



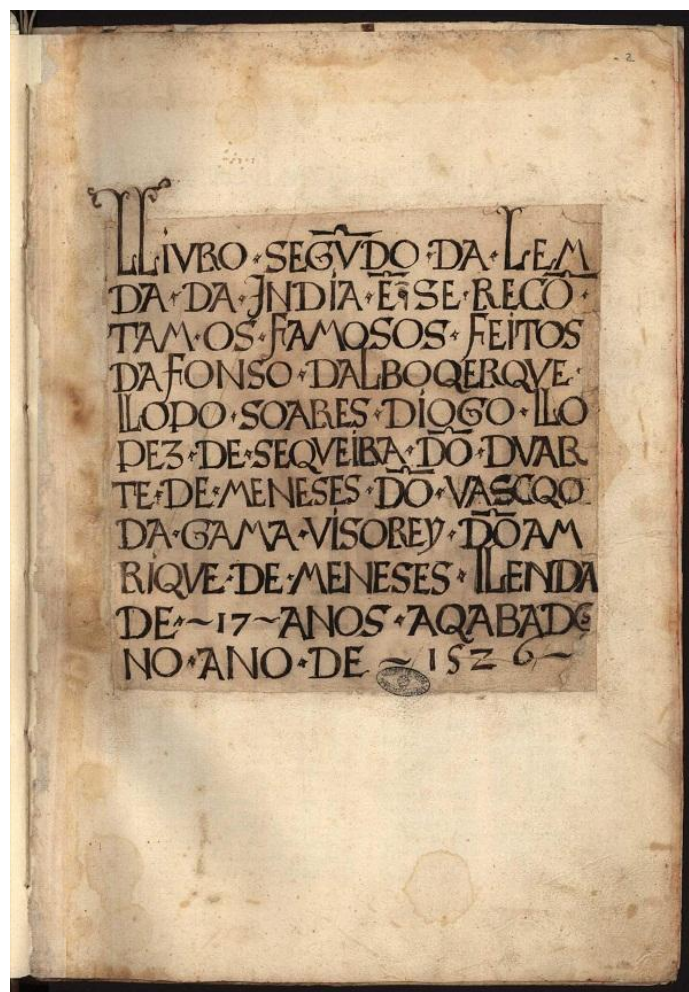
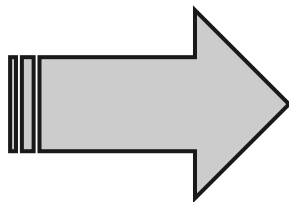
Formatos

SOS Digital: Tópico 4

O que são formatos?

1. conjuntos de informação que podem ser armazenados como **ficheiros**
1. Um formato é a interpretação de bits (0 e 1) de uma forma humanamente perceptível através de um conjunto de **regras sintáticas e semânticas** que transformam unidades binárias (0, 1) em informação.
1. Com um determinado propósito funcional:
 - a. para escrever texto para fazer apresentações, para realizar cálculo, para capturar e armazenar imagem, para registar música, para criar bases de dados, para criar filmes, etc...
 - b. determinando propriedades de apresentação, manipulação, comportamentos ...

ffd8ffe000104a46
4946000102010083
00830000ffed0fb0
50686f746f73686f
7020332e30003842
494d03e90a507269
6e7420496e666f00
0000007800000000
0048004800000000
02f40240ffeeffee
0306025203470528
03fc000200000048
00480000000002d8
0228000100000064
0000000100030...



PRESERVAÇÃO
COMUM de
PATRIMÔNIO DIGITAL





O que são formatos?

- Este conjunto de regras usados para processar e representar 0 e 1 são desenvolvidos por empresas através de I&D.
- Os formatos podem ser proprietários quando as regras criadas por uma empresa estão fechadas ao público, ou seja, não é possível ter conhecimento dessas regras,
- Podem ser proprietários e abertos (o caso mais conhecido é pdf)
 - Aberto significa que a documentação que explicita as regras do formato está publicamente disponível
- Podem ser normalizados e abertos
 - normalizado significa que as regras do formato se baseiam em normas desenvolvidas por agências nacionais e internacionais e não em especificações desenvolvidas apenas por uma empresa ou cluster de empresas



Formatos e software

- Os formatos são suportados por softwares específicos para esse efeito:
 - Por exemplo, o microsoft word suporta formatos docx, doc, txt, odt, pdf, xml, rtf, etc
 - Mas nenhum software suporta TODOS os formatos existentes.
- Há portanto uma relação estreita, por vezes inseparável, entre formatos e software
- Muitas aplicações suportam formatos de formas diferentes:
 - podem permitir abrir e processar a informação num determinado formato
 - podem simplesmente permitir visualizar a informação sem possibilidade de processamento



