

INSTITUTO POLITÉCNICO DO PORTO

ESCOLA SUPERIOR DE ESTUDOS INDUSTRIAIS E DE GESTÃO

Mestrado em Informação Empresarial

Patrícia Alexandra de Sousa Santos

Preservação digital nas médias e grandes empresas portuguesas: estudo de caso e recomendação de boas práticas

2015

Instituto Politécnico do Porto
Escola Superior de Estudos Industriais e de Gestão

Patrícia Alexandra de Sousa Santos

**Preservação digital nas médias e grandes empresas portuguesas:
estudo de caso e recomendação de boas práticas**

Dissertação de Mestrado

Mestrado em Informação Empresarial

Orientação: Professor Doutor Alberto Simões

Coorientação: Professor Doutor Ricardo Queirós

Vila do Conde, julho de 2015

Patrícia Alexandra de Sousa Santos

**Preservação digital nas médias e grandes empresas portuguesas:
estudo de caso e recomendação de boas práticas**

Dissertação de Mestrado
Mestrado em Informação Empresarial

Membros do Júri

Presidente

Professora Doutora Ana Lúcia Silva Terra

Escola Superior de Estudos Industriais e de Gestão – Instituto Politécnico do Porto

Professor Doutor Alberto Manuel Brandão Simões

Escola Superior de Tecnologia – Instituto Politécnico do Cávado e do Ave

Professora Doutora Idalete Maria da Silva Dias

Instituto de Letras e Ciências Humanas – Universidade do Minho

Vila do Conde, julho de 2015

À nossa doce, fiel e inigualável Diana. Foi uma dádiva ter-te na nossa vida durante oito anos e meio, apesar de teres partido prematuramente e sem que ninguém adivinhasse. Tiveste a melhor vida que podias ter tido e fizeste de nós melhores pessoas. Obrigada por teres existido. Descansa em paz nossa Princesa linda.

AGRADECIMENTOS

Terminada esta tão longa jornada é altura de agradecer a todos aqueles que contribuíram, de algum modo, para a sua concretização.

Ao Professor Alberto Simões, orientador, que sempre demonstrou ser uma excelente pessoa e um ótimo profissional. Agradeço o esforço e dedicação, durante estes últimos meses, para me apoiar e orientar sempre que precisei.

Ao Professor Ricardo Queirós, coorientador, agradeço o facto de ter aceiteado mais esta responsabilidade e de me ter ajudado sempre que foi possível.

A ambos agradeço o espírito crítico e nível de exigência: foram as vossas sugestões que permitiram criar, melhorar e finalizar este trabalho.

Às Professoras Milena Carvalho e Ana Terra que deram uma preciosa ajuda no arranque deste trabalho.

Às colegas Nídia e Patrícia que apenas conheci graças ao mestrado. Foi um gosto ter feito esta caminhada ao vosso lado e termos alcançado tantos feitos juntas.

Aos meus Pais que em todas as ocasiões me apoiaram e sempre estiveram presentes. Agradeço a educação que me deram e espero não defraudar as expectativas que têm em mim.

Ao Morito peço desculpa pelos meus momentos de distração e por todas as saídas que não foram feitas. Espero recompensar-te daqui em diante.

Um bem-haja a todos!

RESUMO ANALÍTICO

Este estudo tem como objetivo analisar o estado atual da preservação digital no tecido empresarial português, particularmente nas empresas de média e grande dimensão. Para tal, expôs-se o referencial teórico da preservação digital, necessário para uma melhor compreensão desta temática, nomeadamente através da apresentação de definições de diversos autores, de estratégias e boas práticas, e de algumas das iniciativas internacionais mais relevantes. De seguida apresentou-se a importância que a informação tem para uma organização, sobretudo aquela informação criada no âmbito laboral. Realizou-se, também, uma caracterização dos vários tipos de empresas, de acordo com diretivas europeias, e efetuou-se uma breve análise das empresas portuguesas.

Para atingir esse objetivo elaborou-se um inquérito, por questionário, fazendo-se uso de uma plataforma de questionários *online*. Este foi remetido a 1.358 médias e grandes empresas que atuam em Portugal, com o objetivo de compreender – entre outros – a importância dada, pelas empresas, ao material existente em formato digital, quais as tecnologias responsáveis pelo seu armazenamento e quais os cuidados realizados para assegurar o seu acesso por longos períodos de tempo.

Ao analisar os resultados, e pese embora a baixa taxa de resposta ao questionário, foi possível concluir que quase todas as empresas participantes já procuram substituir os documentos em papel por documentos em formato digital e que se preocupam em manter o acesso prolongado à informação. Não obstante, dois terços dos participantes conhecem o conceito de preservação digital e apenas 44% afirmam aplicar estratégias desse tipo na sua empresa. Apesar de cerca de 70% dos participantes reconhecerem a importância do uso de formatos abertos para a preservação digital, na realidade ainda se verifica uma utilização massiva de formatos proprietários ao nível dos documentos de texto e das folhas de cálculo.

Para além deste contributo analítico sobre o estado da preservação digital no tecido empresarial português, prepararam-se alguns documentos que visam auxiliar a elaboração de um plano de preservação digital e, conseqüentemente, permitir a sustentabilidade a longo prazo da informação digital.

Palavras-chave: Preservação digital; Médias e grandes empresas; Informação digital; Estratégias de preservação digital; Migração; Formatos abertos.

ABSTRACT

This study aims to analyse the current state of digital preservation in the portuguese enterprises, particularly in medium and large-sized enterprises. For this purpose, it was exposed the theoretical background of digital preservation, necessary to understanding this issue, namely by presenting definitions of various authors, strategies and best practices, and some of the most relevant international initiatives. Then, it was showed the importance of information for the organization, particularly the information that is created in the workplace. In addition, it was made a characterization of various types of enterprises, accordingly to european directives, and a brief analysis of the portuguese enterprises.

In order to achieve this goal it was elaborated a questionnaire survey, using an online survey platform. This questionnaire survey was sent to 1.358 medium and large-sized enterprises operating in Portugal, with the goal of understanding – among others – the importance given by companies to the material existing in digital format, which technologies are responsible for their storage and which are the practices used to ensure the access for long periods of time.

During the analysis of the results, and despite the low response rates to the questionnaire, it was possible to conclude that almost all participating companies seek to replace paper documents by digital documents and care to mantain long-term access to information. However, two thirds of participants know the digital preservation concept, but only 44% says that apply such strategies in their business. Although nearly 70% of the participants recognize the importance of the open formats use to the digital preservation, there is still a massive use of proprietary formats in text documents and spreadsheets.

Apart from this analytical contribute about the digital preservation state in portuguese enterprises, it was also prepared some documents to help the development of a digital preservation plan and, consequently, allow the long-term sustainability of digital information.

Keywords: Digital preservation; Medium and large-sized enterprises; Digital Information; Digital preservation strategies; Migration; Open formats.

SUMÁRIO

Lista de tabelas	10
Lista de gráficos	11
Lista de figuras	13
Glossário	15
Introdução	17
Metodologia de investigação	19
Estrutura do documento	28
Capítulo Um – Preservação digital	30
1 – Necessidade de um plano de preservação digital	30
2 – Definições de preservação digital	32
3 – Preservação tradicional e preservação digital	35
4 – Considerações acerca da preservação digital	36
5 – Plano de preservação digital	38
5.1 – Pré-requisitos	40
5.2 – Planeamento	41
5.3 – Implementação	43
6 – Ameaças	44
6.1 – Obsolescência tecnológica	45
6.1.1 – Obsolescência de formatos e <i>software</i>	45
6.1.2 – Obsolescência de suporte e <i>hardware</i>	48
6.2 – Danos físicos	49
6.3 – Ação humana	50
7 – Estratégias de preservação digital	51
7.1 – Preservação de tecnologia	54
7.2 – Refrescamento	55
7.3 – Emulação	56
7.4 – Migração	58
7.4.1 – Migração para suportes analógicos	60
7.4.2 – Normalização	60
7.4.3 – Atualização de versões de formato	62
7.4.4 – Conversão para formatos concorrentes	63

7.4.5 – Migração a-pedido	63
7.4.6 – Migração distribuída	65
7.5 – Encapsulamento	67
7.6 – Pedra de Rosetta digital.....	68
8 – Metainformação	68
8.1 – Metainformação de preservação	71
8.2 – Esquemas de metainformação	73
8.2.1 – Dicionário de Dados PREMIS.....	73
8.2.2 – Dublin Core.....	75
8.2.3 – METS.....	77
9 – Iniciativas	79
9.1 – Modelo de referência OAIS.....	79
9.1.1 – DSpace.....	86
9.1.2 – Fedora	89
9.1.3 – LOCKSS	92
9.1.4 – CRiB	94
9.2 – MoReq2010	98
Capítulo Dois – Médias e grandes empresas portuguesas	103
1 – O papel da informação no mundo empresarial	103
2 – Definição de empresas: micro, pequenas, médias e grandes.....	105
3 – Empresas em Portugal.....	108
Capítulo Três – Estudo de caso	111
1 – Metodologia para a seleção das empresas.....	111
1.1 – Critérios de seleção	111
1.2 – Processo de seleção das empresas	114
2 – Análise ao questionário.....	115
2.1 – Grupo I – Perfil da empresa.....	116
2.2 – Grupo II – Gestão documental.....	117
2.3 – Grupo III – Preservação digital	120
2.4 – Grupo IV – Interesse na receção dos resultados e comentários	122
3 – Análise de resultados.....	122
3.1 – Dimensão da empresa.....	123
3.2 – Tipo de empresa.....	124
3.3 – Localização geográfica e sede social da empresa	124

