

Preservação Digital na Universidade do Porto

Augusto Ribeiro

João Miguel Ferreira Pereira

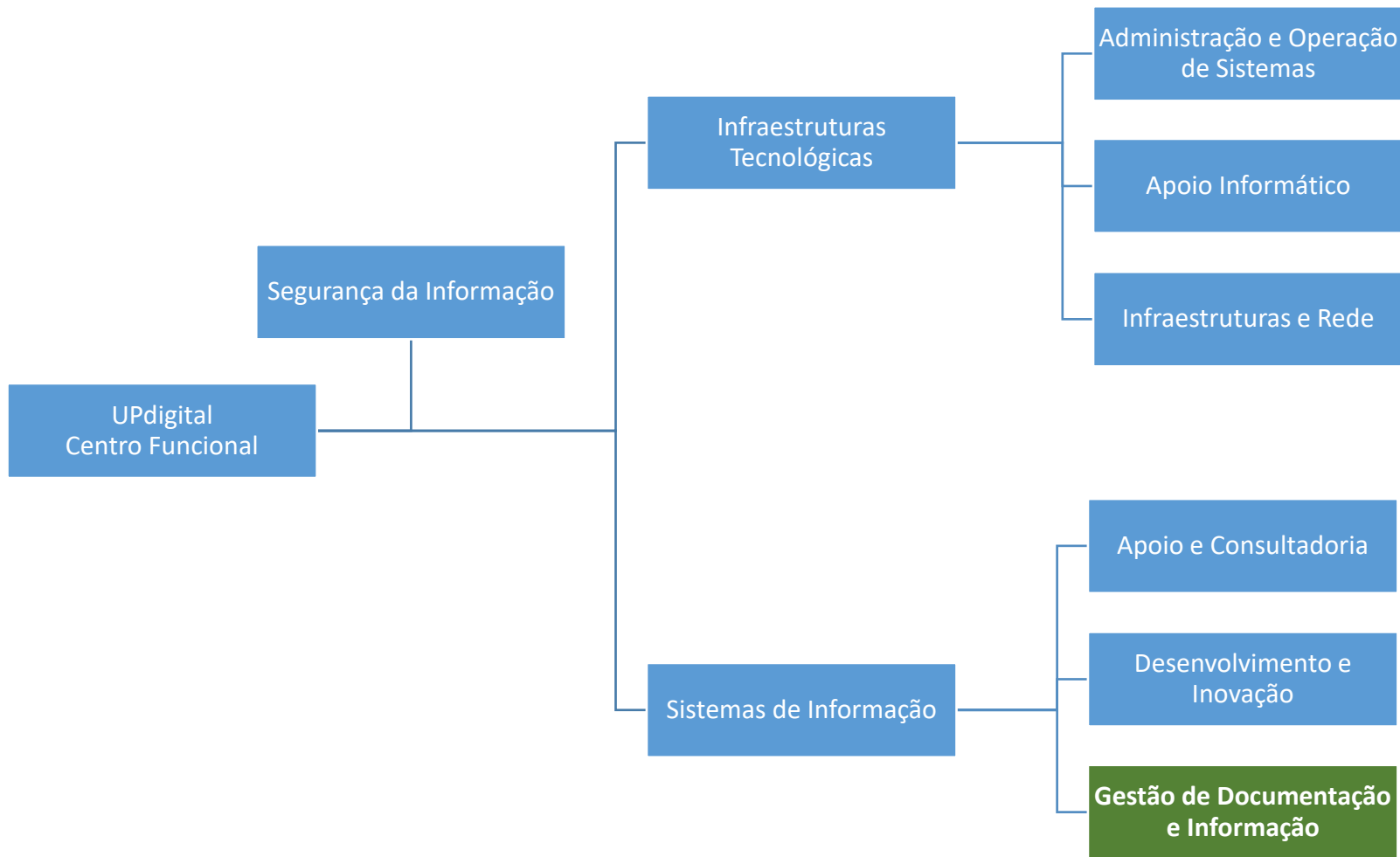
Gestão de Documentação e Informação



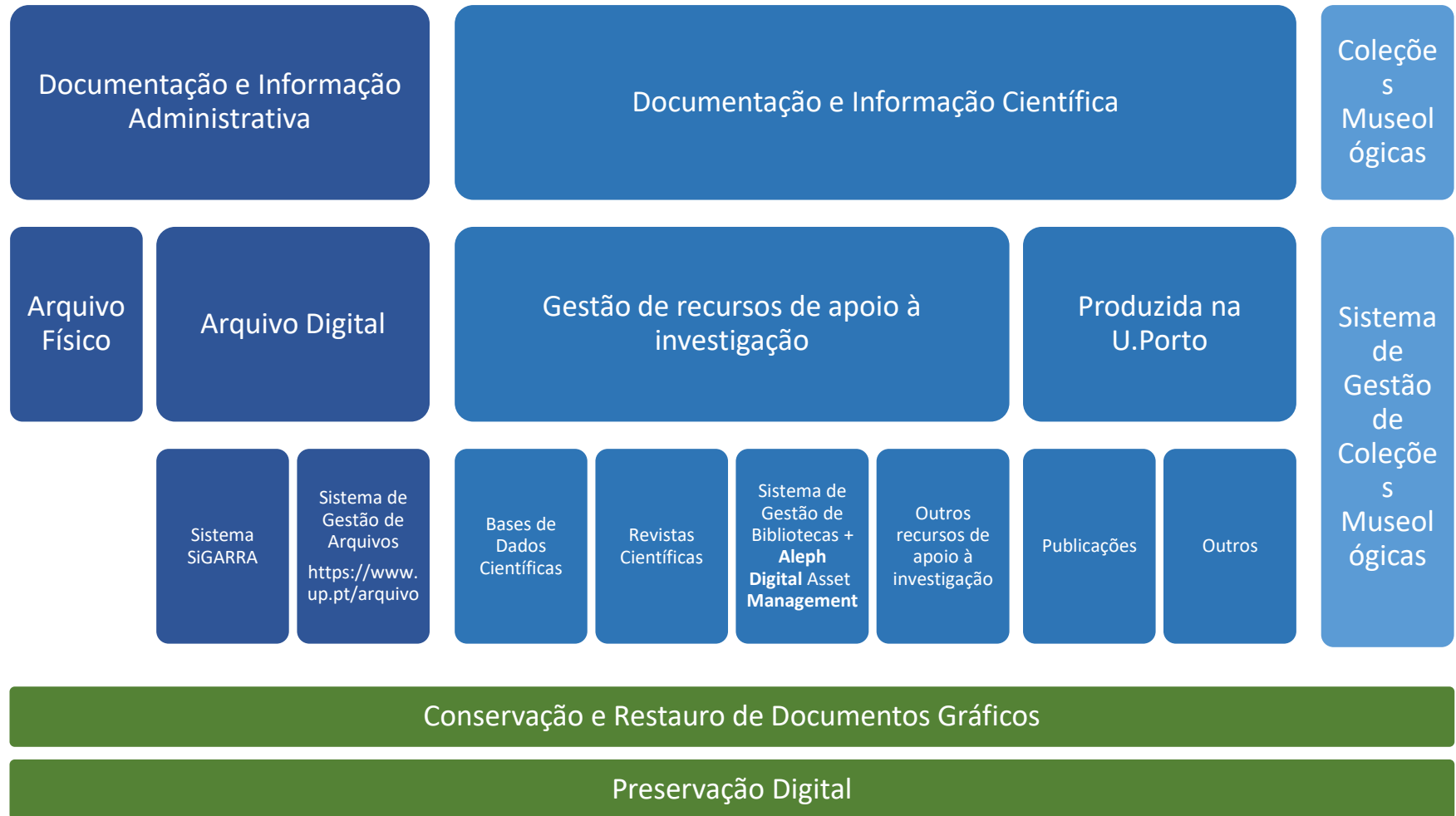
UNIVERSIDADE
DO PORTO
DIGITAL



UPdigital



Gestão da Documentação e Informação (GDI) - Competências



