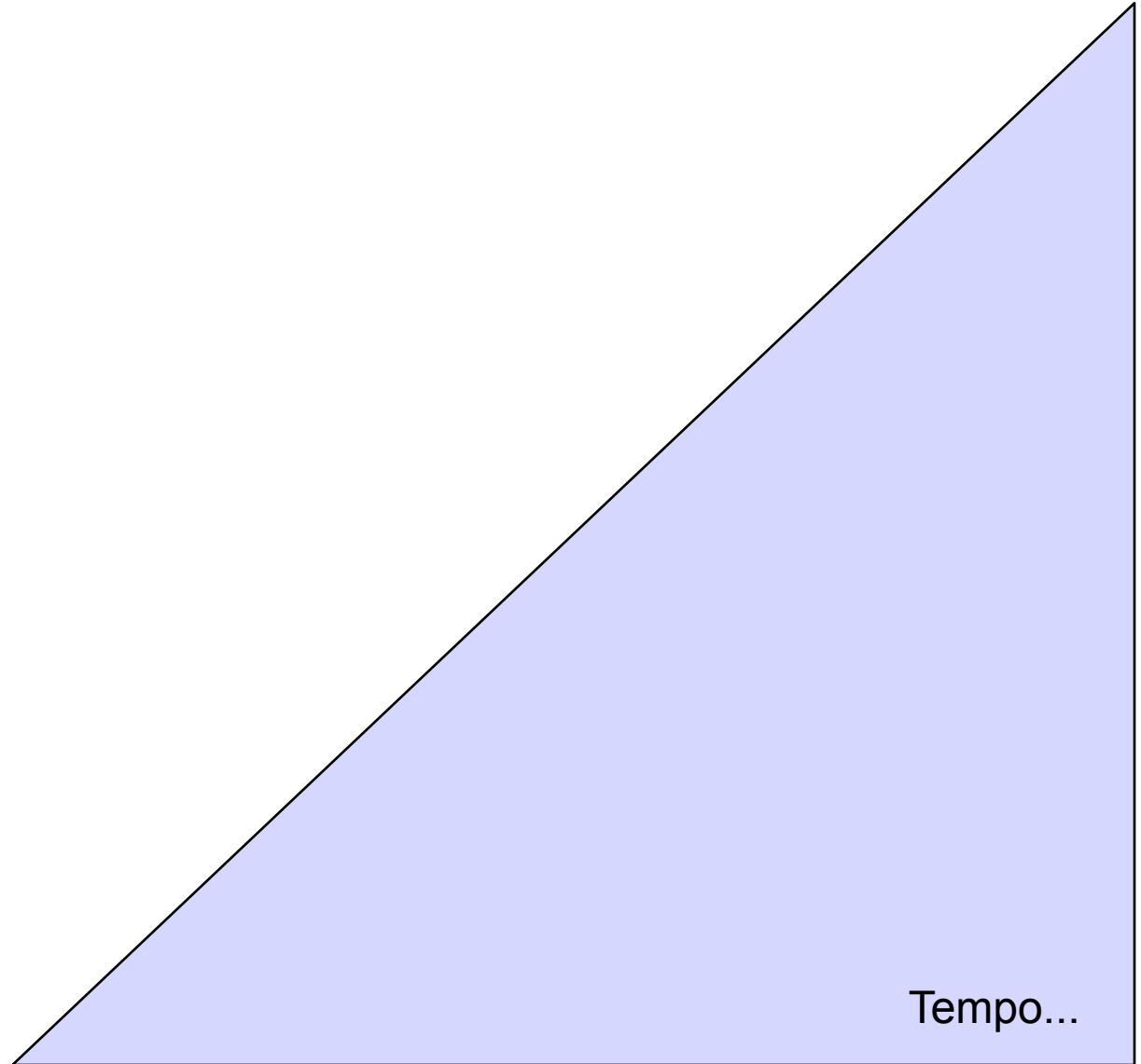


UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

6

Evolução dos suportes audio [1877-2007]

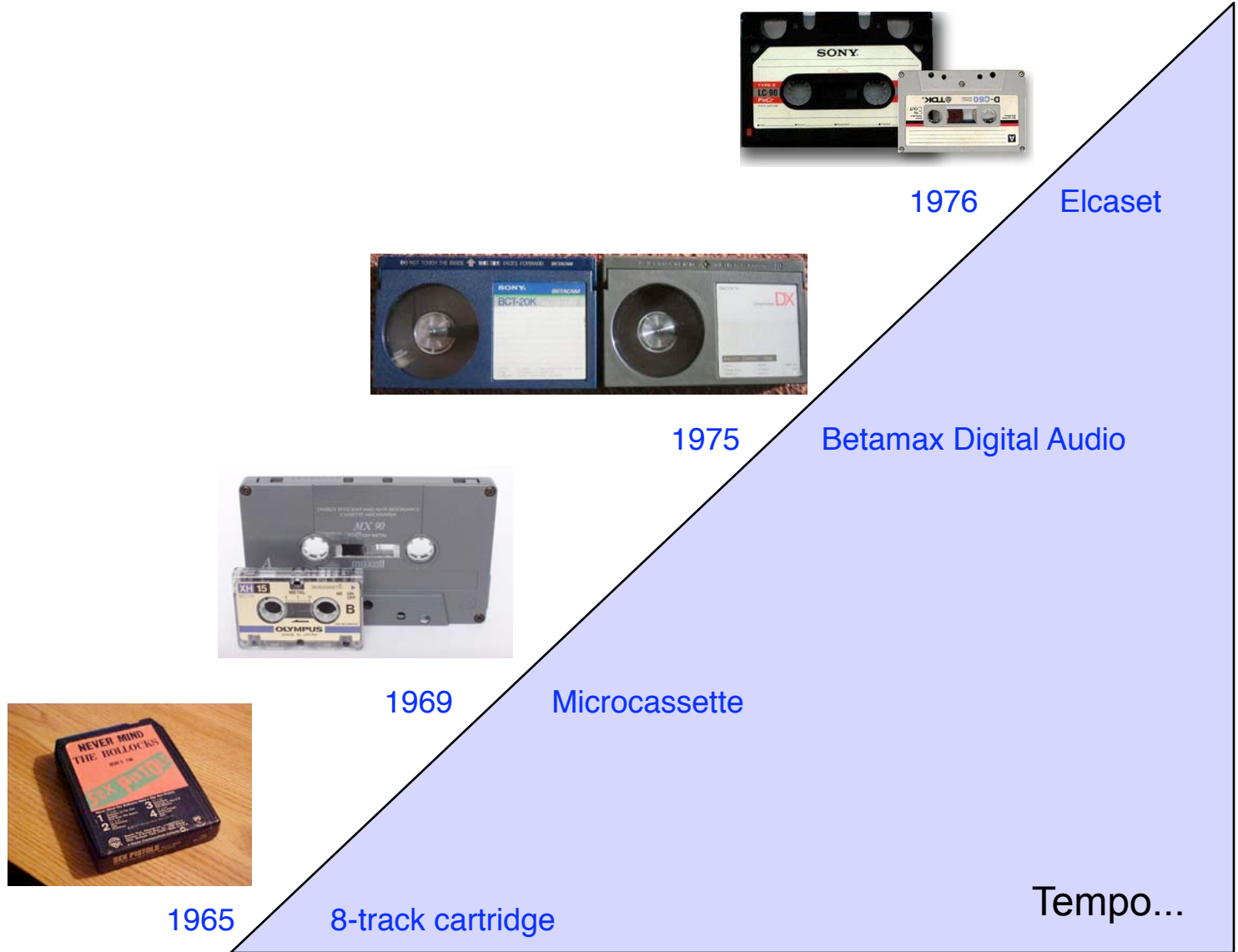




UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Evolução dos suportes audio [1877-2007]