3.4 – Âmbito de atuação da empresa	126
3.5 – Substituição de documentos em papel por documentos digitais	127
3.6 – Utilização de <i>software</i> de gestão documental.....	129
3.7 – Recomendação de formatos.....	130
3.7.1 – Documentos de texto.....	131
3.7.2 – Folhas de cálculo.....	132
3.7.3 – Imagens.....	133
3.7.4 – <i>Emails</i>	133
3.8 – Dispositivos de armazenamento e cuidados na sua utilização	134
3.9 – Acesso continuado à informação	136
3.10 – Aplicação de estratégias de preservação digital.....	138
3.10.1 – Equipa multidisciplinar	139
3.10.2 – Responsável pela preservação digital	141
3.10.3 – Estratégias de preservação digital.....	142
3.10.4 – Tipos de documentos digitais alvo de preservação	143
3.10.5 – Motivação para a preservação digital	144
3.10.6 – Investimento realizado.....	145
3.11 – Desejo de implementação de um plano de preservação digital.....	145
3.12 – Formatos abertos.....	146
4 – Conclusões do estudo.....	148
Conclusão	152
Referências bibliográficas	159
Anexos	179
Anexo A – Boas práticas para o correto armazenamento e manutenção de objetos digitais.....	180
Anexo B – Processo de escolha da estratégia de preservação digital.....	184
Anexo C – Vantagens e desvantagens de estratégias de preservação digital.....	185
Anexo D – Critérios para a seleção de formatos.....	192
Anexo E – Lista de formatos adequados para preservação.....	196
Anexo F – Exemplos de esquemas de metainformação	200
Anexo G – Apresentação do questionário.....	206
Anexo H – Inquérito por questionário.....	207
Anexo I – Recomendação de ações que visam garantir a sustentabilidade da informação digital	217

Lista de tabelas

Tabela 1 – Caracterização de médias, pequenas ou micro empresas	107
---	-----

Lista de gráficos

Gráfico 1 – Proporção de empresas com 10 ou mais pessoas ao serviço que utilizam sistemas operativos de fonte aberta – 2007 a 2010.....	47
Gráfico 2 – Densidade de informação vs. Esperança de vida.....	48
Gráfico 3 – Formas de contacto das empresas selecionadas.....	114
Gráfico 4 – Número de registos do questionário.....	123
Gráfico 5 – Número de respostas por dimensão da empresa.....	123
Gráfico 6 – Número de respostas por dimensão e tipologia de empresas.....	124
Gráfico 7 – Número de empresas por distritos/regiões.....	125
Gráfico 8 – Localização da sede social das empresas privadas.....	125
Gráfico 9 – Número de empresas públicas por atividade económica.....	126
Gráfico 10 – Número de empresas privadas por atividade económica.....	127
Gráfico 11 – Substituição dos documentos em papel por digitais.....	128
Gráfico 12 – Número de empresas que procuram substituir os documentos em papel por digitais.....	128
Gráfico 13 – Empresas que utilizam <i>software</i> de gestão documental.....	129
Gráfico 14 – Programas de gestão documental usados pelas empresas.....	129
Gráfico 15 – Utilização de formatos recomendados.....	130
Gráfico 16 – Número de empresas que recomendam o uso de determinados formatos.....	131
Gráfico 17 – Formatos recomendados para documentos de texto.....	131
Gráfico 18 – Formatos recomendados para folhas de cálculo.....	132
Gráfico 19 – Formatos recomendados para imagens.....	133
Gráfico 20 – Formatos recomendados para <i>emails</i>	133
Gráfico 21 – Dispositivos de armazenamento utilizados nas empresas.....	135
Gráfico 22 – Cuidados com o <i>hardware</i> utilizado nas empresas.....	135
Gráfico 23 – Preocupação em manter o acesso contínuo à informação.....	136
Gráfico 24 – Conhecimento do conceito de preservação digital.....	137
Gráfico 25 – Número de empresas que têm conhecimento do conceito de preservação digital.....	138
Gráfico 26 – Aplicação de estratégias de preservação digital.....	138

Gráfico 27 – Número de empresas que aplicam estratégias de preservação digital	139
Gráfico 28 – Existência de equipa multidisciplinar para preparar e implementar o plano de preservação digital	140
Gráfico 29 – Número de empresas com uma equipa multidisciplinar responsável pela preservação digital	141
Gráfico 30 – Cargos dos responsáveis pela preservação digital.....	141
Gráfico 31 – Estratégias de preservação digital aplicadas nas empresas	143
Gráfico 32 – Documentos sujeitos a estratégias de preservação.....	143
Gráfico 33 – Razões que levaram à implementação do plano de preservação digital	144
Gráfico 34 – Investimento feito pelas empresas para implementar um plano de preservação digital	145
Gráfico 35 – Participantes que gostavam de ter um plano de preservação digital na sua organização	146
Gráfico 36 – Participantes que reconhecem a importância dos formatos abertos para a preservação digital	147
Gráfico 37 – Número de participantes que reconhecem a importância dos formatos abertos para a preservação digital	147

Lista de figuras

Figura 1 – Método quadripolar proposto por De Bruyne, Herman e De Schoutheete	21
Figura 2 – Atos e etapas do procedimento científico propostos por Quivy e Campenhoudt.....	22
Figura 3 – Proposta de classificação de Thibodeau adaptada por Ferreira	53
Figura 4 – Degradação do objeto digital ao longo de sucessivas ações de migração	64
Figura 5 – Migração a-pedido.....	64
Figura 6 – Migração distribuída baseada em serviços Web	66
Figura 7 – Metadados de preservação.....	71
Figura 8 – Interação entre entidades externas e o arquivo OAIS.....	81
Figura 9 – Modelo funcional do OAIS.....	84
Figura 10 – Funcionamento da plataforma DSpace	88
Figura 11 – Modelo de um objeto digital de acordo com o Fedora.....	90
Figura 12 – Especificação dos <i>datastreams</i> reservados	91
Figura 13 – Arquitetura do sistema CRiB	95
Figura 14 – Estrutura do MoReq2010	99
Figura 15 – Serviços do MoReq2010	100
Figura 16 – Estrutura de uma entidade no MoReq2010.....	102
Figura 17 – Principais indicadores económicos das empresas não financeiras de Portugal em 2012	109
Figura 18 – Número, dimensão e volume de negócios das grandes empresas (2008-2012)	110
Figura 19 – Número de empresas exportadas pelo SABI	114
Figura 20 – Fluxograma das questões do Grupo I	117
Figura 21 – Fluxograma das questões do Grupo II	119
Figura 22 – Fluxograma das questões do Grupo III	121
Figura 23 – Fluxograma das questões do Grupo IV	122
Figura 24 – Esquema para seleção da estratégia de preservação digital	184
Figura 25 – Excerto do esquema XML que suporta a implementação do Dicionário de dados PREMIS (1/2)	200

Figura 26 – Excerto do esquema XML que suporta a implementação do Dicionário de dados PREMIS (2/2)	201
Figura 27 – Exemplo de um registo correspondente ao elemento Objeto.....	202
Figura 28 – Exemplo de um registo correspondente ao elemento Evento	202
Figura 29 – Exemplo de um registo Dublin Core.....	203
Figura 30 – Exemplo de um registo METS (1/2)	204
Figura 31 – Exemplo de um registo METS (2/2)	205

Glossário

APDSI – Associação para a Promoção e Desenvolvimento da Sociedade de Informação

ASCII – *American Standard Code for Information Interchange*

b-on – Biblioteca do Conhecimento Online

CAE Rev.3 – Classificação Portuguesa das Atividades Económicas – Revisão 3

CAMiLEON – *Creative Archiving at Michigan and Leeds Emulating the Old on the New*

CCSDS – *Consultative Committee for Space Data Systems*

CD – *Compact Disc*

CE – Comissão Europeia

CEDARS – *CURL Exemplars in Digital ARchiveS*

CRiB – *Conversion and Recommendation of Digital Object Formats*

CSV – *Comma-Separated Values*

DCMI – *Dublin Core Metadata Initiative*

DGARQ – Direção-Geral de Arquivos

DOI – *Digital Object Identifier*

DVD – *Digital Versatile Disc*

ESEIG – Escola Superior de Estudos Industriais e de Gestão

Fedora – *Flexible Extensible Digital Object Repository Architecture*

GIF – *Graphics Interchange Format*

HTML – *HyperText Markup Language*

IDC – *International Data Corporation*

IEC – *International Electrotechnical Commission*

INE – Instituto Nacional de Estatística

IPP – Instituto Politécnico do Porto

ISO – *International Organization for Standardization*

JPEG – *Joint Photographic Experts Group*

LOCKSS – *Lots of Copies Keep Stuff Safe*

METS – *Metadata Encoding and Transmission Standards*

MIT – *Massachusetts Institute of Technology*

Moreq2010 – *Modular Requirements for Records Systems*

NCSA – *National Center for Supercomputing Applications*

NISO – *National Information Standards Organization*

OAI-PMH – *Open Archives Initiative Protocol for Metadata Harvesting*
OAIS – *Open Archival Information System*
OASIS – *Organization for the Advancement of Structured Information Standards*
OCLC – *Online Computer Library Center*
ODF – *OpenDocument Format*
ODS – *OpenDocument Spreadsheet*
ODT – *OpenDocument Text*
PDF – *Portable Document Format*
PIA – *Pacote de Informação de Arquivo*
PID – *Pacote de Informação de Disseminação*
PIS – *Pacote de Informação de Submissão*
PME – *Micro, Pequenas e Médias Empresas*
PNG – *Portable Network Graphics*
PREMIS – *PREservation Metadata: Implementation Strategies*
RAID – *Redundant Array of Independent Disks*
RCAAP – *Repositório Científico de Acesso Aberto de Portugal*
RDF – *Resource Description Framework*
RLG – *Research Libraries Group*
RODA – *Repositório de Objetos Digitais Autênticos*
SABI – *Sistema de Análises de Balanços Ibéricos*
SciELO – *Scientific Electronic Library Online*
SOA – *Service-Oriented Architecture*
TIC – *Tecnologias da Informação e da Comunicação*
TIFF – *Tagged Image File Format*
TOM – *Typed Objects Model*
TSV – *Tab-Separated Values*
UPF – *Formato Universal de Preservação*
USB – *Universal Serial Bus*
W3C – *World Wide Web Consortium*
XML – *eXtensible Markup Language*

Introdução

Desde os primórdios da civilização existe a necessidade da transmissão e registo de informação por parte do ser humano quer para comunicação imediata quer para posterior salvaguarda e memória. As pinturas rupestres, os papiros e, mais tarde, o papel foram algumas das formas utilizadas pela humanidade para expressar e preservar os seus saberes, crenças e atos com o objetivo de os transmitir às gerações vindouras.