Sistema de Preservação Digital: Necessidade

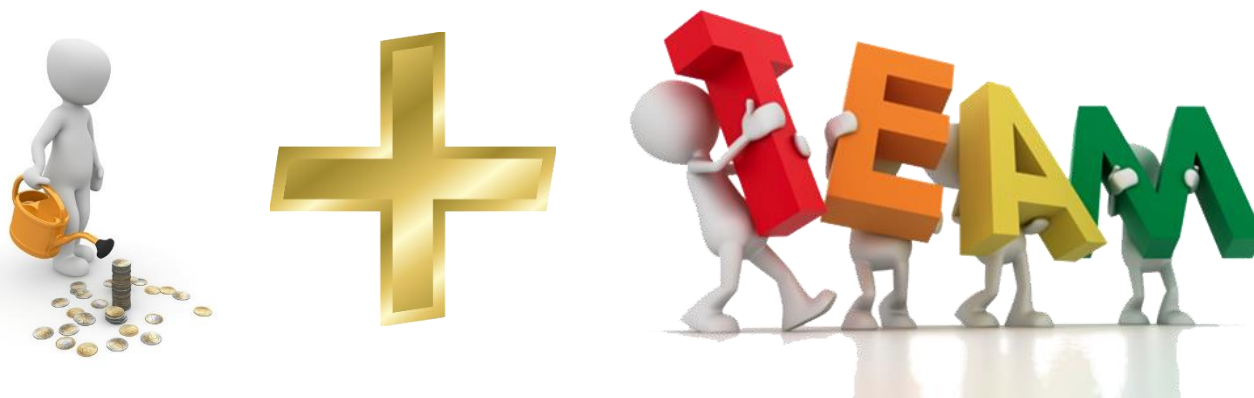
Formatos de documentos digitalizados desde 2004

FORMATO	SUPOORTE	IDADE
TIFF , sem compressão, Escala cinza, 300 DPI	DVD-R 4,7 GB	+ 15 ANOS
PNG , RGB 24 bit, 600 DPI	DVD-R 4,7 GB; DISCOS EXTERNOS 200 GB	+ 15 ANOS