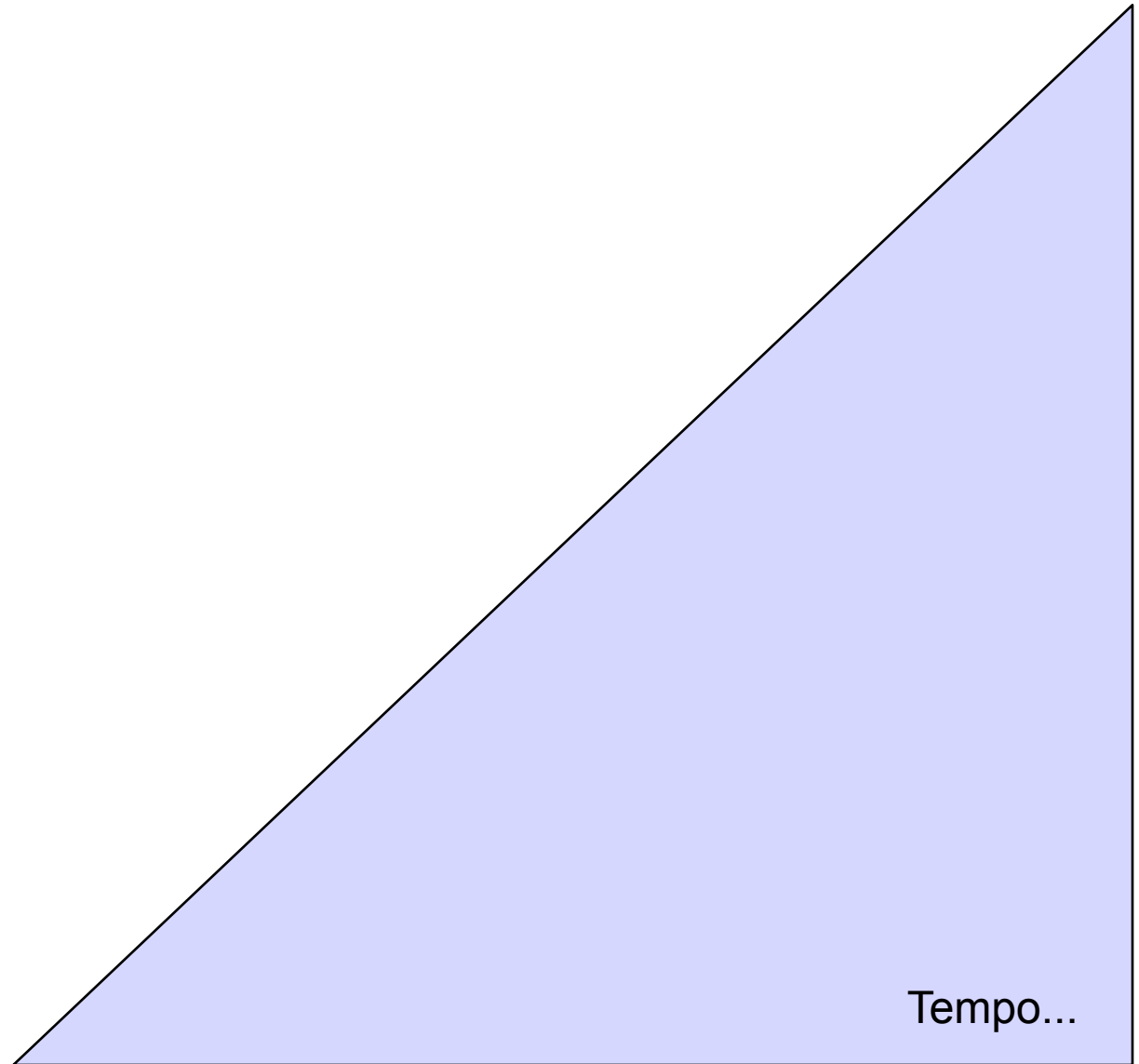


UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

7

Evolução dos suportes audio [1877-2007]

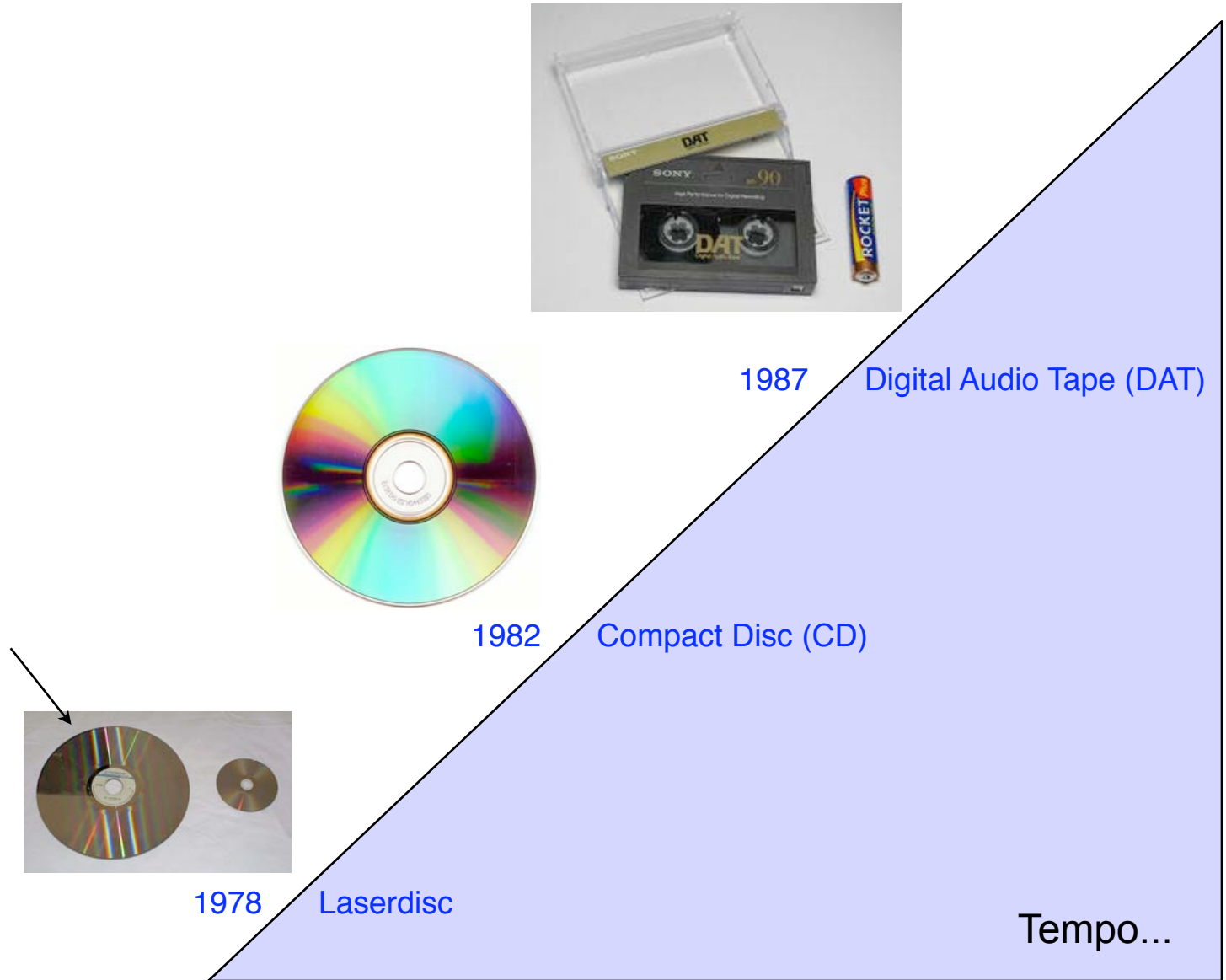




UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Evolução dos suportes audio [1877-2007]



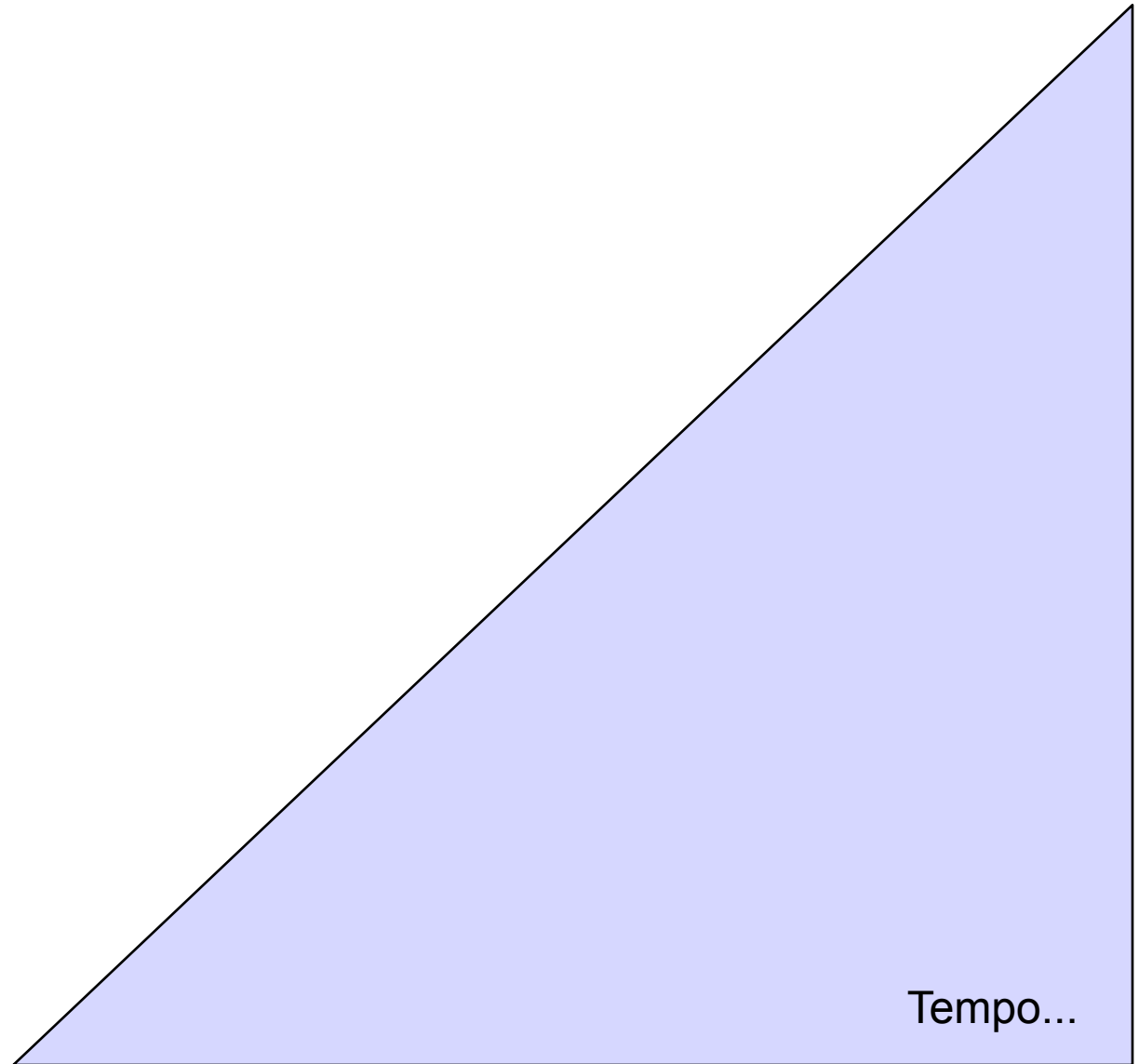


UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

08

Evolução dos suportes audio [1877-2007]