Atualmente não estamos muito longe dessa intenção que os nossos antepassados também tinham de comunicar, partilhar e salvaguardar a informação. Mudaram-se as formas de comunicar e as ferramentas disponíveis para o fazer, mas os objetivos da partilha de informação e criação de conhecimento continuam presentes nas nossas vidas e de forma cada vez mais acentuada.

O aparecimento de ferramentas como o rádio, a televisão, o telemóvel, o computador e a Internet mudaram os paradigmas de comunicação existentes tornando mais fácil a relação, a interação e a partilha de informação entre indivíduos independentemente da sua localização geográfica, dando origem ao conceito de Aldeia Global. Atualmente vivemos numa era marcada pela consolidação das Tecnologias da Informação e da Comunicação (TIC) nas nossas vidas e onde é quase impensável viver, quer seja a nível pessoal como profissional, sem tirar proveito de todas as suas funcionalidades e capacidades. A produção de informação em formato digital é uma temática que tem vindo a ser alvo de diversos estudos em todo o Mundo. Como exemplo, podemos referir o estudo *El Diverso y Explosivo Universo Digital: Un pronóstico actualizado del crecimiento de la información mundial hasta el 2011*, realizado pela EMC¹ e IDC², onde se conclui que a quantidade de informação produzida sobre uma pessoa é maior do que a quantidade de informação produzida por si própria. A este fenómeno atribui-se a denominação de “sombra digital” (Figuerola *et al.*, 2009, p. 266).

A evolução dos dispositivos tecnológicos – como o computador –, de captura e de armazenamento permite, hoje em dia, criar um vasto número de documentos

¹ Disponível em [www: http://portugal.emc.com/index.htm](http://portugal.emc.com/index.htm). Trata-se de uma organização que fornece soluções de infraestruturas de informação.

² Disponível em [www: http://www.idc.pt/](http://www.idc.pt/). A *International Data Corporation* (IDC) é uma empresa consultora para a área das tecnologias da informação.

digitais de forma quase banal. Quanto maior a quantidade de dados digitais produzidos, maior é a necessidade de eficiência por parte dos sistemas informáticos responsáveis por os gerir e processar. Por conseguinte, também a probabilidade de se gerar conhecimento com base nessa informação sobe substancialmente (Galindo, 2012, p. 37). O conhecimento é uma peça fundamental no quotidiano de uma organização, pois permite que aqueles que o detêm ajam da forma mais adequada às suas necessidades, mas, principalmente, tendo em consideração as necessidades e objetivos da organização. O uso generalizado das TIC combinado com um mercado altamente competitivo faz com que as respostas dadas por uma organização aos estímulos exteriores tenham de ser, necessariamente, mais conscientes e céleres de forma a assegurar a sua sobrevivência. Tal só será possível se a organização assumir uma postura que privilegie o conhecimento dos seus colaboradores, de modo a fazer frente às constantes adversidades. Apesar da gestão do conhecimento numa organização possuir muitos outros aspetos que poderiam ser abordados, não é pretensão deste trabalho fazê-lo. Contudo, parece importante, desde logo, clarificar o trinómio dado-informação-conhecimento.

Dados podem ser considerados como o “conjunto de factos discretos e objectivos sobre os acontecimentos”, ou seja, descrevem as características de um dado evento ocorrido mas sem prestar qualquer juízo de valor sobre ele (Serrano e Fialho, 2005, p. 49). Por sua vez, a informação consiste em atribuir significado/contexto aos dados, através da sua organização, classificação e/ou processamento e graças ao estabelecimento de relações entre dados ou outras informações. Por último, surge o conhecimento que consiste na interpretação dos dados e da informação – tendo em conta o saber, experiências, valores e comportamentos daquele que os está a analisar – e que irá originar novas experiências e informações. O conhecimento reside na mente dos indivíduos mas pode ser transmitido a outras pessoas, nomeadamente através de documentos.

Para que as gerações vindouras sejam capazes de aceder à informação agora criada, é fulcral que a sociedade de hoje reconheça e se comprometa em salvaguardar a informação armazenada em âmbito digital e, para que isso se consiga, devem ser concebidas soluções e estratégias de preservação digital com recurso à tecnologia existente. No entanto, o grande desafio estará na manutenção do ambiente de criação

dessa informação digital: tanto *software* como *hardware* sofrem atualizações constantes e aquilo que está acessível hoje poderá já não estar amanhã³.

Assegurar a longevidade deste tipo de objetos não deve ser uma preocupação exclusiva das instituições responsáveis pela gestão e manutenção de arquivos e acervos digitais. Também as empresas capazes de criar grandes quantidades de informação devem assumir uma posição preventiva no que concerne à salvaguarda de documentos digitais ao longo do tempo. Contudo, apesar de existir bastante literatura e estudos relacionados com a preservação digital, nomeadamente em bibliotecas e arquivos, o mesmo não acontece para o mundo empresarial. Assim sendo, decidiu-se que esta dissertação pretende colmatar essa falha e, por isso, pretende-se realizar uma análise ao estado atual da preservação digital nas médias e grandes empresas portuguesas.

Metodologia de investigação

A presente dissertação enquadra-se na área da Ciência da Informação. Para Capurro e Hjørland (2007, p. 187), a Ciência da Informação é uma área de investigação com princípios multidisciplinares e que investiga as ações que tenham por base a informação, como, por exemplo, a sua origem, edição, recuperação, armazenamento, tratamento ou disseminação. Para estes autores, as TIC têm um papel preponderante em todas as perspetivas analisadas pela Ciência da Informação, pese embora o facto de estas tecnologias não serem exclusivas desta ciência e fazerem também parte da Ciência da Computação. Não se pode também esquecer que esta definição é sustentada pelo atual paradigma da Ciência da Informação – o paradigma informacional, científico e pós-custodial – cujo aparecimento está estritamente ligado à emergência das novas tecnologias. Este paradigma teve origem na década de 80 tendo substituído o paradigma custodial, patrimonial, historicista e tecnicista que surgiu depois da Revolução Francesa. O paradigma informacional, científico e pós-custodial caracteriza-se pelos constantes estudos em informação e pela importância da criação, disseminação e posterior acesso à informação (Ribeiro, 2011, p. 62-64).

³ Ver mais considerações acerca desta problemática em Obsolescência tecnológica.

Segundo Pinto (2009^b, p. 23), a Ciência da Informação, que teve a sua origem na Ciência Social, baseia-se num modelo científico-informacional descrito por três características principais: a definição da informação como objeto de estudo; o recurso à Teoria Sistémica como forma de interpretar o fenómeno *informação*; a utilização do método de investigação (o método quadripolar).

Como qualquer ciência, a Ciência da Informação possui um objeto de estudo, isto é, “uma marca apresentada por uma dada ciência e que a caracteriza essencialmente” (Freire e Silva, 2012, p. 169). O objeto de estudo pode alterar-se ao longo do tempo, dependendo então da época e do contexto em que está inserido. Bachelard (Cit. por Freire e Silva, 2012, p. 169) considera que um objeto de estudo científico possui duas características principais: a primeira é não ter origem natural, mas ser desenvolvido pela ciência; e a segunda é o objeto estar “em constante mudança a partir de sua relação com outros conceitos, objetos e métodos”. Pode-se considerar então a informação como sendo o objeto de estudo desta ciência, uma vez que ela analisa a evolução da informação desde a sua origem até à sua transformação em conhecimento.

A Teoria Sistémica teve a sua origem no austríaco Ludwig von Bertalanffy em meados da década de 1920. Esta teoria reivindica “que o todo é maior do que as partes e [...] que as partes devem ser estudadas, necessariamente, a partir da função que desempenham para a manutenção e sobrevivência do todo”, ou seja, defende uma perspetiva em que se pretende compreender um fenómeno na totalidade, e não apenas o comportamento de cada parte que o compõe (Araújo, 2009, p. 195). Recorrendo então à Teoria Sistémica, é possível definir o conceito de sistema de informação:

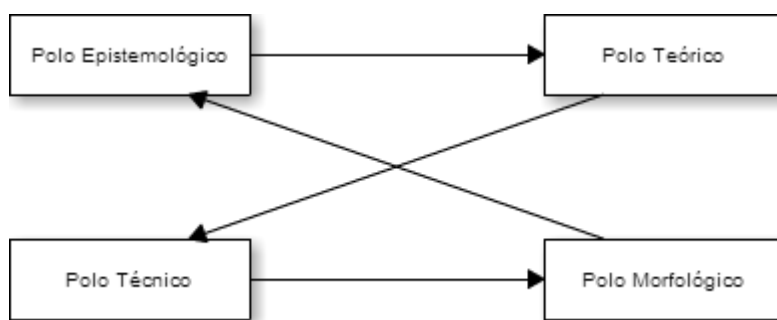
(um sistema de informação é um conjunto de) [...] diferentes tipos de informação registada ou não externamente ao sujeito [...], não importa qual o suporte (material e tecnológico), de acordo com uma estrutura prolongada pela acção na linha do tempo. A estrutura de um SI é um aspecto complexo porque ela é paradoxalmente autónoma e indissolúvel da informação propriamente dita: o sujeito de acção [...] que produz e recebe fluxo informacional é distinto deste, mas é essencial para que este exista.

(Silva, 2006, p. 163)

Ou seja, é importante perceber o contexto em que a informação circula (o todo) para que se possa compreender, de facto, o significado daquela informação (a parte).

Segundo Araújo (2009, p. 196), a Teoria Sistémica permitiu desenvolver estudos acerca dos sistemas de informação, nomeadamente a caracterização e descrição de processos associados ao seu bom funcionamento. Este autor acrescenta ainda que os sistemas de informação são idealizados tendo por base os processos de entrada (obtenção, aquisição e inserção da informação), de processamento (os objetos informacionais devem ser catalogados, classificados, indexados) e de saída (divulgação da informação). Para Mella (Cit. por Ribeiro, 2004, p. 4), a teoria sistémica está de tal forma desenvolvida que considera já apropriado falar-se em “pensamento sistémico” como “um modo de conceber a realidade, seja ela qual for”.