TIFF , sem compressão, RGB 24 bit, 300 / 400 DPI	DISCOS EXTERNOS 1 Tb; DISCOS EXTERNOS 2 Tb	- 15 ANOS
PDF/A-1a , com OCR [PDF v. 1.4]	DISCOS EXTERNOS 1 Tb; DISCOS EXTERNOS 2 Tb	- 15 ANOS

Sistema de Preservação Digital: Oportunidade

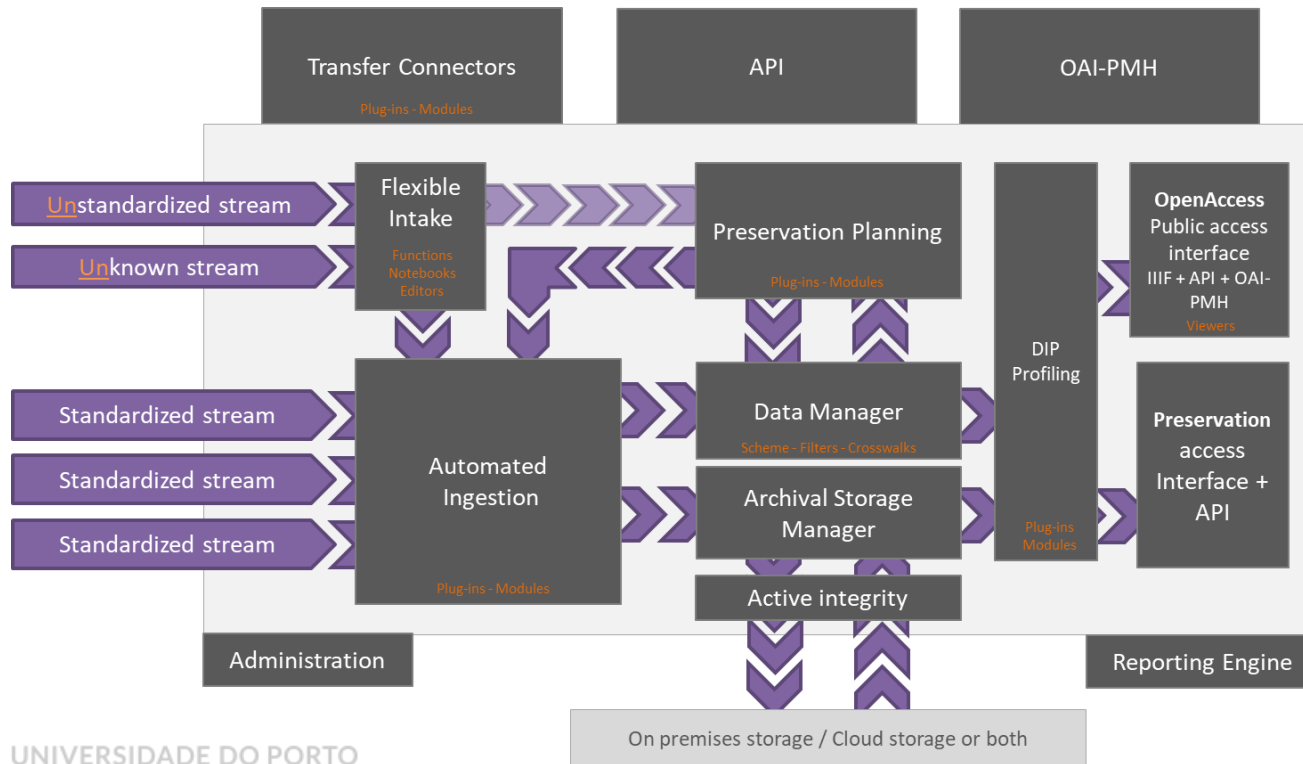
Aquisição:

- Financiamento SAMA, no âmbito do projeto **U.Norte Gateway**;
- Alinhamento das ECs da U.Porto para a convergência de sistemas.



Sistema de Preservação Digital : Requisitos

- Conformidade com a norma ISO 14721 (*Open Archival Information System OAIS*)
- Disponibiliza as características necessárias à implementação da norma ISO 16363 (*Audit and certification of trustworthy digital repositories*)



Sistema de Preservação Digital: Requisitos

- Normas:
 - PREMIS (*PREservation Metadata: Implementation Strategies*)
 - METS (*Metadata Encoding and Transmission Standard*)
 - Dublin Core
 - MODS (*Metadata Object Description Schema*)
 - EAD (*Encoded Archival Description*)
 - ISAD(G) (*General International Standard Archival Description*)
 - MARC21 (*Machine-Readable Cataloging 21*)

Sistema de Preservação Digital: Requisitos

- Criação de *workflows* personalizados para a ingestão de SIP's (*Submission Information Package*)
 - Com controle de versões;
 - Com verificação de existência de ficheiros;
 - Verificação da tipologia de ficheiros;
 - ...
- Ingestão automatizada de SIP's
- Caracterização, validação, conversão de formatos (DROID, JHOVE)
- Criação de DIP's (*Difusion Information Package*)
 - Ex. transformação de TIFF para PDF com OCR a pedido em tempo real

Sistema de Preservação Digital: Requisitos

- Pesquisa no texto integral (*Full-text search*) dos documentos PDF (OCR), texto, etc.
- Produção de relatórios
- Escalável
- Manutenção e apoio