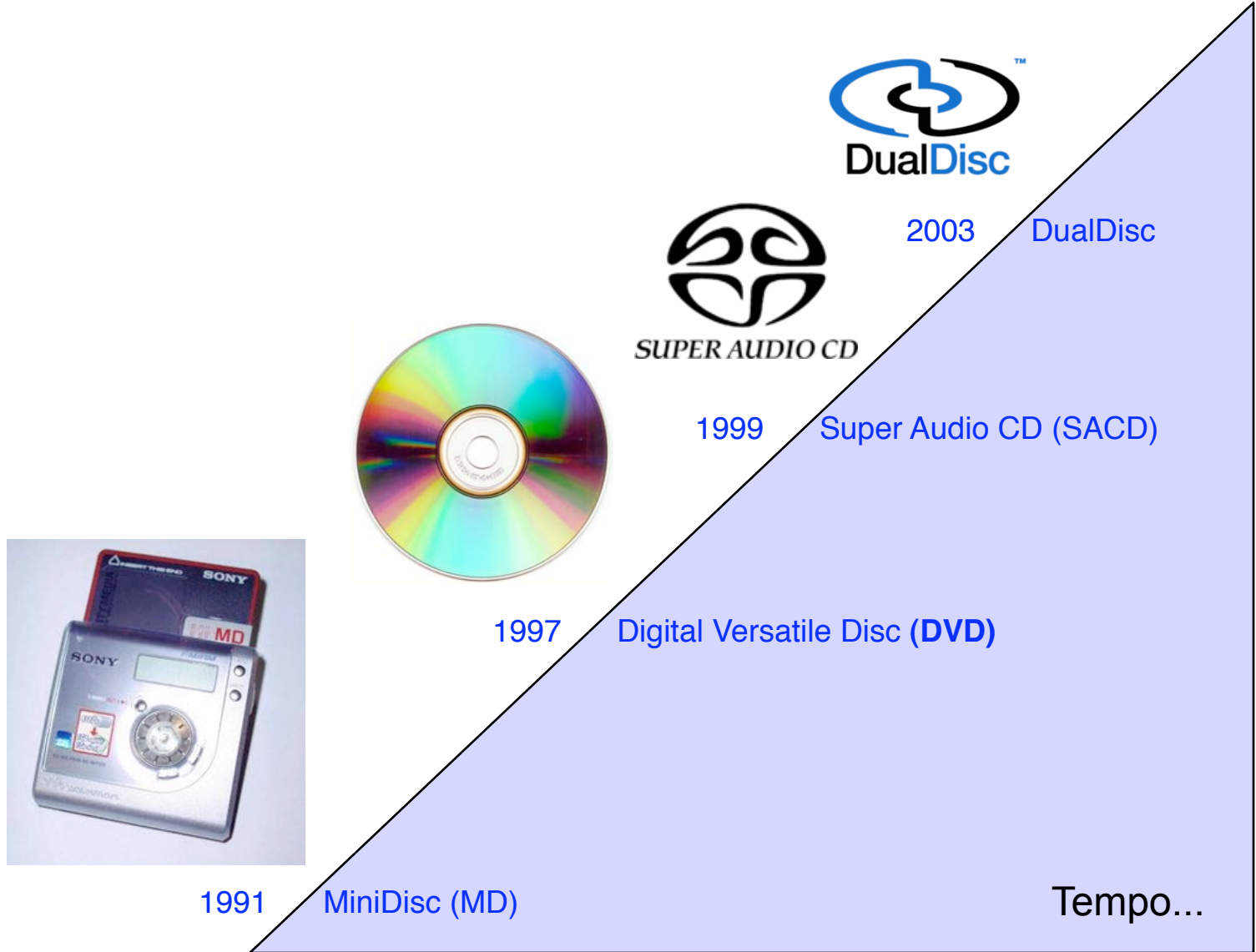




UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Evolução dos suportes audio [1877-2007]



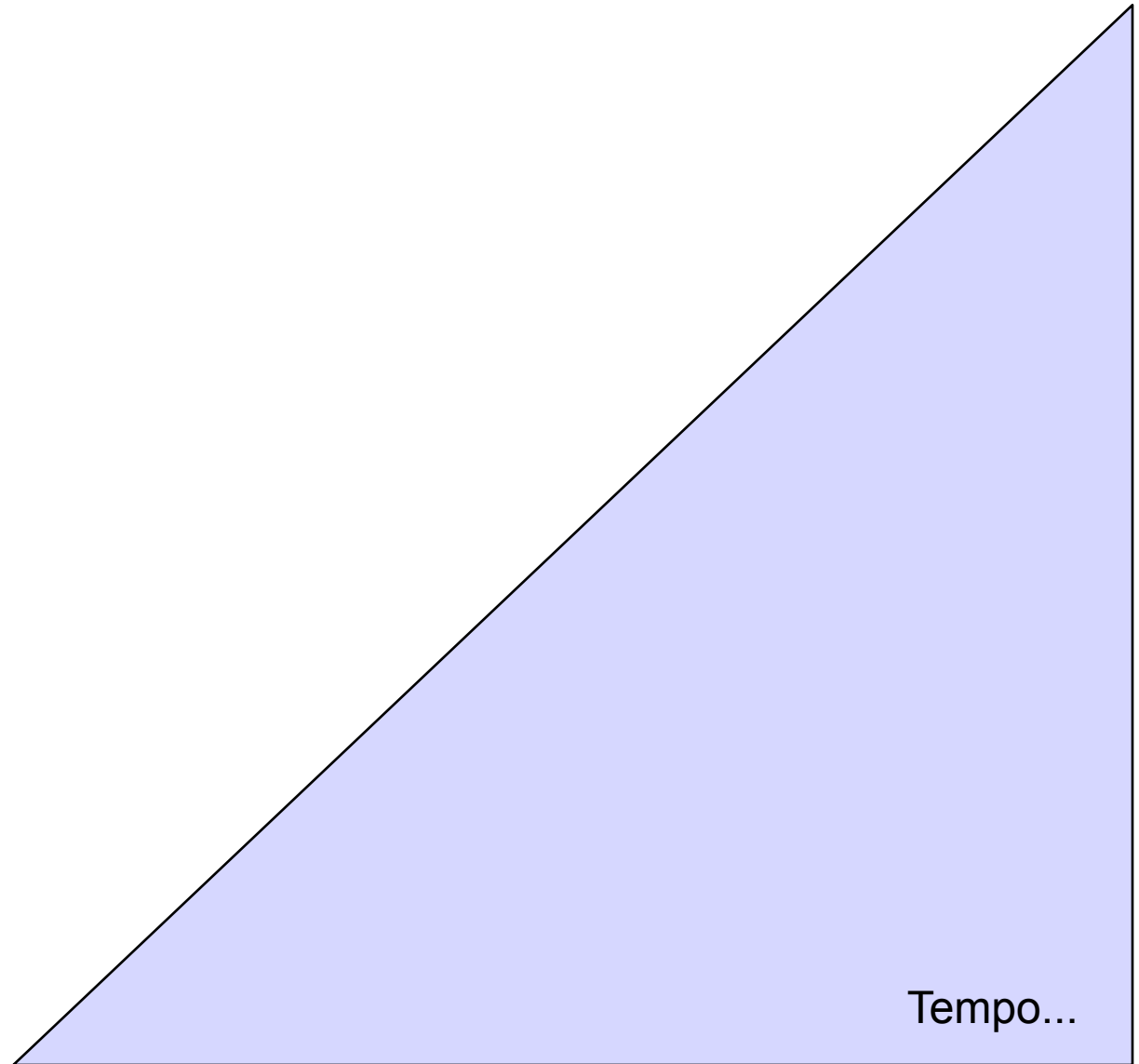


UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

9

Evolução dos suportes audio [1877-2007]