Para que se possa investigar uma ciência, é necessário a existência de métodos definidos que auxiliem a sua análise. No caso da Ciência da Informação, esta é analisada através do método quadripolar que foi desenvolvido por Paul De Bruyne, Jacques Herman e Marc De Schoutheete em 1974. Embora inicialmente o método quadripolar (Figura 1) não tenha sido concebido como auxiliar da Ciência da Informação, o facto de este método conseguir recuperar as etapas de identificação do problema, formulação de hipóteses, adoção de teorias, sua observação através de métodos quantitativos/qualitativos, respetiva análise de resultados e formalização de conclusões, fez com que se tornasse o método utilizado para a Ciência da Informação (Ribeiro, 2011, p. 67).

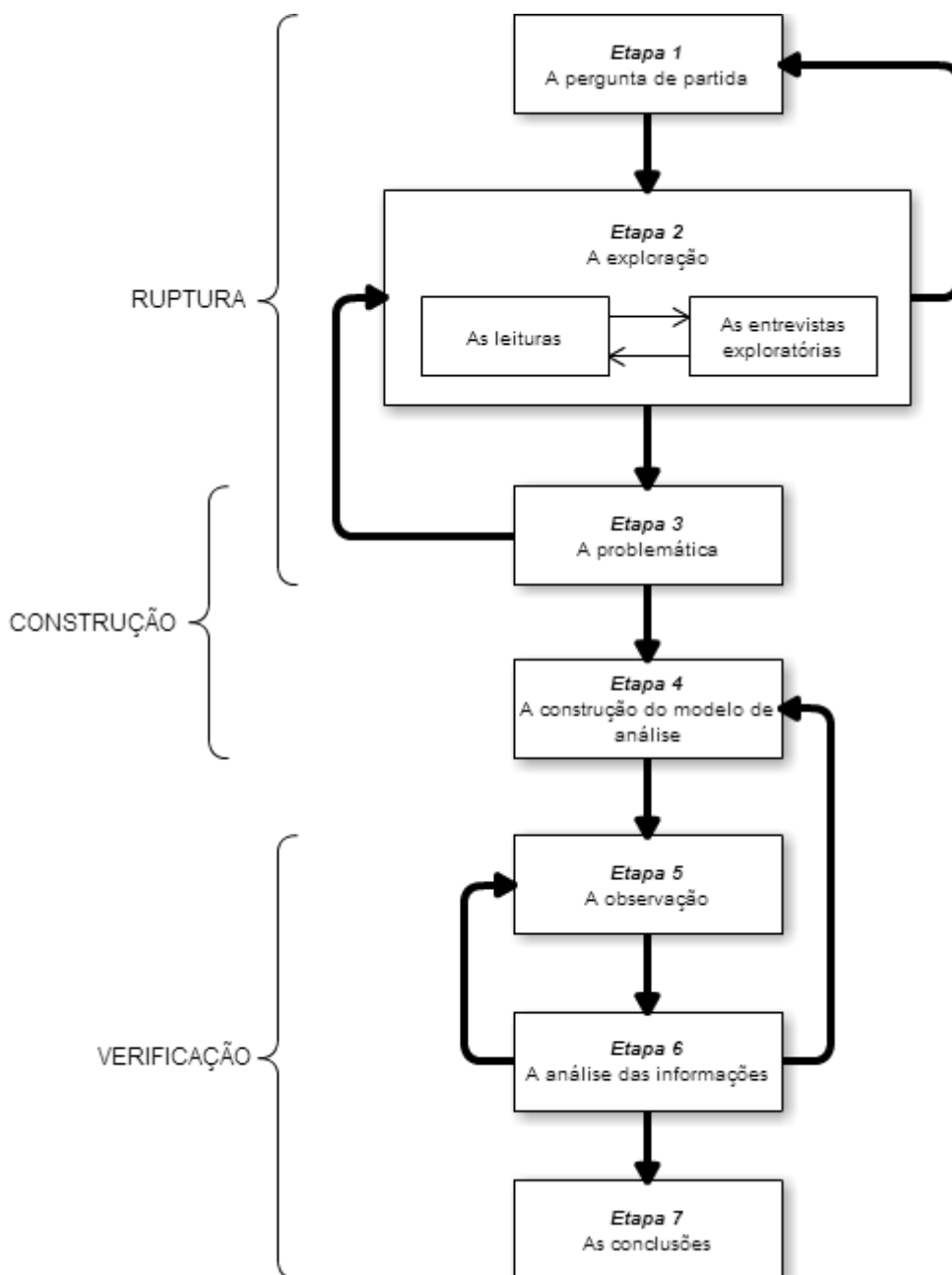


Fonte: Silva, 2006, p. 29

Figura 1 – Método quadripolar proposto por De Bruyne, Herman e De Schoutheete

Enquanto De Bruyne, Herman e De Schoutheete elaboraram uma metodologia geral com linhas orientadoras, Quivy e Campenhoudt descreveram uma perspectiva mais prática propondo procedimentos capazes de auxiliar a investigação

científica (Almeida, Silva e Guimarães, 2013, p. 5). Quivy e Campenhoudt propuseram um conjunto de 3 atos compostos, no total, por 7 etapas (Figura 2).



Fonte: Quivy e Campenhoudt, 2013, p. 27

Figura 2 – Atos e etapas do procedimento científico propostos por Quivy e Campenhoudt

O método quadripolar é composto, como o próprio nome indica, por quatro polos: epistemológico, teórico, técnico e morfológico. O facto de os quatro polos interagirem, como é possível observar na Figura 1, faz com que os paradigmas, teorias

e modelos estejam presentes no processo de pesquisa e na análise de resultados (Lessard-Hébert, Goyette e Boutin Cit. por Silva, 2006, p. 29). De seguida, podemos encontrar uma análise de cada um destes polos, recorrendo, sempre que necessário ao modelo proposto por Quivy e Campenhoudt.

Polo epistemológico

É neste polo que se constrói o objeto científico e se limita a problemática da investigação, ou seja, reformulam-se os paradigmas e critérios que conduzem a investigação (Silva e Ribeiro, 2009, p. 294).

No polo epistemológico pode-se encontrar a Ciência da Informação enquanto ciência orientadora de todo o trabalho sustentando-se no paradigma informacional, científico e pós-custodial.

Polo teórico

Quanto ao polo teórico, é nele que se evidencia a racionalidade do investigador, se formulam leis, hipóteses e conceitos (plano da descoberta) e a consequente confirmação/refutação do contexto teórico elaborado (plano da prova). Este polo pode, também, auxiliar na preparação e orientação da recolha de dados, bem como na interpretação de resultados, já que é nele que se definem as hipóteses da investigação (Silva e Ribeiro, 2009, p. 294).

Neste polo pode encontrar-se a teoria que caracteriza a Ciência da Informação – a Teoria Sistémica – e também a Gestão da Informação como sendo o campo de estudos onde se enquadra a preservação digital, embora esta temática tenha também implicações na Organização e Representação da Informação e no Comportamento Informacional (Pinto, 2009^b, p. 24).

É neste polo que se evidencia a formulação da pergunta de partida que deve ser clara, concisa, unívoca e exequível. Tendo esses aspetos em conta, redigiu-se a seguinte pergunta de partida para esta dissertação, que foi sendo reformulada sempre que necessário ao longo da investigação:

As médias e grandes empresas admitem a adoção de estratégias de preservação digital como forma de garantir a acessibilidade e sustentabilidade da informação a longo prazo?

É especialmente importante que se analise o objeto de estudo pretendido, com o objetivo de se adquirirem novos conceitos e de se solidificar o terreno que será explorado. Este estudo pode ser feito com recurso a revisão bibliográfica e a entrevistas exploratórias.

Com a revisão bibliográfica pretende-se fazer um levantamento de noções que se relacionem com a preservação digital, nomeadamente, a definição do conceito, ameaças, possíveis estratégias e iniciativas existentes.

Para que isso seja possível é necessário, primeiramente, proceder à recolha da informação. Neste trabalho optou-se por obter artigos em formato digital com recurso a bases de dados (como b-on⁴, RCAAP⁵, SciELO⁶, EBSCO⁷) e obter livros em bibliotecas, nomeadamente a biblioteca da Escola Superior de Estudos Industriais e de Gestão (ESEIG). A recuperação de artigos, nas bases de dados digitais referidas anteriormente, foi feita com recurso a expressões portuguesas e inglesas (como “preservação digital” e “digital preservation”), de modo a recolher obras nacionais e internacionais.

As leituras preparatórias são de especial importância, uma vez que permitem identificar o tipo de investigação feita até à data. Assim, foi possível observar uma vasta bibliografia relativa à temática da preservação digital, particularmente a descrição de realidades de arquivos, bibliotecas e de instituições de ensino superior. Destacou-se também uma preocupação na acessibilidade da investigação científica resultado da publicação de revistas da especialidade. Concluiu-se, também, que já existia trabalho realizado sobre esta área, chamando particular atenção os trabalhos nacionais *Recomendações para a produção de Planos de Preservação Digital* e *Introdução à Preservação Digital – Conceitos, estratégias e actuais consensos*. Já a

⁴ Biblioteca do Conhecimento Online. Disponível em [www: http://www.b-on.pt/](http://www.b-on.pt/).

⁵ Repositório Científico de Acesso Aberto de Portugal. Disponível em [www: http://www.rcaap.pt/](http://www.rcaap.pt/).

⁶ *Scientific Electronic Library Online*. Disponível em [www: http://www.scielo.org/php/index.php](http://www.scielo.org/php/index.php).

⁷ Disponível em [www: http://search.ebscohost.com](http://search.ebscohost.com).

nível internacional destacam-se os estudos *Preservation Management of Digital Materials: The Handbook e Guidelines for the Preservation of Digital Heritage*.

Graças às leituras preparatórias foi possível realizar entrevistas exploratórias com vista à descoberta de novos aspetos a ter em consideração e que não tinham sido refletidos até àquela altura. Para que uma entrevista exploratória traga os resultados desejados, é necessário que seja conduzida de forma aberta e flexível, permitindo que o entrevistado divague sobre a temática em causa. Podem ser feitas várias entrevistas exploratórias, particularmente com indivíduos com conhecimento de causa do assunto, sejam eles docentes, investigadores ou peritos.