Formatos e ficheiros

- a informação organizada de acordo com as regras específicas do formato é materializada em ficheiros.
- O formato utilizado para organizar a informação contida no ficheiro é normalmente identificável pela extensão do ficheiro:
 - xxx.**jpg**; xxx.**wav**; xxx.**docx**...xxxx.**odt**, xxx.**midi**; xxx.**htm**, etc
- Ou tipo mime (internet media type) indica o tipo de dados contido no ficheiro
 - video/avi; mage/bmp: application/msword, etc

formatos e software

- Usamos diariamente formatos através de aplicações informáticas específicas para esse efeito.
- Não nos preocupamos em saber quais as regras implícitas nos formatos que usamos porque sabemos que o software que temos à disposição se encarrega desse problema.
- Os formatos vão sendo substituídos bem com as versões das aplicações o que nos permite a sua utilização de forma descontraída!



Formatos, software e problemas...

- Quando precisamos de preservar informação escrita num formato que desapareceu,
- Quando a aplicação que gere o formato se torna obsoleta
- gera-se isolamento tecnológico e
- Se não conhecermos as regras do formato teremos fortes problemas de acesso e, inevitavelmente, de processamento da informação codificada nesse formato específico e suportada por essa aplicação específica



Formatos e preservação

- Considerando que os formatos + software são elementos basilares da informação digital, são também objetos preferenciais de ações de preservação.
- Considera-se boa prática a opção por **formatos abertos**, preferencialmente normalizados mas, sobretudo, com documentação detalhada publicamente disponível.
- É boa prática tomar precauções relativamente a formatos utilizados de forma a evitar consequências de isolamento tecnológico e consequente perda de informação



publicamen Formatos e preservação te

ações aconselháveis:

- Determinar à partida quais os formatos a utilizar na instituição, optando por formatos abertos
 - Para tal definir quais as propriedades e comportamentos significativos que se pretende manter
- Se os formatos já utilizados forem proprietários considerar alterá-los para formatos abertos.
- Se isso não for possível, monitorizar a evolução de mercado dos formatos e do software de maneira a evitar isolamento tecnológico
- No caso de receber objetos digitais provenientes de entidades externas, determinar formatos aceitáveis para esses objetos



RNID - RCM 91/2012

- O regulamento nacional de interoperabilidade digital (RNID - RCM 91/2012) ajuda nesse sentido. Este regulamento estipula um conjunto de formatos abertos de utilização obrigatório ou aconselhável. Tem de ser interpretado juntamente com a Lei 36/2011 que estabelece a obrigatoriedade de usar normas abertas na AP.
- O seu cumprimento ajuda a diminuir o risco relativamente a obsolescência tecnológica.
- O RNID não cobre todas as situações. Apenas as utilizações mais comuns. Não é aplicável no caso de formatos já existentes que não sejam compatíveis com a lista publicada.



Problemas possíveis!

Os formatos abertos podem apresentar alguns problemas:

- Podem não ter tantos recursos e oferecer tantas possibilidades como os formatos proprietários
- Podem ser menos eficientes consumindo mais recursos computacionais para obter o mesmo resultado
- Podem não ter assistência fácil no mercado.



selecionar formatos para preservação

- Há um conjunto de critérios definidos para avaliar formatos, por ex:
 - Estabilidade
 - qualidade da documentação
 - dependências tecnológicas
 - domínio público
 - custo??

Para mais informação consultar os seguintes recursos:

- <http://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>
- <http://sobre.arquivo.pt/how-to-participate/recommendations-for-web-authors-to-enable-web/adequate-formats-for-preservation>
- http://www.interpares.org/display_file.cfm?doc=ip2_gs11_final_report_english.pdf



Outros recursos

Unified Digital Format Registry (UDFR) Em desenvolvimento...

<http://www.udfr.org/>

PRONOM - atualmente com 800 formatos registados.

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx#>

Biblioteca Congresso

<http://www.digitalpreservation.gov/formats/intro/intro.shtml>

Tipos MIME

<http://www.iana.org/assignments/media-types/media-types.xhtml>



Digital Curation Center

<http://www.dcc.ac.uk/r>

Digital Preservation Coalition

<http://www.dpconline.org/advice/technology-watch-reports>