Sistema de Preservação Digital: Referências

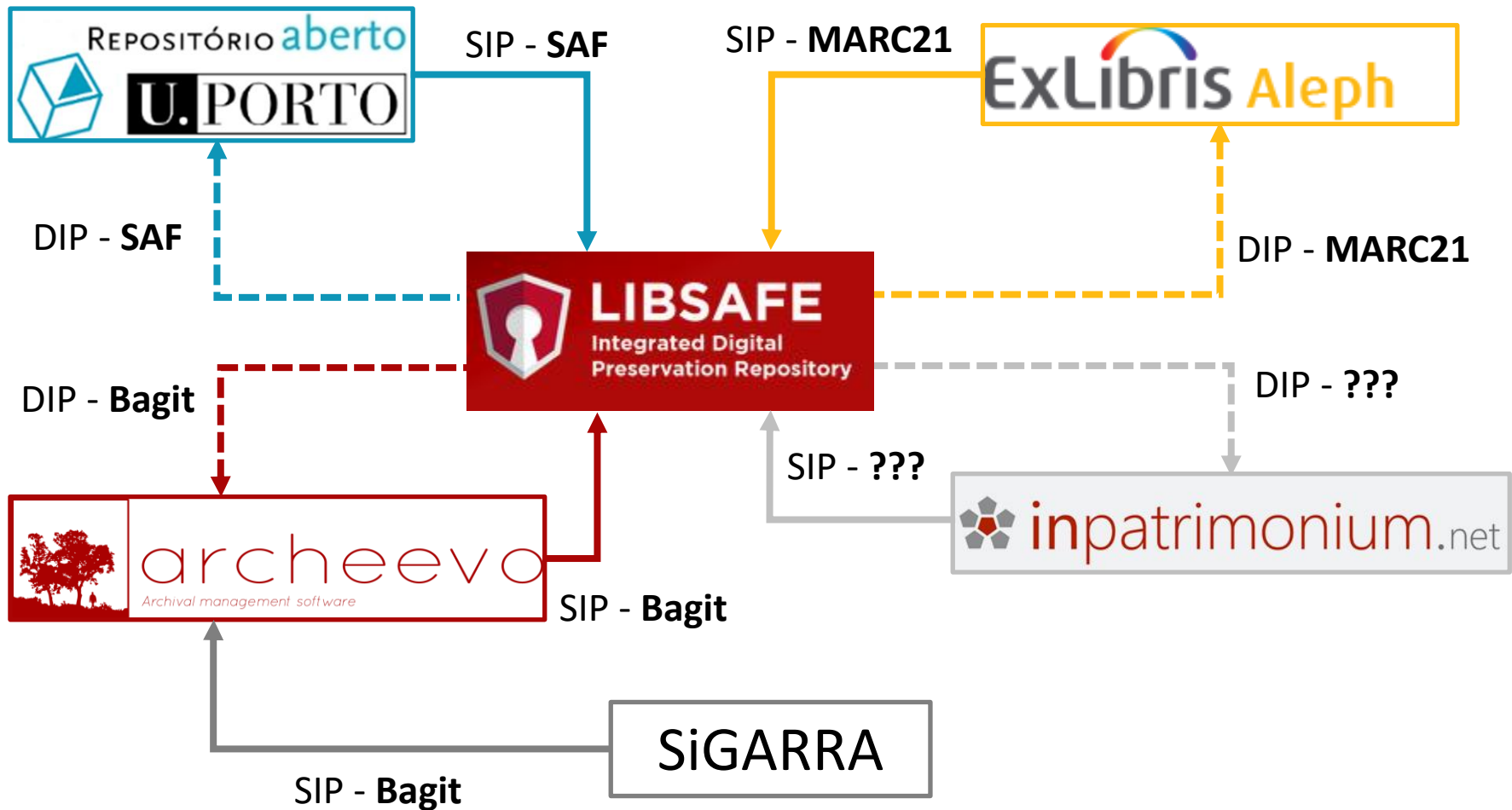


Sistema de Preservação Digital: Decisão

- Consideradas 3 soluções
- Aplicada a análise dos requisitos descritos anteriormente
- Análise económica a 5 anos
- Escolha LibSafe



U.PORTO - Modelo projetado para a Preservação Digital



U.Porto: Políticas & Boas Práticas de preservação digital

Política de Preservação Digital (orientações):

- adopção de esquemas de meta-informação *standard* (Dublin Core, ISAD-G, MARC 21)
- utilização de formatos de ficheiros de preservação
- armazenamento redundante dos ficheiros
- migração de suportes



U.Porto: Políticas & Boas Práticas de digitalização

Guia de Digitalização (orientações):

- nomenclatura de ficheiros
- formato matriz e derivada (quando aplicável)
- resolução
- profundidade de cor
- testes de pré-digitalização

... por tipologia documental



Orientações para a criação correta de nomenclaturas dos ficheiros

1. evitar a utilização de caracteres especiais no nome dos ficheiros

ex. ~ ! @ # \$ % ^ & * () ` ' ; < > ? , [] { } ‘ “”