A primeira entrevista exploratória foi ao Professor Doutor Alberto Simões, orientador desta dissertação. Uma das conclusões possíveis de retirar é que se põe em causa a capacidade das empresas portuguesas aplicarem estratégias efetivas de preservação. Também se chama particular atenção para a utilização de plataformas digitais colaborativas, uma vez que se não forem aplicados os mecanismos de segurança adequados, podem por em causa a privacidade e autenticidade dos documentos armazenados.

A segunda entrevista exploratória realizada foi feita ao Professor Rafael António, tendo sido dada particular atenção ao MoReq2010 (*Modular Requirements for Records Systems*): uma especificação europeia orientada para o desenvolvimento de sistemas de gestão de documentos de arquivo capazes de gerir documentos digitais ao longo de todo o seu ciclo de vida.

Com a realização da revisão bibliográfica e com a abertura de novas perspetivas trazidas pelas entrevistas exploratórias foi possível elaborar as hipóteses que pretendem ser possíveis respostas à pergunta de partida e que irão suportar a elaboração da dissertação.

As hipóteses podem ser construídas através de dois processos: o indutivo e o dedutivo. O método hipotético-indutivo produz conceitos operatórios através da experiência, enquanto o método hipotético-dedutivo, construído por raciocínio abstrato, produz conceitos sistémicos, para os quais se procura correspondentes na realidade (Quivy e Campenhoudt, 2013, p. 144, 150).

Para este trabalho foram redigidas duas hipóteses, que pretendem dar resposta à pergunta de partida anteriormente formulada, com base no modelo

hipotético-dedutivo, uma vez que se parte de um conceito geral – a preservação digital – para a comprovação do seu estado nas médias e grandes empresas portuguesas (conceito particular). São elas:

A estratégia de preservação digital mais utilizada pelas médias e grandes empresas portuguesas é a migração.

As médias e grandes empresas portuguesas reconhecem a importância dos formatos abertos para a preservação digital.

Polo técnico

Relativamente ao polo técnico, ele “estabelece a relação entre a construção do objecto científico e o mundo dos acontecimentos” (De Bruyne *et al.* Cit. por Léssard-Hébert, Goyette e Boutin, 2005, p. 5). No polo técnico entra-se em contacto, por via experimental, com a realidade que se pretende investigar, procurando, assim, confirmar ou refutar as hipóteses, teorias e conceitos que foram sendo elaborados (Silva, 2006, p. 155).

Assim sendo, é neste polo que se procura entrar em contacto com a realidade que se pretende estudar: as médias e grandes empresas de Portugal. Com a observação pretende-se compreender se essas empresas já estão sensibilizadas para a temática da preservação digital, quais as estratégias mais utilizadas e quais os tipos de documentos que as organizações mais interesse têm em preservar.

Para conseguir atingir esse propósito é necessário escolher o método de observação mais adequado e proceder à seleção dos dados das empresas.

Relativamente à observação, existem duas categorias: a observação direta e a indireta. Na observação direta o investigador efetua a recolha dos dados para análise, sem contactar com os sujeitos envolvidos, não tendo os últimos qualquer tipo de intervenção na produção da informação. A observação indireta já implica contacto entre o investigador e o sujeito analisado, onde o sujeito responde a uma série de questões previamente estruturadas intervindo, assim, na produção da informação (Quivy e Campenhoudt, 2013, p. 164).

Uma vez que na presente dissertação se pretende analisar as médias e grandes empresas de Portugal então o tipo de observação mais adequada é a indireta.

Não obstante, na observação indireta existem dois tipos de instrumentos de recolha de dados: o questionário ou a entrevista.

O inquérito por questionário consiste em colocar questões (que devem ser claras e unívocas), já previamente elaboradas, a um sujeito (o inquirido) acerca do objeto de estudo do investigador e que este considere pertinente para o trabalho que tem vindo a realizar, nomeadamente a verificação das hipóteses que se pretendem ver confirmadas/refutadas. Este tipo de inquérito permite recolher informação de um grande número de indivíduos e permite, também, uma fácil análise estatística das respostas recebidas. No entanto também existem contras, uma vez que a resposta aos questionários depende muito do conhecimento, vontade e capacidade da pessoa inquirida (Quivy e Campenhoudt, 2013, p. 188-190).

No inquérito por entrevista, o contacto entre investigador e entrevistado é privilegiado. O investigador lança perguntas-guias (abertas), com vista ao teste das hipóteses de trabalho elaboradas, às quais se pretende que o entrevistado divague. O investigador deve deixar a entrevista fluir, orientando-a, sempre que necessário, para os objetivos que pretende atingir. A entrevista permite recolher um vasto número de informações que posteriormente terão de ser analisadas, mas o facto de essas informações não estarem disponíveis de imediato implica grande capacidade analítica por parte do investigador (Quivy e Campenhoudt, 2013, p. 191-195).

Tal como já foi referido anteriormente, dado o grande número de médias e grandes empresas (amostra) que se pretende analisar e a sua vasta extensão geográfica, a técnica de recolha de dados mais adequada é o inquérito por questionário.

No que concerne à recolha das informações das empresas portuguesas, será utilizada a base de dados SABI⁸ disponível na ESEIG. Esta base de dados reúne informações financeiras de empresas portuguesas bem como espanholas. Nos critérios de pesquisa devem-se apenas selecionar as empresas de Portugal e com estado ativo, e deve-se também recuperar informação acerca do último número de funcionários, volume de negócios e balanço total. Estas informações são fulcrais para posteriormente distinguir micros, pequenas, médias e grandes empresas.

⁸ Sistema de Análises de Balanços Ibéricos. Disponível em [www: http://www.bvdinfo.com/en-gb/our-products/company-information/national-products/sabi](http://www.bvdinfo.com/en-gb/our-products/company-information/national-products/sabi).

Polo morfológico

Por fim, surge o polo morfológico onde se procura estruturar o objeto científico, nomeadamente, através da organização e apresentação dos resultados obtidos e da elaboração de uma explicação de causalidade que permita relacionar esses resultados com a fundamentação teórica realizada anteriormente (De Bruyne *et al.* Cit. por Léssard-Hébert, Goyette e Boutin, 2005, p. 25-26). A análise dos dados pode ser feita através de métodos quantitativos ou qualitativos.

No polo morfológico são então apresentadas as conclusões do estudo realizado e de toda a informação recolhida ao longo da elaboração da dissertação (Pinto, 2009^b, p. 23-24). Para que seja possível apresentar conclusões acerca dos dados recolhidos nos inquéritos por questionário, é necessário adotar o método quantitativo que consiste numa análise estatística dos dados, com recurso a aplicações informáticas que permitam o tratamento de grandes volumes de informação (Quivy e Campenhoudt, 2013, p. 222-225). Não obstante, pretende-se também redigir a recomendação de um conjunto de boas práticas que devem ser aplicadas para garantir a longevidade da informação armazenada em formatos digitais.

Estrutura do documento

Esta dissertação divide-se em três capítulos principais.

No Capítulo Um apresenta-se a revisão da literatura acerca da preservação digital: abordam-se e comparam-se definições de diversos autores; estabelecem-se diferenças e semelhanças entre a preservação tradicional e a digital; e discutem-se as características que um plano de preservação digital deve possuir para que a sua implementação seja bem-sucedida. Distinguem-se, também, as várias ameaças a que estão sujeitos os objetos armazenados em formato digital e referem-se algumas das estratégias e boas práticas mais referenciadas na literatura. É, ainda, evidenciada a importância da metainformação para a preservação digital e referem-se alguns dos esquemas mais conhecidos. Por fim, descrevem-se duas especificações: o modelo de referência OAIS (*Open Archival Information System*) e o MoReq2010 (*Modular Requirements for Records Systems*).

O Capítulo Dois é dedicado às empresas. Inicialmente, procura-se evidenciar a importância da informação e do conhecimento no seio de uma organização, dando particular destaque àqueles que podem ser registados em formato digital e que refletem as atividades laborais efetuadas. De seguida, e à luz das recomendações europeias, distinguem-se os diversos tipos de empresas que existem, de modo a que seja possível definir os critérios necessários para classificar uma organização como média ou grande empresa, e que são essenciais para a concretização do estudo realizado. Por último, é feita uma breve síntese do estado das empresas em Portugal, de forma a compreender-se a sua importância para o desenvolvimento do país.

No Capítulo Três é apresentado o estudo de caso e a forma como foi realizada a seleção das empresas que foram convidadas a participar neste estudo. Este capítulo inclui uma breve análise ao questionário formulado e apresenta diagramas que ilustram a sequência de perguntas. De seguida, analisam-se os resultados e tecem-se as conclusões do estudo onde se procura verificar/refutar as hipóteses consideradas para este trabalho.

Para além dos três capítulos mencionados, e desta Introdução, o documento termina com a Conclusão que reflete os pontos essenciais abordados ao longo deste projeto e sugerem-se algumas propostas para trabalho futuro.

Não menos importante, este trabalho inclui vários Anexos, que, para além de incluírem informação complementar à dissertação, como seja a totalidade do questionário enviado às empresas, inclui também um conjunto de resumos que refletem os trabalhos levados a cabo pela comunidade de interesse na área da preservação digital, como sejam uma síntese das boas práticas para o armazenamento de objetos digitais; os prós e contras de algumas das estratégias habitualmente usadas para a preservação digital; os critérios para a seleção de formatos para o armazenamento dos objetos digitais e, também, uma lista de formatos adequados para a preservação. Finalmente, foi também elaborado um documento que visa reunir um conjunto de recomendações para auxiliar a preparação de modelos de preservação digital tendo em conta a necessidade de acessibilidade, a longo prazo da informação em formato digital.

Capítulo Um – Preservação digital

Neste capítulo introduz-se a temática da preservação digital, graças à revisão da literatura consultada.

