

Ficheiros & Metadados



Ficheiros digitais



1. Ficheiros digitais

1.1 Formatos abertos e proprietários

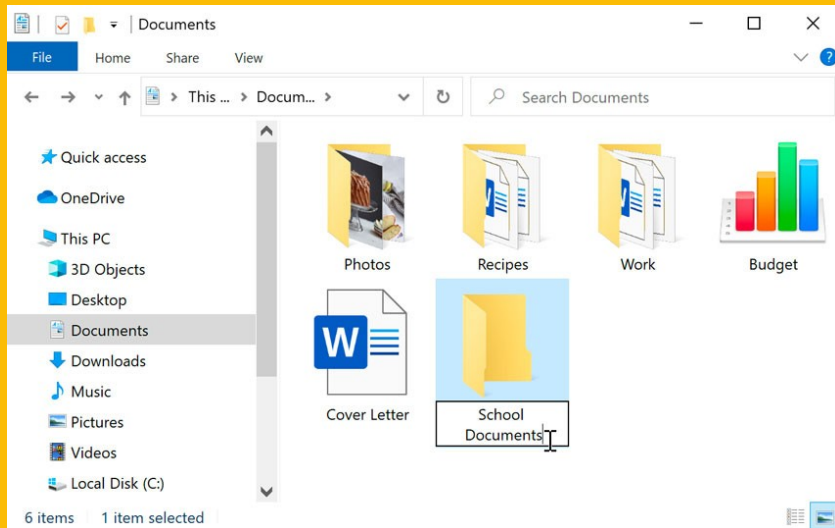
1.2 Formatos e extensões comuns em diferentes áreas

1.3 Operações e comandos úteis para edição em lote

1.4 Normas e algumas recomendações

1.5 Preservação de longa duração

1. Ficheiros digitais



São a **unidade básica de armazenamento de informação** digital manipulada por software e hardware. São conjuntos estruturados de dados armazenados em formato eletrónico. Pode conter diferentes tipos de dados (texto, imagens, áudio, etc.), tem um formato e é identificável por um nome e geralmente uma extensão (ex.: .pdf, .mp3, .jpg).

Até a pasta vazia tem metadados

1.1 Formatos **abertos** e proprietários

Formatos abertos são tipos de ficheiros (extensões) cujas **especificações técnicas são públicas e livremente utilizáveis**, sem restrições de licenças proprietárias. Isso significa que qualquer pessoa ou empresa pode criar software para ler, editar ou processar esses ficheiros sem pagar royalties ou depender de um único fornecedor.

Especificações públicas: Documentação técnica acessível a todos (ex.: PDF/A para documentos).

Livre de patentes: Não exige pagamento para uso (ex.: PNG para imagens).

Interoperabilidade: Funciona em múltiplos sistemas (Windows, Linux, macOS).

Longevidade: Garante que os dados permaneçam acessíveis no futuro (evita obsolescência).

E o Aberto (público) + Proprietário?

1.1 Formatos abertos e **proprietários**



Formatos **controlados por uma empresa** ou entidade específica, com especificações não públicas ou que exigem software ou licenças pagas para serem totalmente acessíveis. Esses formatos são frequentemente otimizados para funções exclusivas de um programa, criando dependência do ecossistema do fabricante.

1.1 Formatos abertos e **proprietários**

Vendor lock-in: dependente de um único fornecedor de software, hardware, serviço ou tecnologia) e enfrenta dificuldades técnicas, legais ou financeiras para mudar para outro.

Obsolescência: Se a empresa abandonar o formato, os dados podem ficar inacessíveis.

Falta de transparência: Risco de backdoors ou incompatibilidade futura.

Restrições de uso: Podem ter DRM (Digital Rights Management) ou exigir pagamento de royalties.

Código fechado: As especificações técnicas não são divulgadas ou são patenteadas.

Dependência de software: Só funcionam corretamente com programas específicos (ex.: PSD no Photoshop, só o AutoCAD abre DWG sem erros).

Custos: Licenças dispendiosas

Interoperabilidade limitada

Formato...	Livre	Proprietário
Office (Texto, cálculo, apresentação)	ODF (ODT/ODS/ODP), <i>PDF</i>	<i>DOC/XLS/PPT</i> <i>DOCX/XLSX/PPTX</i>
Áudio (codec/container)	FLAC, AAC, <i>MP3</i> , WAV, OGG (Vorbis)	WMA, M4A,
Vídeo (codec/container)	WebM, MKV, OGG (Theora)	MP4, AVI, WMV, MOV
SIG	GPKG, Geojson, GeoTIFF, GML, GPX,KML, DXF	SHP (shapefile), DWG, GDB (geodatabase)
Raster	TIFF, XCF, PNG, WebP, JPEG, JPEG2000, GIF, DNG	RAW, PSD, BMP
Vetorial	SVG	AI, CDR
Outros	HTML/XHTML, XML, MD TXT, CSV/TSV, RDF, JSON	

1.2 Formatos e extensões comuns em diferentes áreas

Standards (Padrões Formais): definidos por organizações de normalização, governos ou consórcios, com processos transparentes e abertos à participação.