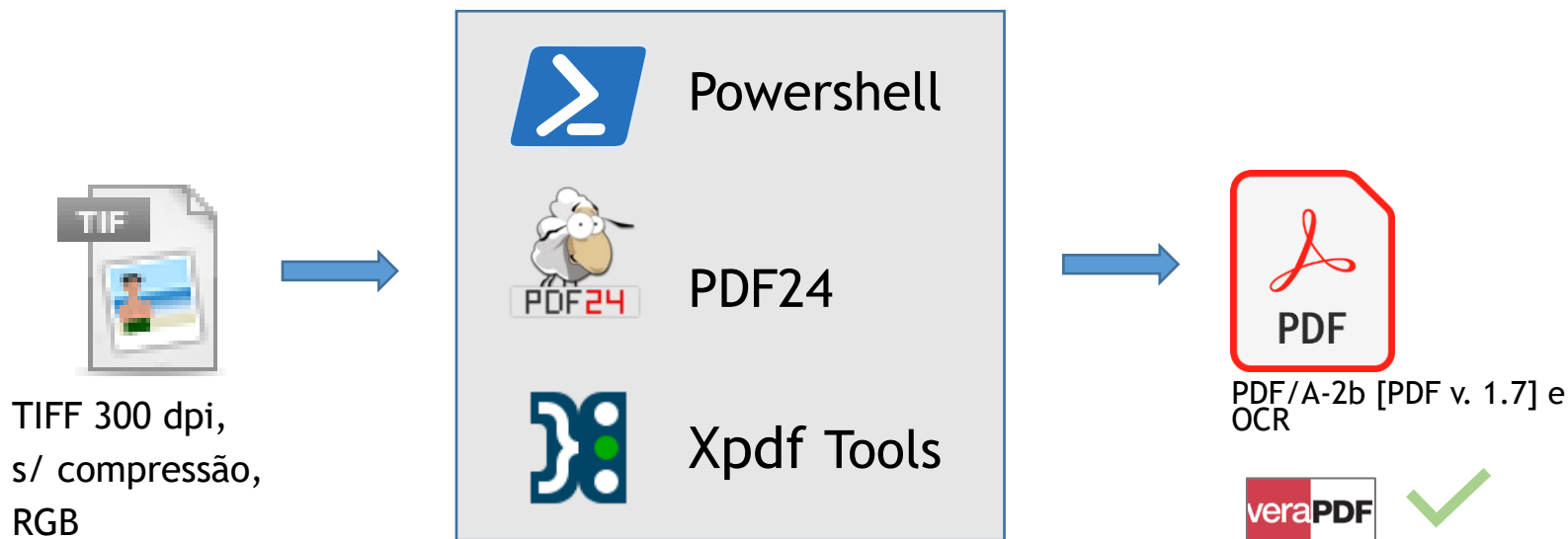
2. evitar nomes muito compridos

Forma incorreta	Forma correta
<i>João Pereira.pdf</i>	Joao_Pereira.pdf ou JoaoPereira.pdf
<i>Anuário da Universidade do Porto - Ano escolar de 1958-1959.pdf</i>	AUP_13.pdf

U.Porto: Políticas & Boas Práticas de digitalização

Promoção da utilização do formato PDF/A nos projetos de digitalização

Utilização de ferramentas para a criação de ficheiros no formato PDF/A-2b



Sistema de Preservação Digital: Plano de Preservação Digital

Objetivo

Preservação digital dos ficheiros de imagem resultado de processos de digitalização da documentação que constitui o Arquivo da Academia Politécnica do Porto e séries com valor histórico/científico do Arquivo da Reitoria da U.Porto.

Análise de risco

- imagens digitalizadas entre 2004 e 2013 que se encontravam em suporte DVD-R e discos externos

Formato(s) principais de ficheiros:

- TIFF, escala cinza a 300 dpi (matriz)
- PNG, RGB, 24 bits a 600 dpi (matriz)
- TIFF, RGB, 24 bits a 300 dpi (matriz)
- PDF/A (derivada)

Sistema de Preservação Digital: Plano de Preservação Digital

Esquema de metainformação descritiva escolhido

- Dublin Core com crosswalk para ISAD(G)