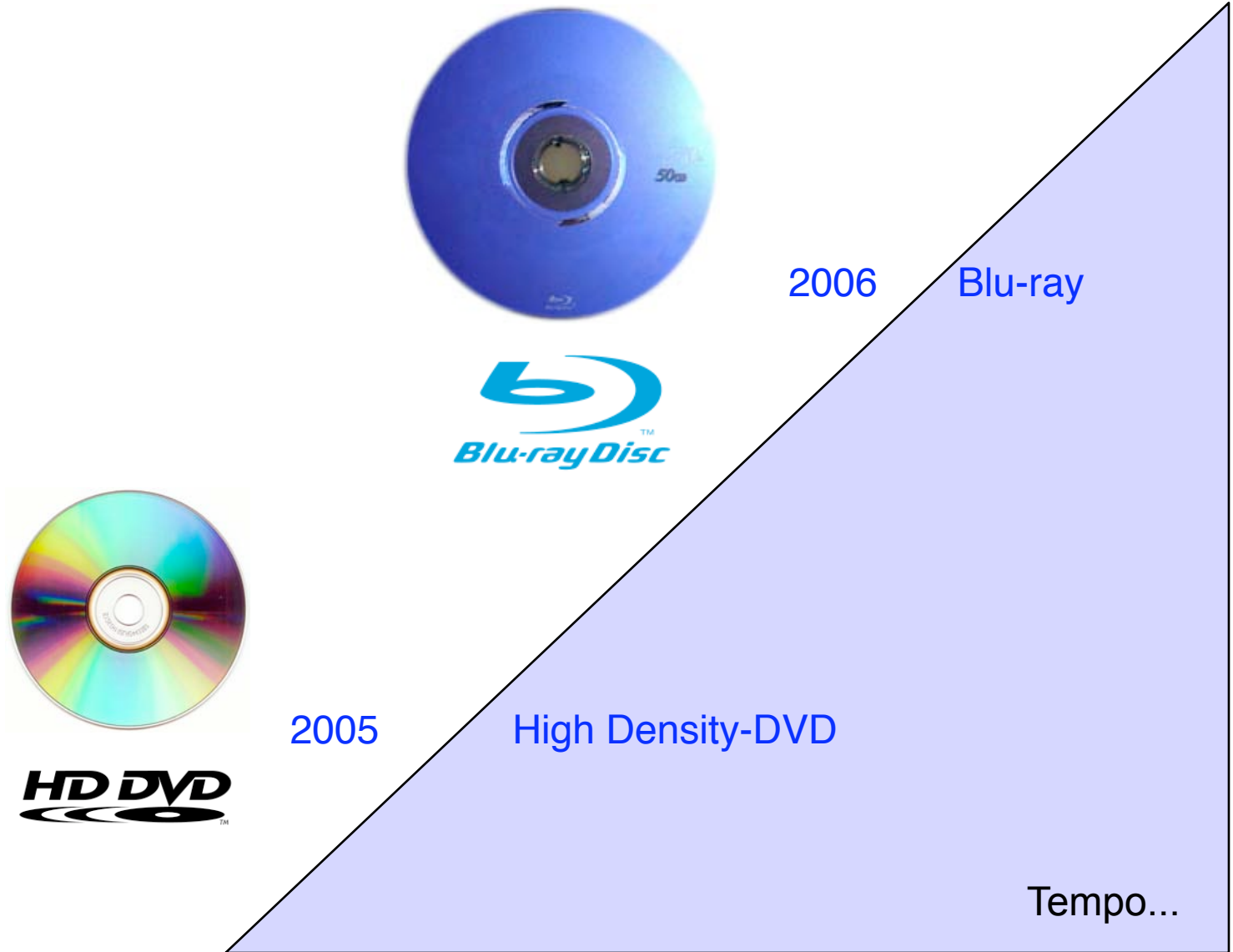




UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Evolução dos suportes audio [1877-2007]

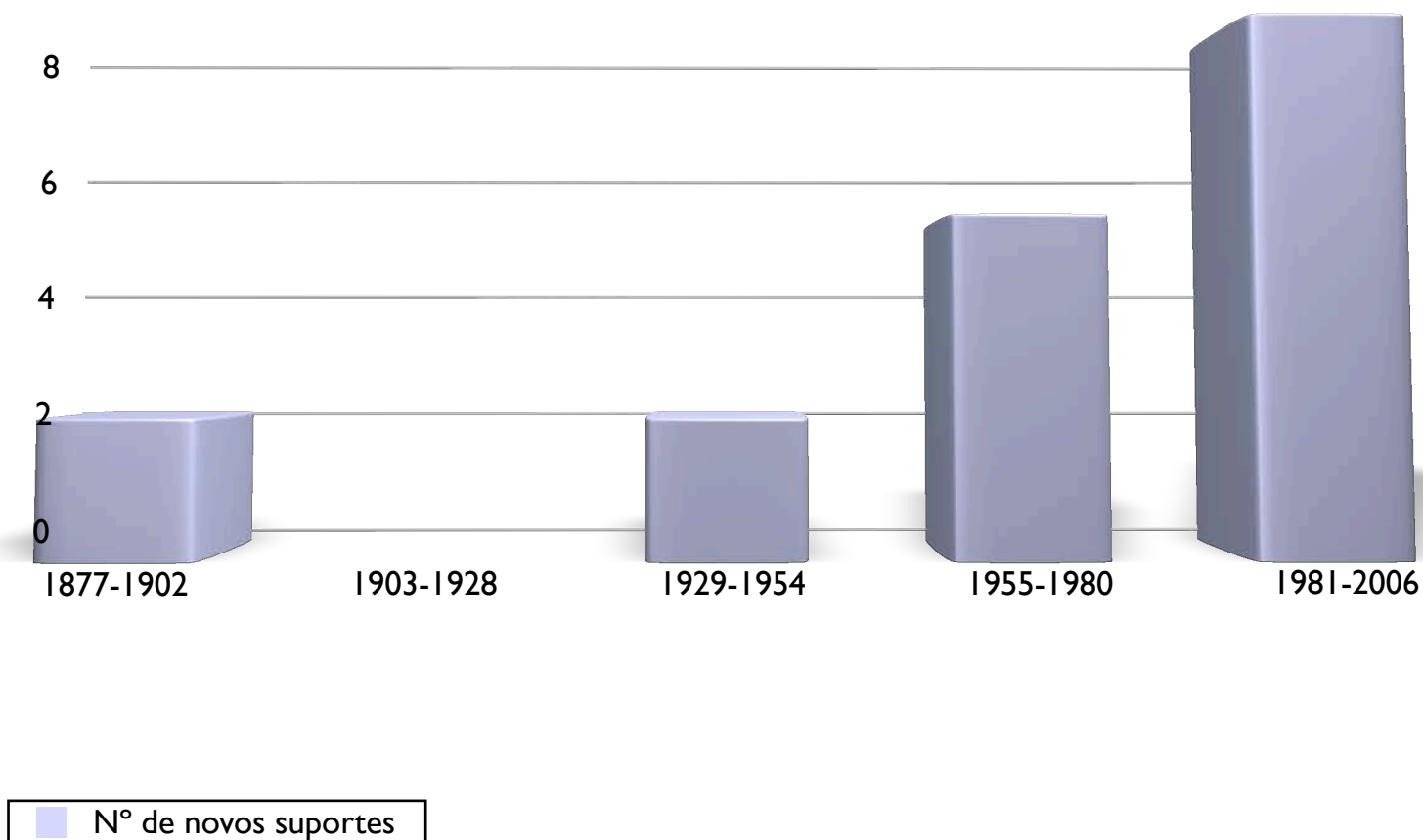




UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Aparecimento de novos suportes audio





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





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**

Webb, C., Bradley, K., Campbell, D., Clifton, G., Corbould, M., O'Connor, M., Phillips, M. and Whiting, J. Guidelines for the Preservation of Digital Heritage, Information Society Division - United Nations Educational, Scientific and Cultural Organization, 2003.



Objecto, material ou recurso digital

Tudo aquilo que possa ser **representado em formato digital**

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

▶ **Exemplos:**

- bases de dados
- folhas de cálculo
- documentos de texto
- páginas Web
- vídeo, audio, imagens
- mapas,
- modelos 3D
- software
- etc...



Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

TER24AGO
EDIÇÃO PORTO
24 de Agosto de 2004
Ano XV • Nº 1251
450 000 VOLS. ANUALMENTE
Diretor JOSÉ MANUEL FERREIROS
Secretaria-geral NUNO FERRAZ
e MANUEL CARVALHO
www.publico.pt

SINÓNIMOS E ANTÓNIMOS
Dicionário
COM MAIS 10 VOLUMES
10000 PALAVRAS EM 10000 EXEMPLOS

PÚBLICO
www.publico.pt

TRÊS MILHÕES POR DISTRIBUIR ÀS VÍTIMAS DOS INCÊNDIOS DE 2003

Um ano depois das campanhas de solidariedade para com as vítimas dos incêndios florestais de 2003, ainda restam por pagar 3,2 milhões de euros dos 3 milhões de euros. O presidente da Associação Nacional de Municípios, Fernando Ruas,

Verdes das campanhas de solidariedade do ano passado podem ir para casas deste ano

Se um utilizador não autoriza, "não lhe será permitido" o acesso a certos conteúdos. É preciso saber a quem se refere esta declaração e a quem se aplica. O que se refere ao acesso para as vítimas dos incêndios de 2003.