Inicialmente, faz-se uma demonstração da necessidade da aplicação de ações de preservação nos objetos existentes em formato digital. Apresentam-se definições de diversos autores e procura-se evidenciar as semelhanças e diferenças entre elas.

Uma vez que os documentos digitais, tal como os armazenados em formato papel, são afetados com o decorrer do tempo, faz-se uma breve comparação entre as ameaças e os objetivos da preservação de informação em suportes digitais e em papel.

De seguida avança-se para a apresentação de considerações sobre o plano de preservação digital, essencial para a implementação de ações e boas práticas que visem garantir a longevidade da informação digital. São ainda nomeadas algumas das ameaças a que estão sujeitas, como é o caso da obsolescência tecnológica.

Posteriormente, são apresentados e explanados dois fatores que contribuem para o sucesso da preservação digital numa organização: as estratégias e a metainformação.

Por último, apresentam-se duas das iniciativas internacionais com mais relevo para a preservação digital: o modelo de referência OAIS (*Open Archival Information System*) e o Moreq2010 (*Modular Requirements for Records Systems*).

1 – Necessidade de um plano de preservação digital

Hoje em dia, grande parte da informação criada tem a sua origem em ambiente digital, não chegando sequer a ser transferida para ambiente analógico. Contudo, muita da informação existente em papel passa a existir, também, digitalmente, seja através de uma estratégia de preservação ou para que o seu acesso seja facilitado. As Tecnologias da Informação e da Comunicação (TIC) vieram modificar a forma como o ser humano lida com a informação e os documentos. É com extrema facilidade que se consegue criar, replicar, editar, armazenar ou partilhar objetos digitais através destes novos dispositivos.

Convém, antes de mais, clarificar o conceito de objetos digitais: tratam-se de objetos passíveis de serem representados através de código binário. O que equivale dizer que se tratam quer de informações nascidas através de dispositivos digitais (nado-digital), bem como informações digitalizadas através de suportes analógicos. Podem representar qualquer tipo de informação e em qualquer tipo de formato (Ferreira, 2006, p. 21; Pinto, 2009^a, p. 326). Um objeto digital possui características referentes a três dimensões: física, lógica e conceptual. O objeto físico consiste num conjunto de sinais inscritos num suporte físico. O *hardware* transforma esse conjunto de sinais num conjunto de dados – organizados de acordo com as especificações do formato – capazes de ser interpretados e processados pelo *software* (objeto lógico). Por sua vez, o *software* prepara o objeto lógico para ser apresentado ao utilizador através dos periféricos de saída. O objeto conceptual (ou intelectual) consiste na imagem reconhecida e percebida pelo utilizador (Ferreira, 2006, p. 22-23; Thibodeau, 2002, p. 6).

Porém, a utilização em massa das TIC não traz apenas vantagens. É sabido que toda a informação criada no seio de uma organização está dependente das plataformas tecnológicas existentes (*software*, *hardware* e suportes de armazenamento) no momento da sua criação, isto é, para que um documento possa ser recuperado e utilizado no futuro, devem existir dispositivos e programas compatíveis que permitam a sua leitura e processamento⁹. O facto de a tecnologia estar em constante evolução e mutação leva a que se torne obsoleta rapidamente e torna-se então uma ameaça ao acesso a médio e longo prazo de documentos digitais.

Urge então a necessidade de preservar a informação armazenada em dispositivos digitais para que se possa assegurar o seu acesso (idealmente, por um período de tempo prolongado), procurando minimizar os efeitos que a obsolescência tecnológica, as mudanças no ambiente legal do país e as mudanças de gestão da organização são capazes de provocar. Márdero Arellano (2004, p. 16) considera que o desafio da preservação digital não se debruça tanto sobre a componente técnica, mas sim ao nível social e institucional, uma vez que se trata de uma abordagem que está dependente de organizações que podem sofrer diversas mudanças ao longo do

⁹ Como exemplo, podemos encontrar as disquetes de 3,5 polegadas. Atualmente, os computadores já não são vendidos com esse tipo de suportes integrados, no entanto (ainda) é possível adquiri-los.

tempo, desde ao nível da estrutura hierárquica, missão e objetivos ou recursos financeiros.

Não obstante, existem ainda outros desafios que devem ser ultrapassados, nomeadamente uma mudança na cultura das organizações, de modo a que todos os envolvidos compreendam os fluxos informacionais existentes e se implementem os mecanismos necessários para uma eficaz gestão da informação e da documentação que é criada no dia-a-dia com o objetivo de auxiliar na realização das tarefas a que essa organização se propõe. Nesses mecanismos – que muitas das vezes acarretam elevados custos financeiros – podemos encontrar o desenvolvimento ou a adoção de um sistema informático que atenda às características da preservação digital e também a implementação de políticas efetivas de manutenção e acesso a longo prazo à informação que visem assegurar a sua autenticidade, fidedignidade e integridade durante todo o seu ciclo de vida. Ou seja, pretende-se manter, a longo termo, a informação completa, utilizável e inalterável, assegurando a preservação dos detalhes da sua criação (como o propósito, produtor ou remetente) e garantindo que se trata de uma representação fiável das atividades desenvolvidas.

2 – Definições de preservação digital

São inúmeros os autores que têm vindo a abordar a temática da preservação digital e toda a sua envolvente. Vejamos as seguintes definições:

A preservação digital é usada para descrever os processos envolvidos na manutenção da informação e outros tipos de património existente em formato digital.

(National Library of Australia, 2003, p. 20, tradução nossa)

A preservação digital consiste nos processos destinados a assegurar o acesso continuado aos materiais digitais. Para isso, devem-se encontrar formas de rerepresentar o que foi originalmente apresentado aos utilizadores através de uma combinação de ferramentas de *software* e *hardware* que atuam sob os dados.

(National Library of Australia, 2003, p. 34, tradução nossa)

A preservação digital pode ser vista como o conjunto dos processos que visam assegurar a continuidade dos objetos digitais durante o período de tempo em que eles se mostrem necessários.

A mais significativa das ameaças para a continuidade digital prende-se com a perda dos meios de acesso. Os materiais digitais não podem ser preservados se os meios de acesso se tenham perdido e o acesso se tenha tornado

impossível. O objetivo de preservar os materiais digitais é manter a sua acessibilidade: a capacidade de aceder à essência da mensagem ou propósito de forma autêntica.

(National Library of Australia, 2003, p. 34, tradução nossa)

A preservação digital consiste na capacidade de garantir que a informação digital permanece acessível e com qualidade de autenticidade suficiente para que possa ser interpretada no futuro recorrendo a uma plataforma tecnológica diferente da utilizada no momento da sua criação.

(Ferreira, 2006, p. 20)

As definições propostas pela *National Library of Australia* e por Ferreira seguem a linha de pensamento de que a preservação digital se trata de um conjunto de ações que visam manter um acesso contínuo aos objetos existentes em plataformas digitais, enquanto estes tenham utilidade para um indivíduo ou organização. Ações essas que têm como objetivo resolver o problema da evolução tecnológica, permitindo que através das plataformas atuais se interpretem, apresentem e mantenham autênticos os objetos digitais criados em plataformas mais antigas.

Pode definir-se como o conjunto de ações relativas à preservação de documentos em formato digital. Assim, se a entendermos como um conjunto de atividades, entre as quais se inclui a conservação, e que se destinam a que um objeto perdure o maior tempo possível no seu estado original, a preservação digital procura levar a cabo estas tarefas sobre a informação em formato digital, preocupando-se não só com a manutenção do objeto, mas também (e principalmente) com o seu conteúdo informativo.

(Romero, 2006, p. 128, tradução nossa)

Romero tem também uma visão semelhante às anteriores, no entanto acrescenta que, para além de se manter um objeto ao longo do tempo, a prioridade deverá ser o conteúdo da informação nele existente. Isto é, não adianta realizar ações de preservação num objeto se estas originarem alterações indesejadas pondo em causa a sua autenticidade e integridade.

Refere-se à série de atividades de gestão necessárias para garantir o acesso contínuo a materiais digitais durante o tempo que forem necessários. A preservação digital [...] refere-se a todas as ações necessárias para manter o acesso aos materiais digitais para além dos limites das falhas do *hardware* ou alterações tecnológicas. Esses materiais podem ser registos criados durante o dia-a-dia de uma organização; materiais nado-digitais criados para um dado propósito [...]; ou produtos de projetos de digitalização.

(Beagrie e Jones, 2008, p. 24, tradução nossa)

Por sua vez, Beagrie e Jones enriquecem as definições anteriores fazendo referência às possíveis causas de perda de acesso a objetos digitais, nomeadamente eventuais falhas dos dispositivos de armazenamento e as consequências da evolução tecnológica. Não obstante, acrescentam ainda três possíveis origens dos objetos: a criação de registos no âmbito de funções organizacionais, os objetos nado-digitais criados para um determinado fim ou, ainda, os criados através do processo de digitalização.

Como um processo contínuo, o conceito de preservação digital inclui planeamento constante, alocação de recursos e aplicação de métodos e técnicas necessários para garantir que os recursos de informação digital de valor contínuo permaneçam acessíveis e utilizáveis para as gerações atuais e futuras.

(Anbu e Chibambo, 2009, p. 47, tradução nossa)

A preservação digital, enquanto um conjunto de atividades voltadas para garantir o acesso aos conteúdos digitais por longo prazo, é, ao mesmo tempo, um desafio técnico e organizacional que se desenrola permanentemente no tempo e no espaço; seus objetivos exigem processos que portem uma intencionalidade contínua, dado que os objetos digitais não sobrevivem inercialmente, como sobrevivem as plaquetas de argila de cinco mil anos encontradas casualmente no deserto. Não existe absolutamente essa possibilidade para os objetos digitais.

(Sayão, 2010, p. 7)

A Preservação é assim, intrínseca à função de Gestão, seja a nível institucional, seja a nível intermédio [...], devendo ser pensada no longo prazo e em termos de políticas, planos e programas, recursos e estrutura orgânica/funcional que os suporte, tendo, conseqüentemente, implicações quer na fixação da Missão da Organização, quer nos objetivos (estratégicos e operacionais), quer nas metas fixadas, quer, ainda, nas acções/actividades e projectos planeados para os efectivar. A Preservação projecta-se ao nível da gestão da organização, da gestão do serviço de informação e da gestão do sistema de informação organizacional, afirmando-se como uma variável indiscutível da Gestão da Informação.