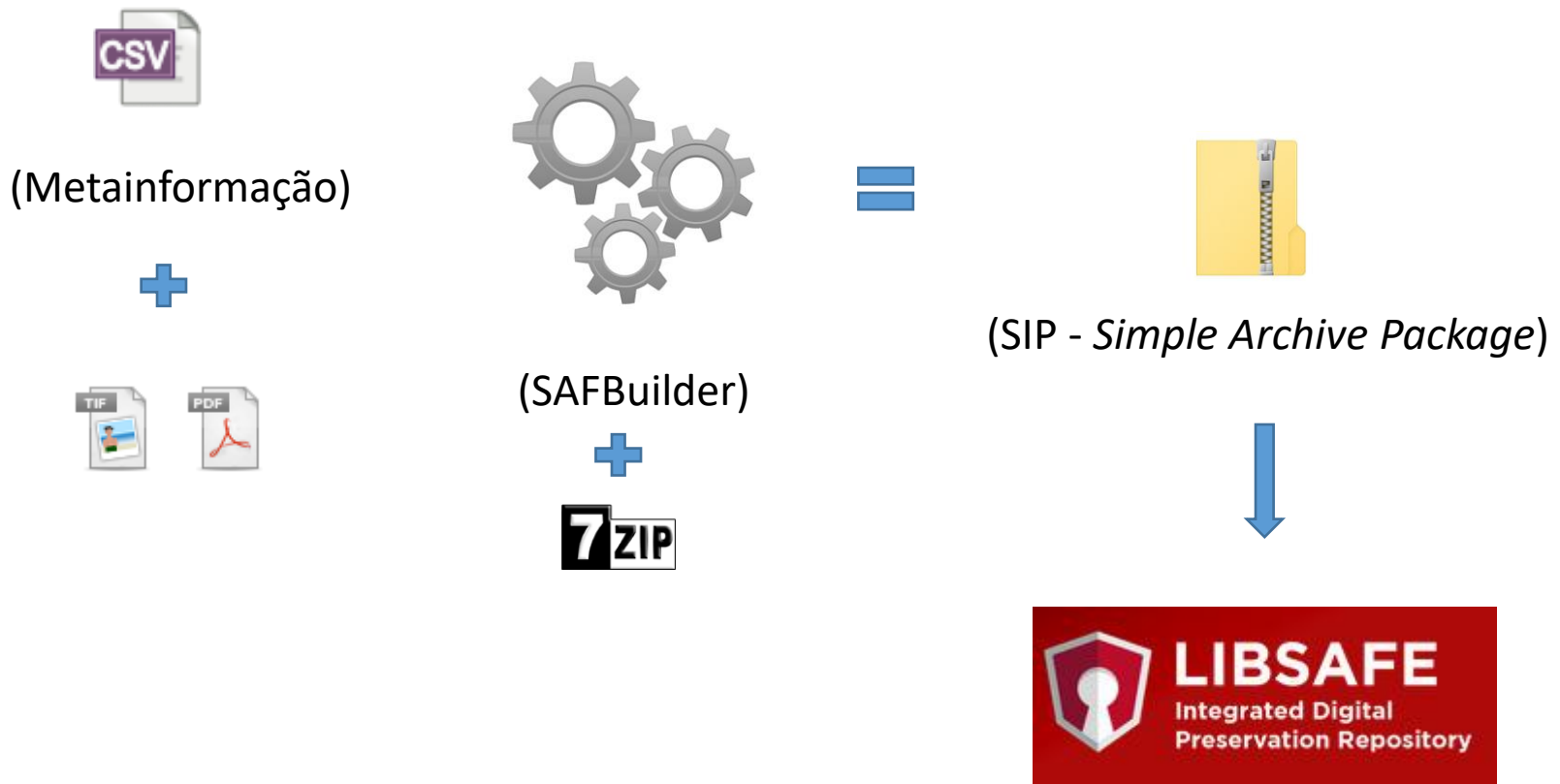
Formato do *Submission Information Package* (SIP)

Utilização do *Simple Archive Format* no processo de ingestão no LIBSAFE (<https://wiki.duraspace.org/display/DSPACE/Simple+Archive+Format+Packager>).

Utilização da ferramenta SAFBuilder (<https://github.com/DSpace-Labs/SAFBuilder>) para a criação do SIP.