Documentação pública: Especificações acessíveis a todos.

Aprovação por **entidades reconhecidas**:
Ex.: ISO, IEEE, W3C.

Objetivo: Garantir **interoperabilidade** universal, durabilidade.

Exemplos: HTML5 (W3C), PDF/A (ISO 19005)

De Facto Standards (Padrões de Fato): são dominantes no mercado. Tornam-se "padrão" devido ao uso massivo.

Popularidade: Adotados por conveniência ou monopólio.

Pode ser **aberto ou fechado**

Riscos: **Dependência** de um único fornecedor.

Exemplos: Qwerty, MP3 e Windows

1.2 Formatos e extensões comuns em diferentes áreas

OGC = Open Geospatial Consortium

W3C = World Wide Web Consortium

ISO = International Organization for Standardization

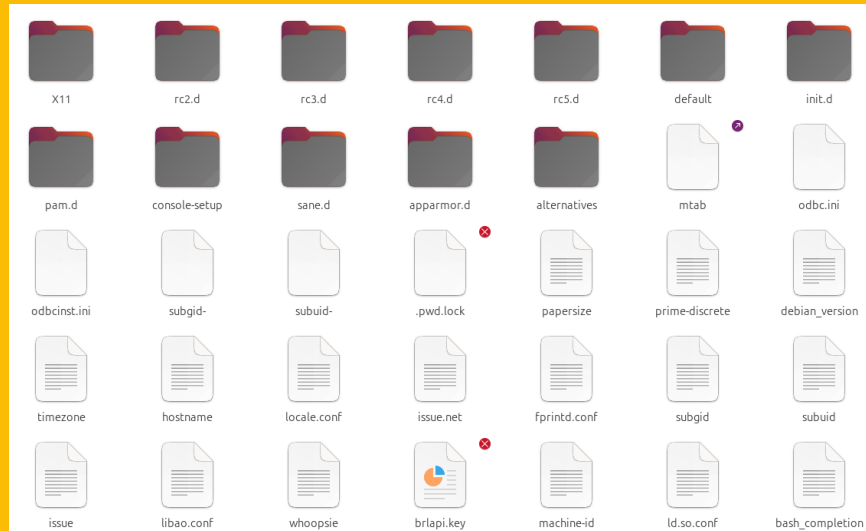
IEFT = Internet Engineering Task Force

IEEE = Institute of Electrical and Electronics Engineers

IEC = International Electrotechnical Commission

Formato...	Standards	De facto Standards
Office (Texto, cálculo, apresentação)	ODF (ISO), OOXML (ISO), PDF/A (ISO)	DOC/XLS/PPT DOCX/XLSX/PPTX
Áudio (codec/container)	FLAC (Xiph), AAC (ISO/MPEG), MP3 (ISO/MPEG), WAV (MS/IBM/ISO),	MP3 (ISO/MPEG), WMA (MS), AAC (ISO/MPEG), OGG Vorbis (Xiph)
Vídeo (codec/container)	MP4 (ISO/IEC), VPg (Google) h264 e h265 (ISO/IEC), AV1 (IEFT), MKV (IEFT)	MOV (Apple), WebM (Google), VPg (Google), AV1 (IEFT)
SIG	Geojson (IEFT); GPKG, GeoTIFF, GML e KML (OGC)	GDB e SHP (ESRI), GeoTIFF e GPKG (OGC), Geojson (IEFT)
Raster	TIFF (Adobe/ISO), PNG (W3C/ISO), JPEG e JPEG2000 (ISO/IEC), WebP (IEFT)	JPEG (ISO/IEC) RAW (Canon, Nikon, Sony, etc.), PSD (Adobe), GIF, HEIF/HEIC (Apple)
Vetorial	SVG (W3C), PDF (ISO), DWG (ISO)	AI (Adobe), CDR (Corel), DXF (Autodesk), WMF/EMF (Microsoft)
Outros	HTML, XML, RDF (W3c)	HTML, XML (W3c), CSV

1.3 Operações e comandos úteis para edição em lote



A edição em lote (ou *batch editing*) refere-se ao processamento automático de múltiplos ficheiros de uma só vez, aplicando as mesmas operações ou transformações a todos eles, sem necessidade de intervenção manual para cada arquivo individual.

1.3 Operações e comandos úteis para edição em lote

Pode ser útil instalar o Microsoft Powertoys (PoerRename) ou usar Ubuntu (GNU/Linux) mas entretanto vamos experimentar a **linha de comandos!**

TAREFA!

1-ficheiros-batch.pdf

1.4 Normas e algumas recomendações

Digitalização em formatos abertos, estáveis e amplamente suportados (TIFF, PDF/A, CSV, XML).

Uso de metadados padronizados (Dublin Core, CIDOC CRM, LIDO).