CULTURA
Polícia procura do DNA dos ladrões dos quadros de Manet

A polícia portuguesa está a melhorar os métodos de "O Crime" e "Máscaras", os crimes de quadrado de Manet, roubados no dia 20 de maio, à procura de impressões digitais. Já se fez a identificação do crime. A ministra da Cultura, Irineu Gomes, critica a falta de investigação crítica sobre

Assessor de Rui Rio ganha mais que Jorge Sampaio

Finanças com débitos no reembolso de IVA ao BCP

A Inspeção-Geral de Finanças multou o Banco de Portugal por uma falta de reembolso de IVA ao BCP. O BCP não quis reconhecer o erro, mas considerou que, entre outros, se trata de diferentes interpretações da lei sobre a distribuição de resultados.

FUTEBOL
Gilberto Madal
admitido chamado de Derki à seleção nacional

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

Francis Obikwela 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

A festa de vitória no Estádio Olímpico



Reportagem
de um dos
participantes
de Angola, Leticia
Muniz, sobre o
evento em
Beira, Moçambique

Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

TER24AGO
EDIÇÃO PORTO

24 de Agosto de 2004
Ano XV • Nº 1251
€0,80 (IVA incluída)

Director JOSÉ MANUEL FERREIROS
Secretário-geral NUNO FERRAZ
e MANUEL CARVALHO

PÚBLICO
www.publico.pt

SINÓNIMOS E ANTÓNIMOS
Dicionário
COM MAIS 10 VOLUMES
HOJE PRECISAMOS VOLTAR PÁG. 63 A 67

TRÊS MILHÕES POR DISTRIBUIR ÀS VÍTIMAS DOS INCÊNDIOS DE 2003

Um ano depois dos combates de salitre, o Estado tem em mãos os 3 milhões de euros necessários para distribuir 3 milhões de euros às vítimas dos incêndios de 2003. A distribuição será feita em 12 parcelas durante os próximos meses. O presidente da Associação Nacional de Municípios, Fernando Ruas,

Veja as empresas de solidariedade do ano passado podem ir para casa deste ano

Se um cidadão não estiver, "isto é sentido que o Estado não vá ao encontro do cidadão. É preciso trabalhar para a criação de uma cultura de cidadania". O que vai ser de 2003 repete-se para as vítimas dos incêndios de 2003. **REPÚBLICA**

CULTURA
Polícia à procura do DNA dos ladrões dos quadros de Manet

A polícia portuguesa está a analisar as moléculas de "O Criar" e "Madame", as últimas obras de Manet roubadas no mundo, à procura de impressões digitais. Já se fez a identificação do crime. A ministra da Cultura, Irina Cabaco, critica a falta de investigação crítica antes da aquisição.

Assessor de Rui Rio ganha mais que Jorge Sampaio

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

A Inspeção-Geral de Finanças multou o Banco Comercial Português por um reembolso de IVA pedido pelo 3.º Milionário BCP por valor superior a 40 milhões de euros, apesar de o BCP não ter sido autorizado a fazer esse reembolso. A Inspeção-Geral de Finanças não autorizou o reembolso.

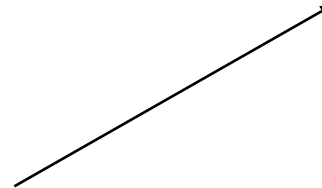
FUTEBOL
Gilberto Madaladante chamada de Derki à selecção nacional

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

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

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

Como Angola defrontou os gigantes norte-americanos do basquetebol



Central
Oklahoma's
Premiere
Shopper's Guide
Since 1977

Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Exclusividade

publicopt

2 de Maio de 2006 - 22h50

PLUS

Utilizador:

HOME

CANAIS

- Edição Imprensa
- Última Hora
- Última Hora UL
- Carga e Transportes
- Última Hora **h35**
- Bolsa
- Desporto
- Consultórios
- Podium
- Guia do Lazer
- Cinecartaz

Dez detidos e 19 armas apreendidas no Bairro da Torre

Dez detidos e 19 armas de fogo apreendidas é o balanço provisório da operação policial hoje realizada no Bairro da Torre, em Camarate, segundo o Comando Metropolitano da PSP de Lisboa.

- › **Moradores do Bairro da Torre queixam-se da actuação da polícia**
- › **Operação policial no Bairro da Torre resultou em várias detenções**

Protesto inflamado

Um apicultor sul-coreano salta para cima de uma bandeira japonesa e de uma imagem do primeiro-ministro Junichiro Koizumi, em Seul, Ahn

Conselho de Segurança continua dividido sobre crise nuclear iraniana

China e Rússia opõem-se a aprovação de sanção vinculativa contra o Irão

- O meu analista avisou-me

ESCREVA AO PROVEDOR DOS LEITORES

Find: Find Next Find Previous Highlight all Match case Phrase not found

http://www.publico.clix.pt/sites/blogues/

MI2006 MI2006-0.2.ppt PUBLICO.PT - Mozilla ...