Sistema de Preservação Digital: Plano de Preservação Digital - Criação do SIP

Ex. de criação do *Simple Archive Format*



Sistema de Preservação Digital: Estrutura do SIP

Collection_APP_50273

APP37

contents

dublin_core.xml

PT-UP-APP-37_0001.tif

PT-UP-APP-37_0002.tif

PT-UP-APP-37_C.pdf

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<dublin_core>
```

```
<dcvalue element="contributor" qualifier="uporto">Reitoria da Universidade do Porto</dcvalue>
```

```
<dcvalue element="contributor">Academia Politécnica do Porto, 1837-1911</dcvalue>
```

```
<dcvalue element="date" qualifier="issued">1876-01-08 - 1886-12-31</dcvalue>
```

```
<dcvalue element="description" qualifier="abstract">Série constituída por correspondência oficial
```

dirigida ao Diretor da Academia Politécnica do Porto, proveniente do Ministério do Reino, maioritariamente da Direção Geral de Instrução Pública e da Repartição de Contabilidade, no período compreendido entre 8 de janeiro de 1876 e 31 de dezembro de 1886. Contém 260 ofícios numerados, incluindo o programa para a representação dos estabelecimentos de ensino superior na Exposição Universal de Paris de 1878.</dcvalue>

```
<dcvalue element="description">Documento simples</dcvalue>
```

```
<dcvalue element="format" qualifier="medium">Livro; Papel; Manuscrito; Razoável</dcvalue>
```

```
<dcvalue element="format" qualifier="medium">Disco; Magnético; Bom</dcvalue>
```

```
<dcvalue element="format">1 livro</dcvalue>
```

```
<dcvalue element="identifier" qualifier="archeevo">210139</dcvalue>
```

```
<dcvalue element="identifier" qualifier="uodoc">APP37</dcvalue>
```

```
<dcvalue element="identifier" qualifier="uf">RUPTESSES11</dcvalue>
```

```
<dcvalue element="identifier" qualifier="uf">RUPTESSES11B</dcvalue>
```

```
<dcvalue element="identifier" qualifier="uf">APP_37</dcvalue>
```

```
<dcvalue element="identifier" qualifier="uri">http://hdl.handle.net/10405/50273</dcvalue>
```

```
<dcvalue element="language" qualifier="iso">por</dcvalue>
```

```
<dcvalue element="relation" qualifier="ispartofseries">Academia Politécnica do Porto :
```

Correspondência do Governo (1876-1886)</dcvalue>

```
<dcvalue element="rights">open access</dcvalue>
```

```
<dcvalue element="subject">Academia Politécnica do Porto</dcvalue>
```

```
<dcvalue element="subject">Correspondência recebida</dcvalue>
```

```
<dcvalue element="title">Academia Polythecnica do Porto : Correspondencia do Governo : 1876-1886</dcvalue>
```

```
<dcvalue element="type">Ofício</dcvalue>
```

```
</dublin_core>
```

Sistema de Preservação Digital: Documentação Preservada

- Arquivo da **Academia Politécnica do Porto, (1837-1911)**
 - 31 séries documentais
- Arquivo da **Reitoria da Universidade do Porto**
 - 3 séries documentais

Total de objectos digitais preservados: 9058

Total de ficheiros: 66010

Espaço ocupado: 1437.76 GB x 2 áreas de armazenamento (*datacenters*) geograficamente separadas

PRESERVAÇÃO DIGITAL - Recomendações

Software

DROID - <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

JHOVE - <https://jhove.openpreservation.org/>

Exiftool - <https://www.sno.phy.queensu.ca/~phil/exiftool/>

Exiftoolgui - <https://exiftool.org/gui/exiftoolgui516.zip> (download)

SafBuilder - <https://github.com/DSpace-Labs/SAFBuilder>

DPF Manager - <http://dpfmanager.org/>

PDFinfoGUI - <https://www.dcmembers.com/skwire/download/pdfinfogui/>

veraPDF - <https://verapdf.org/home/>

TweakPNG - <http://entropymine.com/jason/tweakpng/>

PDF24 - <https://pt.pdf24.org/>

PRESERVAÇÃO DIGITAL - Recomendações

Para leitura de 10 minutos

Ledoux, T. (2019, November 7). **JPEG got the blues**. Open Preservation Foundation. Disponível em: <https://openpreservation.org/blogs/jpeg-got-the-blues/> Acesso em 22 de Outubro de 2022.



5. Questões



OBRIGADO!

Augusto Ribeiro (UPdigital / GDI) - aecr@uporto.pt

João Pereira (UPdigital / GDI) - jmiguel@uporto.pt