(Pinto, 2009^b, p. 34)

[...] não devemos considerar a preservação digital como um processo isolado, mas como um componente de um conjunto de serviços, políticas e especialistas que constituem o contexto do ciclo de vida da informação digital.

(Boeres e Cunha, 2012, p. 107)

Por último, tanto Sayão como Pinto introduzem nas suas definições o fator organizacional. Contudo, Pinto dá uma visão mais específica das implicações que a preservação digital traz a uma empresa. Tendo em conta o âmbito desta dissertação, essa é uma das definições de maior relevância.

A implementação de um plano de preservação digital envolve a criação de uma equipa multidisciplinar composta por colaboradores que terão como funções planejar e definir procedimentos de modo a manter a informação disponível ao longo do seu ciclo de vida. A concretização de tais procedimentos poderá (e deverá) afetar o dia-a-dia da organização, uma vez que são necessárias atitudes renovadas para que se possam obter melhores resultados a longo prazo. O plano requer também a atribuição de fundos monetários – não apenas por uma vez, mas repetidamente ao longo do tempo – de modo a fazer face aos eventuais investimentos necessários, quer ao nível da infraestrutura tecnológica como para a contratação de pessoal especializado. Por conseguinte, torna-se fulcral a inserção das ações de preservação digital na gestão estratégica da organização para que, independentemente das mudanças administrativas ocorridas, os recursos para a preservação digital estejam salvaguardados.

Através da revisão da literatura, é possível concluir que as definições de preservação digital seguem duas interpretações: aquelas que dão ênfase às ações que devem ser realizadas para combater a obsolescência tecnológica ou falhas técnicas, e aquelas que dão mais importância às mudanças a nível organizacional necessárias para albergar todos os requisitos da preservação digital. Porém, todas as definições convergem para a necessidade de se planearem ações adequadas que permitam que a informação armazenada nos objetos digitais esteja acessível por um período de tempo prolongado (enquanto for necessária), e mantenha as características de autenticidade e integridade.

3 – Preservação tradicional e preservação digital

Apesar dos suportes de armazenamento da informação dita normal (em suporte papel) e da informação digital serem bastante diferentes, ambos sofrem vários tipos de ameaças, sendo algumas delas semelhantes. Os acervos tradicionais são afetados pela degradação do papel, enquanto a informação armazenada no meio digital é afetada quer pela obsolescência dos dispositivos de armazenamento quer dos próprios formatos usados. Quer nos acervos tradicionais como nos digitais devem existir cuidados com os níveis de temperatura, humidade e poluição no ambiente onde se encontram, e também com a prevenção de catástrofes (incêndios, inundações,

terramotos). O uso excessivo ou incorreto destes dois tipos de acervos podem também ser lesivos para a informação existente (Sant'Anna, 2001, p. 129).

Porém, ao contrário do que acontece com a informação registada nos suportes tradicionais, no mundo digital é possível a decomposição entre o suporte e o conteúdo intelectual nele registado. Tal situação facilita a partilha de informação sem que haja perda de dados, contudo também torna a informação mais suscetível de ser modificada de forma inapropriada, quebrando assim as suas características de integridade e autenticidade.

O objetivo da preservação da informação em suporte tradicional ou digital é de assegurar que esta seja recuperada de forma autêntica independentemente das mudanças que possam ocorrer nos seus suportes ou na gestão da organização que a detém (Grácio e Fadel, 2009, p. 276). Se na preservação tradicional se procura apenas conservar o suporte físico, na digital também se procura salvaguardar a informação existente garantindo que o seu acesso será feito de forma autêntica, independentemente do suporte. Para que isso seja possível devem ser aplicadas estratégias de preservação e efetuar uma correta documentação de todo o processo.

4 – Considerações acerca da preservação digital

A preservação digital procura minimizar, e idealmente evitar, os danos causados pela evolução tecnológica, pela fragilidade dos suportes físicos e do meio digital com o objetivo de assegurar que essa informação que se pretende ver preservada mantenha a autenticidade, integridade e fiabilidade. Dessa forma, as organizações conseguirão garantir a salvaguarda da sua memória.

Márdero Arellano (2004, p. 17) faz distinção entre dois tipos de preservação: preservação prospetiva e retrospectiva. Enquanto a preservação prospetiva se foca na prevenção e redução dos danos, a preservação retrospectiva tem como objetivo reparar os documentos já danificados. Porém, para Galindo (2012, p. 40), a tentativa comum de antecipação de danos/minimização de perda de dados já não é suficiente. Para este autor, o atual contexto da preservação digital requer instrumentos mais eficientes que sejam capazes de proceder a avaliações dos riscos a que os recursos estão sujeitos, de prever o impacto que essas ameaças trarão e de planear e implementar estratégias.

Já Beagrie e Jones (2008, p. 25) consideram três tipos de preservação: longo, médio ou curto prazo. A preservação de longo prazo foca-se no acesso contínuo e por tempo indeterminado aos objetos digitais ou pelo menos à informação neles existentes. Relativamente à de médio prazo, refere-se ao acesso contínuo – mas não por tempo indefinido – a objetos digitais mesmo existindo alterações tecnológicas. Por último, a preservação de curto prazo trata-se do acesso aos objetos digitais por um período de tempo definido enquanto ainda são utilizados (mas não para além de um futuro próximo) e/ou até que se torne inacessível devido aos avanços tecnológicos.

Para que um documento não esteja dependente da plataforma tecnológica existente na data da sua criação devem ser tomadas medidas e, dessa forma, assegurar não só o seu acesso prolongado mas também as suas características iniciais. Contudo, preservar não é simples nem barato: é necessário disponibilizar os recursos humanos, tecnológicos e materiais capazes de planear e colocar em prática um plano de preservação digital. Para além disto, a escolha das iniciativas de preservação deve-se fundamentar em alguns aspetos organizacionais, entre os quais: o tipo de empresa que está a conceber a informação digital; o sistema de informação existente, bem como a sua gestão e relação com os vários departamentos; normas próprias; ou a legislação do país em que se insere (Pinto, 2009^b, p. 157). É, por isso, premente que as decisões de preservação sejam enquadradas na gestão estratégica da organização e que a equipa responsável pelo desenvolvimento da política de preservação digital esteja ciente dos ambientes externo e interno, que influenciam o ambiente laboral da empresa.

Mas mesmo antes da elaboração e implementação dessa política de preservação digital, a organização deve assumir um papel ativo na gestão da informação. Devem ser definidos procedimentos que orientem os funcionários para os momentos de criação e difusão de objetos digitais de modo a garantir que a informação se mantenha autêntica e fidedigna. Apenas aquelas organizações que já tenham percebido a real importância da informação, dos fluxos informacionais e também do conhecimento que será gerado, é que estarão melhor preparadas para assumir a responsabilidade de desenvolverem um plano que permita salvaguardar a sua informação por um período de tempo prolongado.

Um claro sinal da pouca importância dada pelas organizações ao acesso continuado de documentos digitais é a afirmação feita pela Direção-Geral de Arquivos

(DGARQ)¹⁰ (2011, p. 6), com base em conhecimento empírico, em que refere que “a informação é produzida com carácter de utilização imediata sem serem consideradas necessidades operacionais sobre essa mesma informação a médio ou longo prazo” e esta inércia “resulta na perda, muitas vezes irreversível, de informação com consequências [...] para a instituição que a perdeu”.

Num mundo ideal, o profissional da informação deveria ser uma das peças fundamentais no ato de escolha ou desenvolvimento do sistema informático a ser utilizado na organização, uma vez que é ele que conhece os fluxos informacionais existentes. Este colaborador deve especificar os requisitos a que o sistema informático deve dar resposta, com o objetivo de que seja implementada a melhor solução para satisfazer as necessidades da sua organização.

Torna-se, então, importante que as organizações elaborem um documento estratégico onde estejam definidos os procedimentos a realizar para que se evite a perda de informação. A esse documento atribui-se a denominação de *plano de preservação digital*.

5 – Plano de preservação digital

A DGARQ (2011, p. 8) define o *plano de preservação digital* como “um documento estratégico que contém políticas e procedimentos orientados para a constituição de uma estrutura técnica e organizacional que permita preservar de forma continuada documentos de arquivos electrónicos através de acções realizadas sobre os objectos digitais que os compõem”.

Este documento é de relevante importância – pois pode ser considerado como o principal meio para promover e arrecadar fundos para a preservação da informação – e pressupõe que os objetivos da organização estejam bem definidos. Ele deve ser elaborado por uma equipa multidisciplinar composta por profissionais de várias áreas da organização. As áreas da gestão documental e informática assumem um papel preponderante na realização deste plano, uma vez que são os profissionais destas áreas que irão implementar as estratégias que serão estipuladas. Enquanto o profissional da informação possui conhecimento acerca dos documentos e suas

¹⁰ Atual Direção-Geral do Livro, dos Arquivos e das Bibliotecas (DGLAB). A DGLAB resultou da fusão entre a Direção-Geral do Livro e das Bibliotecas e a DGARQ. Disponível em [www: http://dglab.gov.pt/](http://dglab.gov.pt/).

formalidades, o profissional da informática tem conhecimento dos sistemas eletrônicos. No entanto, a realização do plano de preservação digital deve também unir esforços de outros departamentos, como o financeiro, e todos os outros que produzam informação digital (DGARQ, 2011, p. 9). Embora seja muito importante que todos os departamentos anteriormente referidos tenham uma voz ativa no processo e demonstrem as suas necessidades informacionais para poderem realizar as suas tarefas quotidianas, é importante que a responsabilidade pela preservação digital seja atribuída a um colaborador específico: o gestor. Essa pessoa designada estaria, então, encarregue por orientar o planeamento do projeto, pela administração das tarefas e pela seleção de pessoal. O gestor teria também de se manter constantemente atualizado acerca das tecnologias existentes e emergentes, das iniciativas e estratégias de preservação digital e, por fim mas não menos importante, possuir conhecimento ao nível da gestão da informação.