Armazenamento em múltiplas localizações (3-2-1 backup rule).

Monitorização da integridade e migração periódica de formatos obsoletos.

Plano de preservação digital (com políticas claras).

1.4 Normas e algumas recomendações – **Armazenamento Regra 3-2-1**

Garantir redundância e dispersão geográfica!

3 cópias dos dados (original + 2 backups) + **2** tipos diferentes de suporte (ex.: HD externo + nuvem) + **1** cópia armazenada em local externo (fora do museu/instituição).

Cópia 1 (Original): Fotos (TIFF) no servidor interno [SSD] do museu

Cópia 2 = HD externo (armazenado em cofre à prova de fogo)

Cópia 3 = Cloud (Amazon, Archivematica, etc.) num datacenter (outro país?)

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Imagens

TIFF (**sem compressão** ou com compressão lossless)

PNG (para imagens raster com transparência).

JPEG 2000 (se necessário compressão sem perdas).

RAW (**mas exige conversão periódica**). > Converter para DNG

SVG (para gráficos vetoriais).

Evitar: JPEG (com perdas), GIF (limitado a 256 cores).

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Documentos e Textos

PDF/A (padrão ISO para preservação de documentos).

TXT (texto simples, sem formatação).

ODT (OpenDocument, alternativa aberta ao DOCX).

XML (para textos estruturados).

Evitar: DOCX, PPTX (podem ter dependências proprietárias).

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Áudio

WAV (sem compressão)

FLAC (compressão sem perdas, boa para economia de espaço).

AIFF (alta qualidade, mas menos comum).

Evitar: MP3, AAC (com perdas).

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Vídeo

MKV (container flexível para vídeo sem perdas).

FFV1 (codec sem perdas, recomendado para arquivo).

Motion JPEG 2000

MXF com JPEG 2000 lossless

Evitar: formatos proprietários.

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Metadados

XML (padrão para metadados estruturados).

JSON (leve e legível).

CSV (para dados tabulares simples).

RDF/OWL (para linked data e semântica).

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Metadados

XML (padrão para metadados estruturados).

CSV (para dados tabulares simples).

RDF/OWL (para linked data e semântica).

Bases de Dados

SQLite (para bases de dados leves e portáteis).

CSV/XML (exportação periódica de dados).

1.4 Normas e algumas recomendações – **Formatos dos ficheiros**

Mas há mais em...

Formatos_comparacao.ods

TAREFA!

1. Assinalar aqueles que já usa ou sabe criar (verde).
2. Assinalar aqueles que podem ser úteis mas não sabe como criar (vermelho)

1.5 Preservação de longa duração - **Plano de Preservação Digital**

Plano de Preservação Digital (PPD) é um documento estratégico que define as **políticas, responsabilidades, métodos e recursos** necessários para garantir a *longevidade, autenticidade, integridade e acessibilidade* de objetos digitais ao longo do tempo.

Para quê? Para quem? Como? Quando? E quanto custa!

regulamentado por uma combinação de normativos legais, diretrizes nacionais e padrões internacionais

1.5 Preservação de longa duração - **Plano de Preservação Digital**

1. Objetivos e Âmbito

1.1 Propósito (ex.: acesso público, pesquisa, conformidade legal) e objetivos da preservação (ex.: preservar acervos digitais por 50+ anos)

1.2 Coleções incluídas

2. Formatos de Arquivo

2.1 Lista de formatos recomendados

2.2 Documentação das escolhas

2.3 Critérios para migração de formatos obsoletos

1.5 Preservação de longa duração - **Plano de Preservação Digital**

3. Metadados

3.1 Padrões adotados (ex.: Dublin Core, METS, PREMIS)

3.2 Metadados obrigatórios para cada objeto digital.

3.3 Formatos preferenciais

4. Armazenamento e Segurança

4.1 Estratégias de backup

4.2 Verificações de integridade (ex.: checksum)

4.3 Infraestrutura tecnológica

1.5 Preservação de longa duração - **Plano de Preservação Digital**

5. Ações de Preservação

5.1 Migração de formatos obsoletos

5.2 Normalização para formatos de preservação

5.3 Documentação de migrações e alterações

6. Gestão de Risco

6.1 Identificação de ameaças (ex.: obsolescência tecnológica, corrupção de dados)

6.2 Planos de contingência (ex.: migração periódica para novos suportes)

1.5 Preservação de longa duração - **Plano de Preservação Digital**

7. Políticas de acesso

7.1 Dados restritos e dados públicos

7.2 RGPD

7.3 Política de licenciamento

8. Responsabilidades e Recursos

8.1 Equipa responsável (curadores, IT, arquivistas)

8.2 Orçamento e recursos necessários para infraestrutura e atualizações

1.5 Preservação de longa duração - **Plano de Preservação Digital**

9. Monitorização e Revisão

9.1 Calendário para avaliação periódica do plano (ex.: bienal)