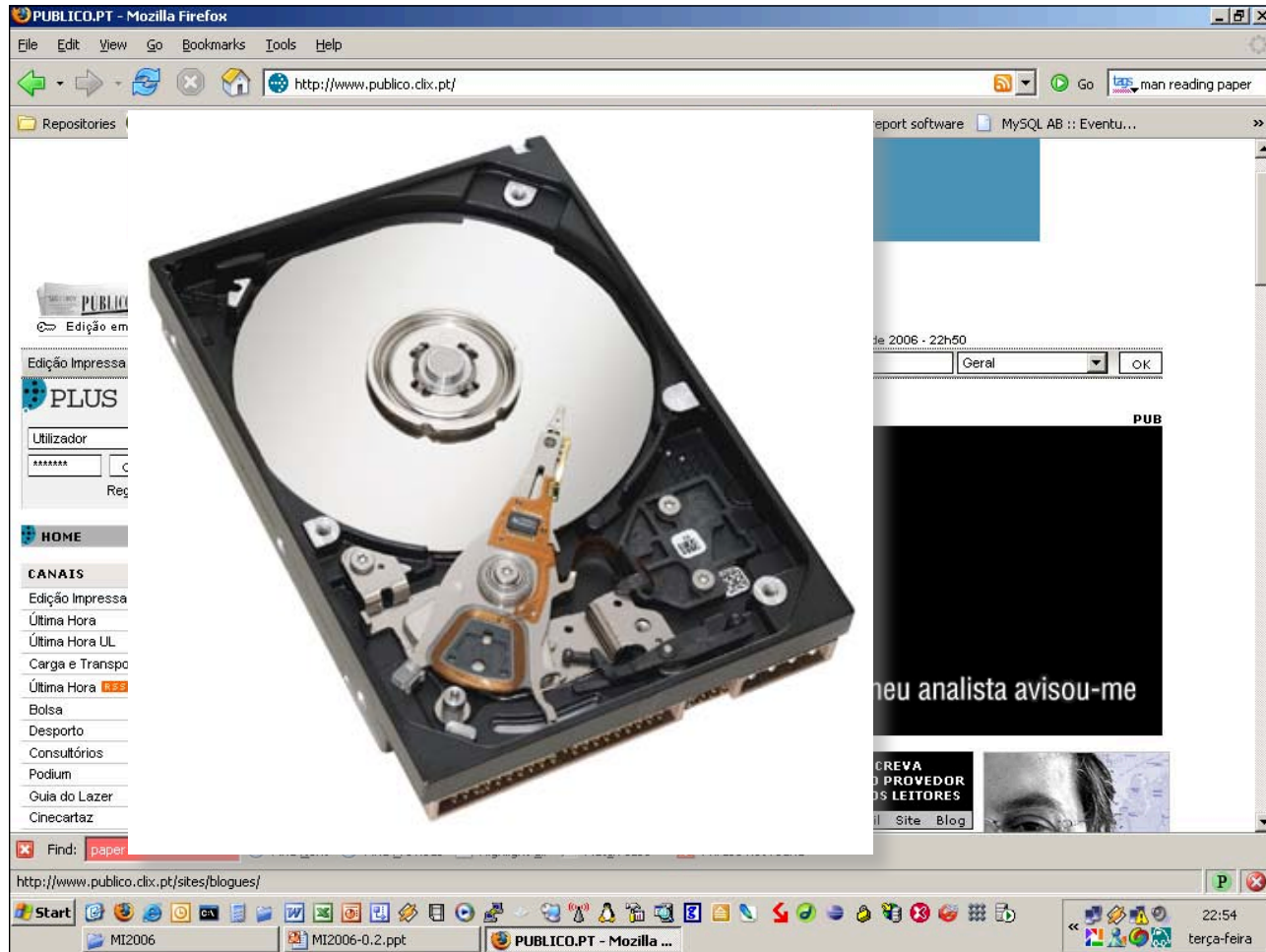
22:54 terça-feira

Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital



Analógico vs Digital



UNIVERSIDADE
DO
MINHO

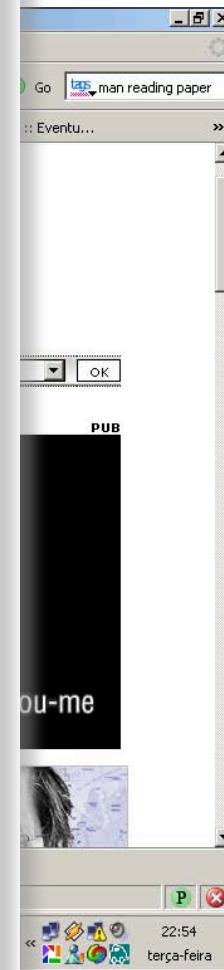
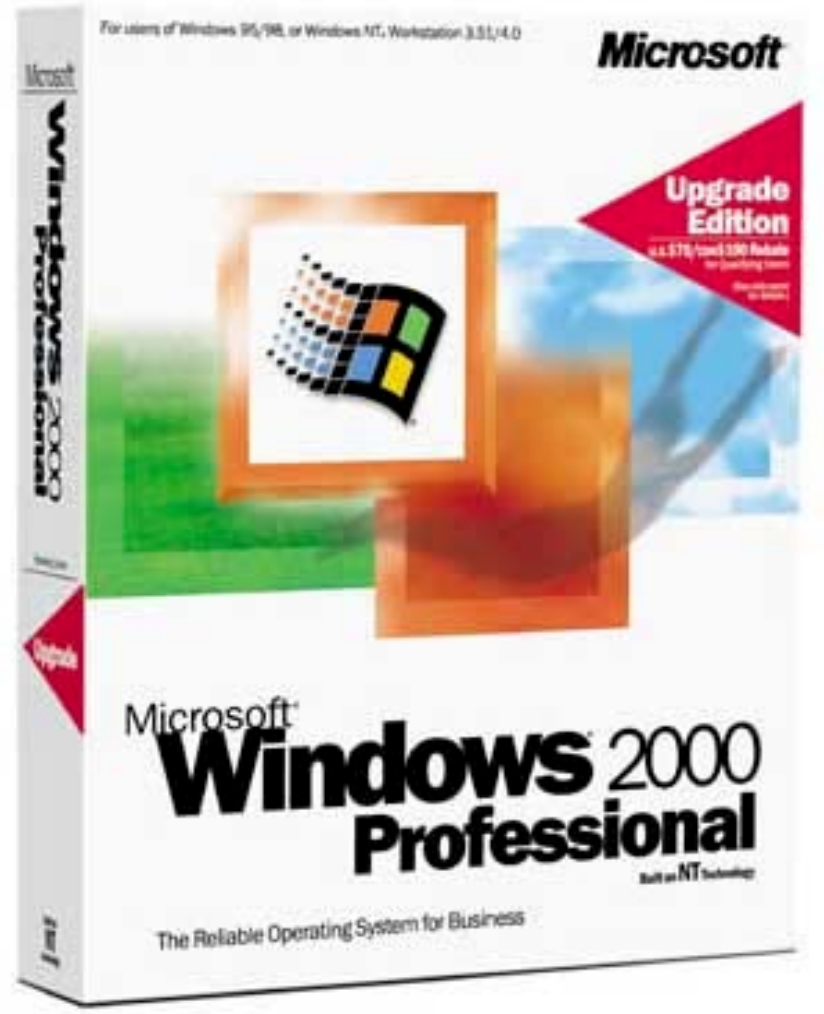
Jornadas de Trabalho
A gestão da informação na era digital

Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital



Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital



Analógico vs Digital



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

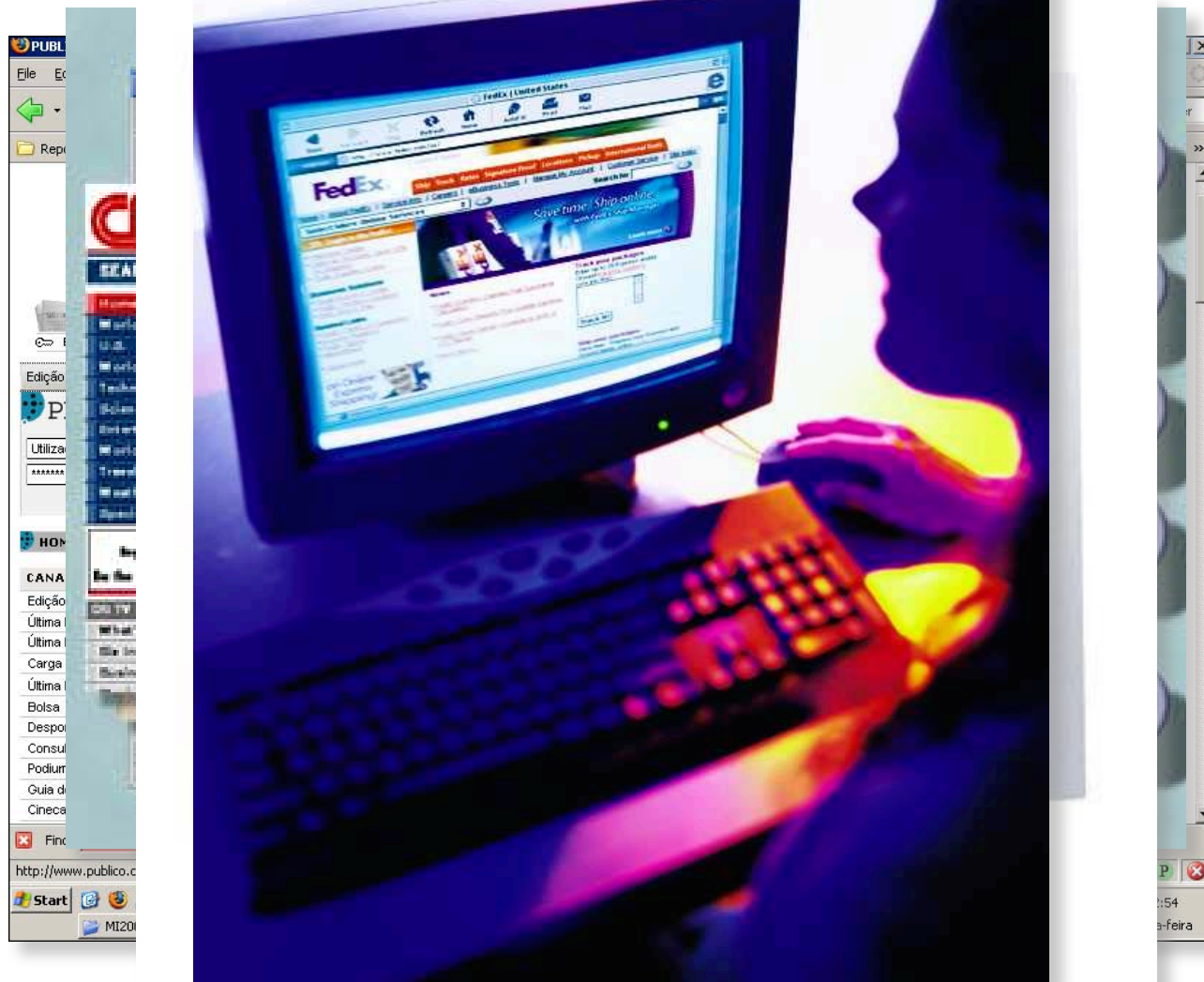


Analógico vs Digital



UNIVERSIDADE
DO
MINHO

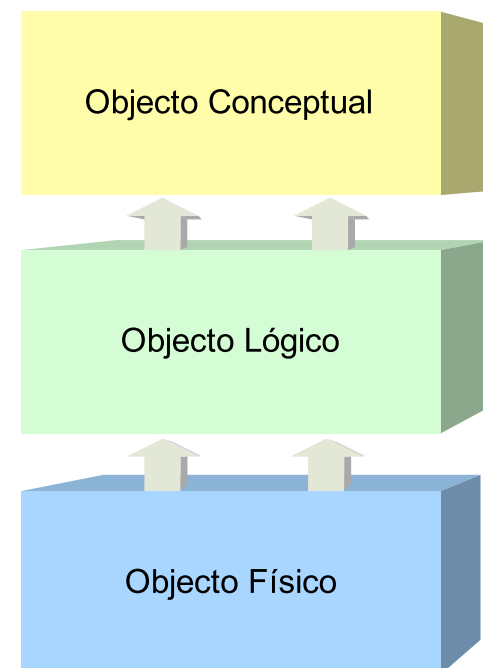
Jornadas de Trabalho
A gestão da informação na era digital





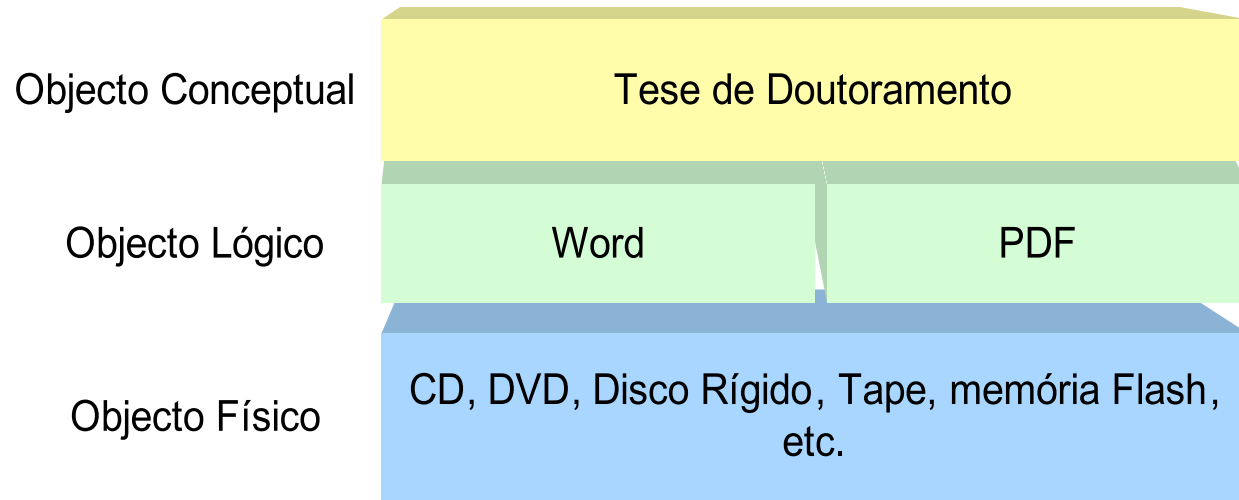
A anatomia de um objecto digital

- **Nível Conceptual**
 - Aquilo com que lidamos no **mundo real**; o que reconhecemos como uma unidade significativa de informação
 - e.g. jornal, livro, mapa, fotografia
- **Nível Lógico**
 - **Codificação** da informação
 - Tarefa realizada pelo software
 - Trata-se do formato do objecto
- **Nível Físico**
 - **Sinais** inscritos em suporte físico
 - e.g. disco rígido, CD, Pen-drive





A anatomia de um objecto digital



Se um dos níveis se tornar obsoleto o objecto conceptual deixa de estar acessível



Estratégias de preservação

Duas vias possíveis

Preservação de
tecnologia

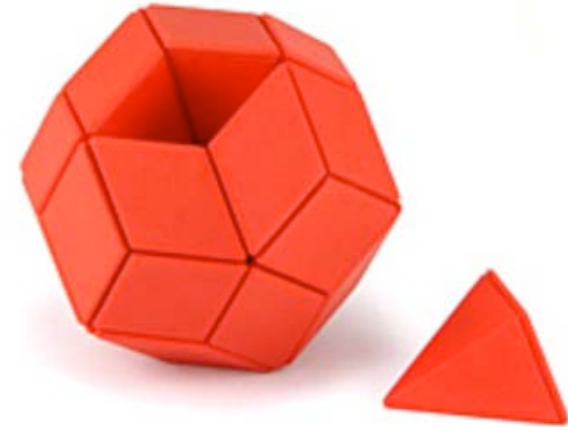
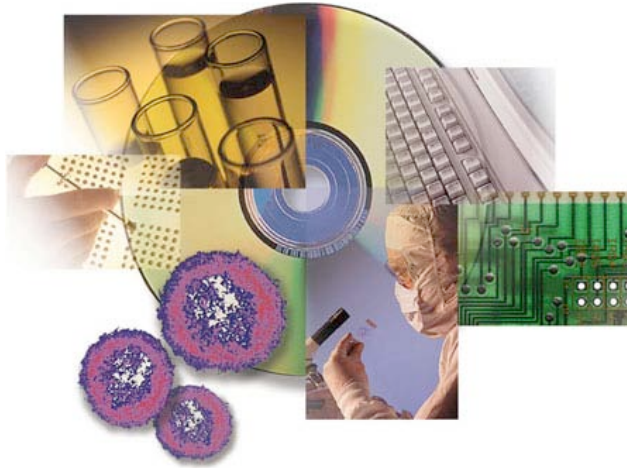


Preservação da
informação





Estratégias de preservação



Preservação de tecnologia

- ▶ **Preservar** a informação nos seus **formatos** lógicos e/ou físicos **originais (preservação do bitstream)**
- ▶ 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!



UNIVERSIDADE
DO
MINHO

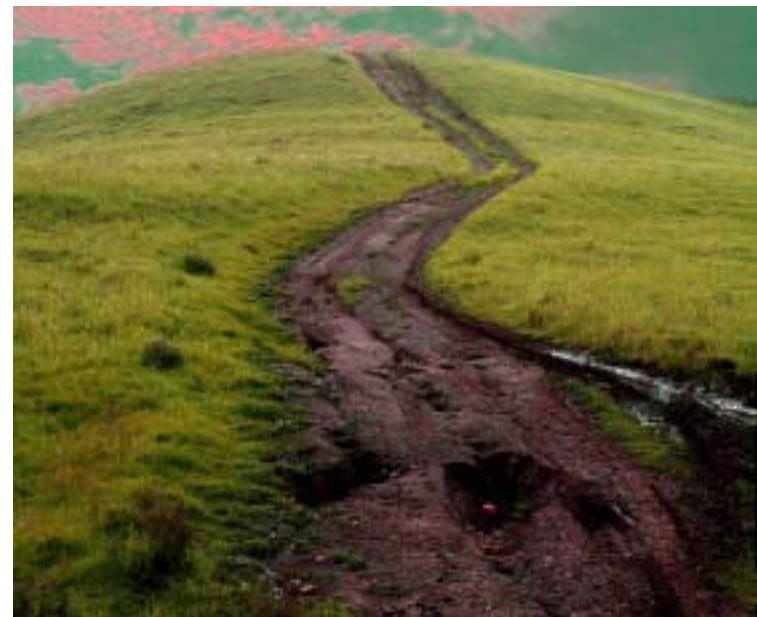
Jornadas de Trabalho
A gestão da informação na era digital

Estratégias de preservação

Preservação de tecnologia



Preservação da informação





UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Estratégias de preservação

Preservação de tecnologia





Preservação de tecnologia

▶ **Museus** de hardware

- Vantagens
 - **Preservação** do **contexto** de criação/
interpretação **original**
- Desvantagens
 - Demasiado **dispendioso**
 - Dificuldade em **mater o hardware operacional** a longo-prazo
 - **Restrições** ao nível do **acesso** à
informação
 - **Dificuldade** na **reutilização** da informação



Emulação

Software capaz de **reproduzir o comportamento** de uma **plataforma** de hardware e/ou software

▶ Vantagens

- Recriação do **contexto tecnológico original**
- Preservação do **look & feel** do objecto

▶ Desvantagens

- Os próprios **emuladores** estão **sujeitos a obsolescência**
- Os **utilizadores** são obrigados a **conhecer sistemas** há muito **desaparecidos**
- A **construção** de **emuladores** é um processo **complexo**
- Problemas ao nível dos **direitos de autor** das plataformas
- Preservar um sistema operativo inteiro apenas para **visualizar** um **documento** parece uma **tarefa exageradamente pesada**
- Não garante a **reutilização da informação**

Emulação [exemplo]



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

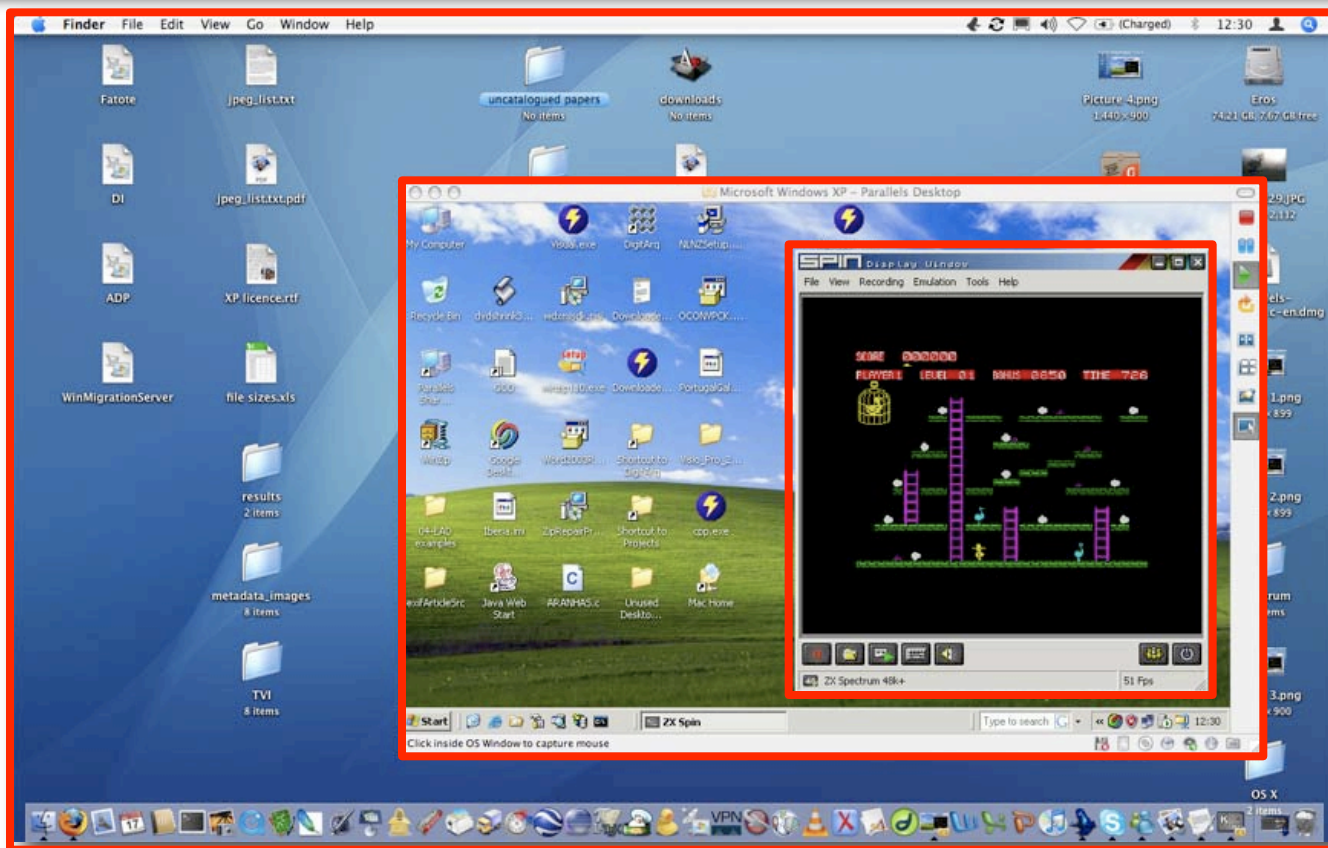


Emulação [exemplo]



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital





Encapsulamento

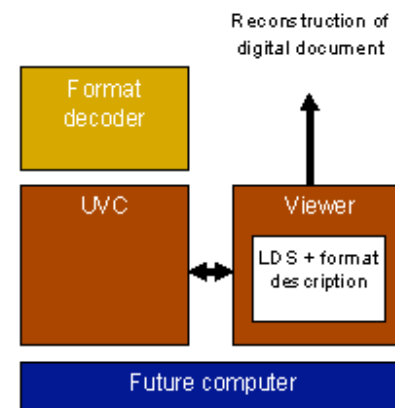
Preservação da **sequência de bits original** juntamente com **metainformação** que permita a sua interpretação no futuro

- ▶ Vantagens
 - Permite **adiar a responsabilidade** de preservação
 - Orientado a objectos que apenas serão acedidos num futuro longínquo
 - Futuro desenvolvimento de **visualizadores, migradores** ou **emuladores**
- ▶ Desvantagens
 - **Objectos complexos** possuem **especificações complexas**
 - Uma especificação incompleta poderá ter um efeito desastroso



Computador Virtual Universal [UVC]

- ▶ **Variante** da **Emulação** e Encapsulamento
- ▶ Preservação do **bitstream original** + **descodificador** compilado para uma máquina virtual
 - e.g. Java Virtual Machine
- ▶ A máquina virtual deverá ser facilmente reproduzida qualquer computador futuro
 - Está limitada a um conjunto finito de operações
- ▶ Estratégia ainda em estudo
- ▶ Necessita de um **visualizador**





UNIVERSIDADE
DO
MINHO

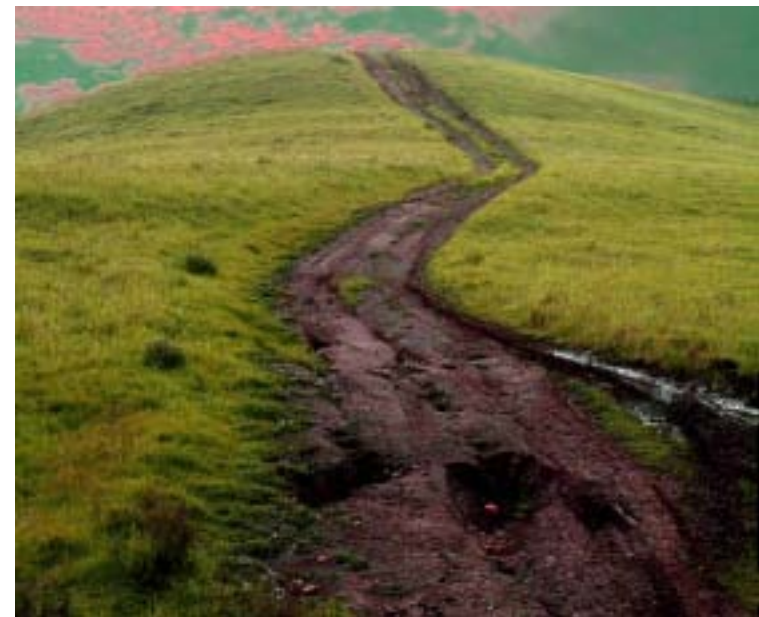
Jornadas de Trabalho
A gestão da informação na era digital

Estratégias de preservação

Preservação de
tecnologia



Preservação da
informação



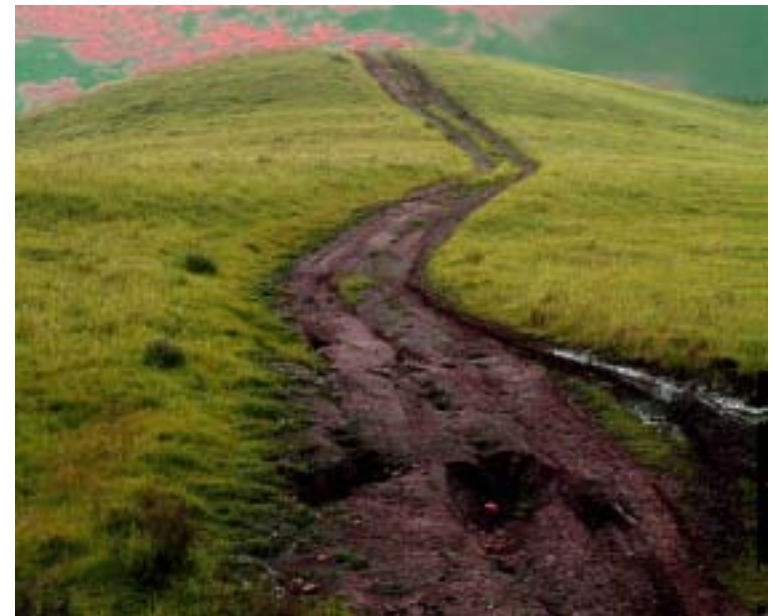


UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

Estratégias de preservação

Preservação da informação





Migração de formatos

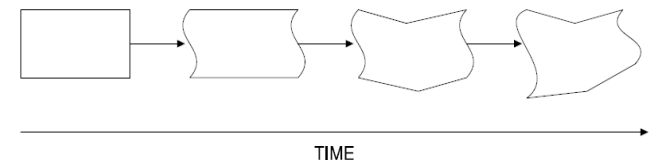
Transferência periódica de **material** digital de uma **configuração** de hardware/software para outra mais **actual**

▶ 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 contínua**
- **Dispendiosa** a longo-prazo





Variantes da Migração

- ▶ Print-out
 - **Transferir** o objecto digital para um meio **analógico** mais persistente (e.g. imprimir em papel)
- ▶ Actualização de versões
 - **Actualização** dos **formatos** utilizando software **retro-compatível**
 - e.g. utilização do Word 6 para actualizar um documento criado com o Word 5
- ▶ Migração para formatos semelhantes
 - Conversão para **formatos concorrentes**
 - e.g. converter um documento Word para PDF
- ▶ **Normalização**
 - Migração para um **número reduzido de formatos** compatíveis



Notas finais

- ▶ **Não existe uma solução única** e adequada a todos os tipos de material
- ▶ A decisão sobre que estratégia de preservação adoptar **depende de vários factores**:
 - Simplicidade de implementação a curto e longo-prazo
 - Custo
 - Restrições legais
 - Satisfação da comunidade de interesse
 - Tipo de material
- ▶ **Emulação** versus **Migração**
 - **Emulação** para preservação de objectos **dinâmicos/interactivos** (e.g. software, animações flash)
 - **Migração** para a preservação dos restantes **objectos digitais** (e.g. documentos texto, imagens, bases de dados)



Notas finais

▶ Autenticidade

- Capacidade de se conseguir **demonstrar** que um objecto digital **é aquilo que se propõe ser**
- Capacidade de **identificar** os **elementos diplomáticos** que permitem aferir se um dado objecto é autêntico
- Metainformação de preservação - **PREMIS**

▶ Avaliação/selecção vs Arquivar tudo

- O **preço/Mb decresce** com o tempo
- O preço do **homem/mês aumenta** com o tempo
- “Prefiro ser condenado por algo que fiz do que por algo que não fiz” (Lesk, Internet Archive)



Notas finais

- ▶ **Avaliação** da preservação
 - Identificação das **propriedades significativas** que deverão ser mantidas intactas para que o objecto possa ser considerado autêntico
 - O **conjunto das propriedades** significativas **não é absoluto** nem estático, depende de:
 - natureza da organização,
 - características dos objectos
 - requisitos da comunidade de interesse
- ▶ Quanto **mais propriedades** definirmos **maiores** serão os **requisitos** e os custos de preservação
- ▶ Grupos de propriedades significativas:
 - **Aparência, conteúdo, contexto, estrutura**



UNIVERSIDADE
DO
MINHO

Jornadas de Trabalho
A gestão da informação na era digital

... para mais informação...



Disponível em Acesso-Livre

[HTTP://HDL.HANDLE.NET/1822/5820](http://hdl.handle.net/1822/5820)