Nesse sentido, é crucial que os gestores de topo estejam conscientes de que a implementação de uma política de preservação vai trazer grandes mudanças na vida da organização e que será necessário disponibilizar verbas para adquirir toda a infraestrutura necessária. Esta equipa multidisciplinar irá precisar de muitas horas de trabalho até que os seus elementos cheguem a um consenso e até que, de facto, seja possível observar os resultados efetivos e que irão provocar mudanças ao nível da cultura da organização.

Teixeira (2011, p. 273-274) considera a cultura organizacional como sendo um conjunto único de características que distinguem uma instituição das demais e que são transmitidas aos seus colaboradores através de rituais, valores e crenças. Por conseguinte, tais normas de comportamento devem ser incutidas nos novos colaboradores de modo a que compreendam o ambiente de trabalho, de comunicação e de relacionamento existente naquela organização.

Da mesma forma que um novo colaborador se adapta às normas e valores da sua organização, também todos os colaboradores se devem adaptar às políticas de preservação digital. Porém, não estamos perante uma aceitação e adaptação imediata, até porque, dependendo da situação em que a organização se encontra no que toca à gestão da informação e da documentação, pode envolver grandes mudanças e, conseqüentemente, gerar resistência à sua implementação por parte dos trabalhadores. Cabe, então, à equipa multidisciplinar responsável pela elaboração do plano de preservação digital, desenvolver manuais e realizar ações de formação junto

dos restantes colegas com o objetivo de explicar os objetivos, as mudanças e as novas práticas que serão adotadas e que devem ser aplicadas a partir daquele momento.

5.1 – Pré-requisitos

Para que uma empresa assimile a necessidade de preservar a sua informação é necessário que, previamente, já tenha entendido a importância da gestão da informação e da gestão documental. Apenas uma organização consciente das vantagens que a informação lhe pode trazer, nomeadamente a geração de novo conhecimento, estará apta e será capaz de desenvolver os esforços necessários para o desenvolvimento e implementação de políticas de preservação digital.

Para a elaboração de um plano de preservação digital existem dois elementos essenciais: o plano de classificação e a tabela de seleção.

O plano de classificação promove um conjunto de regras para a gestão eficaz de documentos de arquivo criados ou utilizados por uma organização e reflete o levantamento da documentação produzida por cada departamento e das atividades por eles levadas a cabo para a prossecução das suas funções. Esta ferramenta permite perceber relações entre documentos; garantir que um documento mantém a sua denominação em todo o seu percurso; auxiliar no processo de recuperação e utilização de documentos; e facilitar o processo de avaliação e seleção de documentos a preservar ou suprimir (DGARQ, 2011, p. 12-13).

A tabela de seleção reflete o processo de avaliação de documentos e para cada tipo de documento, define o seu prazo de preservação e/ou eliminação dos documentos de arquivo, baseando-se nas normas da organização e na legislação vigente no país. Se esta ferramenta tiver sido desenvolvida tendo como alicerce a utilidade de um tipo de documento para as atividades da organização e a sua validade temporal, então o processo de preservação digital será facilitado, pois podem-se escolher, de uma forma simples, quais os documentos e sistemas de informação que deverão ser alvo de ações de salvaguarda. Apenas aqueles documentos cuja utilidade operacional supere o horizonte de obsolescência¹¹ é que deverão ser objetos de

¹¹ O horizonte de obsolescência é o período estimado para que o meio ambiente tecnológico (geralmente, formatos de documentos) existente no momento de produção do objeto digital se torne ultrapassado. O valor deste horizonte depende do prazo de retrocompatibilidade garantido pelos fabricantes, ou seja, uma nova versão é compatível com a versão anteriormente lançada (DGARQ, 2011, p. 14-15).

ações de preservação, uma vez que é inviável preservar toda a informação criada no seio da organização (DGARQ, 2011, p. 13-14).

Campos (2002, p. 9) considera que o tema da preservação digital levanta ainda outras questões para além da seleção do que deve ser preservado, nomeadamente a forma como a preservação será efetuada, quem é o responsável, quais os custos, quem os paga, quais as estratégias a aplicar, quem estará apto a aceder aos objetos e por quanto tempo eles estarão acessíveis.

5.2 – Planeamento

Segundo a DGARQ (2011, p. 21-25), durante a preparação do plano de preservação digital, que pode englobar uma ou várias estratégias a serem aplicadas, existem vários aspetos que a equipa multidisciplinar deve considerar para realizar as suas escolhas, nomeadamente:

- O investimento necessário:
 - Que dispositivos tecnológicos existem, verificar se podem ser reaproveitados ou se existe necessidade de substituição;
 - Se implicam mudanças na estrutura física da organização, ou seja, se pressupõe que se passe a ter ainda maiores cuidados com o meio ambiente que envolve os dispositivos eletrónicos;
 - Se os recursos humanos existentes incluem pessoas capacitadas para fazer funcionar o plano de preservação digital ou se terão de existir novas contratações;
- Os custos de manutenção: devem-se prever quais os custos das soluções que se pretendem aplicar e, nesse sentido, procurar entender se se trata de uma planificação financeira e humanamente viável a longo prazo;
- Possíveis alterações que possam vir a ocorrer nas atividades levadas a cabo pela organização.

Tendo em conta os aspetos supracitados, torna-se necessário que os gestores, que realmente estejam preocupados com a salvaguarda da memória da sua instituição, considerem – de uma forma eficaz – as iniciativas de preservação digital como um aspeto a ter em conta na gestão estratégica da organização. Tal assunção

pode levar à retificação da missão, objetivos e metas que a organização se propõe atingir e também das ações que estavam previstas para que tal acontecesse.

Dependendo dos fluxos informacionais da organização, um objeto digital, ao longo do seu ciclo de vida, pode estar associado a vários *stakeholders*. É fulcral que seja estabelecida comunicação entre eles, com vista à preservação de longo prazo, de modo a que a equipa multidisciplinar seja capaz de reunir as necessidades informacionais, organizacionais e legais dos utilizadores e possa planear e implementar as ações mais adequadas para as colmatar.

É importante que todas as ações a realizar ao longo do ciclo de vida¹² da informação e respetivos responsáveis estejam contemplados neste documento, desde o momento da sua criação até à altura em que deixa de ser necessária. As ações de preservação devem ser aplicadas desde a fase de produção do documento, tendo, portanto, o seu criador, uma responsabilidade acrescida. É, então, desejável que, no momento de criação de um documento, a organização já tenha pré-definido o formato em que ele será armazenado e preservado, evitando-se assim gastos e possíveis perdas de informação desnecessários em conversões posteriores.

Pinto (2009^b, p. 157) considera que a organização deve adotar normas e especificações abertas e com provas dadas acerca da sua capacidade e eficácia no que concerne ao desenvolvimento, armazenamento e difusão de objetos digitais, que permitam garantir a interoperabilidade entre diferentes plataformas e, também, a longevidade da informação existente. Esta autora refere ainda que as organizações devem colaborar com as empresas que desenvolvem sistemas informáticos, criando

¹² O ciclo de vida de um documento é composto por três momentos: fase corrente, intermédia e permanente. Esta divisão é feita com base na frequência de uso e na utilidade que o documento tem para o serviço que o criou. De seguida, são explicadas, de forma sucinta, as três idades pelas quais um documento pode passar (Sant'Anna, 2001, p. 125-126; Decreto-Lei nº 16/93, 1993, p. 264):

- O documento encontra-se na fase corrente se a atividade que lhe deu origem ainda decorre, ou seja, é consultado com frequência e deve ser de fácil e rápido acesso. O documento digital é, geralmente, armazenado no computador do seu produtor ou armazenado em servidores da empresa. No caso de documentos em suporte físico (nomeadamente, em papel) deve-se encontrar perto do local de trabalho;
- O documento encontra-se na fase intermédia se a atividade que lhe deu origem já terminou, ou seja, deixa de ser necessário para garantir as tarefas do dia-a-dia da organização. Os documentos digitais podem ser transferidos para outros suportes, enquanto os documentos tradicionais podem ser guardados em arquivos centrais, ficando a aguardar o cumprimento dos prazos legais estipulados na tabela de seleção;
- O documento encontra-se na fase permanente quando deixa de ter vigência administrativa, isto é, deixa de ter valor para a organização. São conservados permanentemente devido ao ser carácter probatório, histórico ou informativo.

assim sinergias essenciais para conceber soluções que deem resposta às suas necessidades informacionais e aos aspetos legais em vigor.

Não obstante, existem também outros aspetos técnicos a ter em consideração no momento de escolher o *hardware* e *software* necessários para preservar e aceder à informação em meio digital, como a necessidade de sistemas de armazenamento de alta capacidade e com mecanismos de segurança, estruturas fiáveis de cópias de segurança e sistemas de deteção e recuperação automática de falhas.

Relativamente ao fator humano, é necessário a alocação de verbas que permitam que esta equipa multidisciplinar esteja capacitada a gerir a infraestrutura tecnológica existente. Porém, dado o rápido avanço da tecnologia é crucial que a equipa esteja atenta às mudanças no ambiente digital – quer para prever quais dispositivos ou formatos se irão tornar obsoletos como para registar os emergentes – e em constante aprendizagem para que possa, de forma eficaz, readaptar a infraestrutura existente e, conseqüentemente, as ações de preservação que já estavam definidas.

Também a metainformação – principalmente a metainformação de preservação – assume um papel de grande relevância na implementação da solução de preservação a longo prazo numa organização, pois permite descrever o contexto de criação de um documento, bem como dados de possíveis edições futuras ou ações de preservação efetuadas.

Tendo em conta que a preservação digital está intrinsecamente ligada à evolução das TIC, é fulcral que o plano de preservação digital seja revisto periodicamente para que possa ser ajustado às mudanças que vão ocorrendo não só a nível tecnológico mas também organizacional e legal.

5.3 – Implementação

Antes das estratégias definidas no plano de preservação digital serem postas em prática na sua totalidade, devem ser feitas cópias de segurança dos objetos digitais que se pretendem salvaguardar. Estas cópias devem ser armazenadas em local seguro e sobre elas não devem ser efetuados quaisquer tipos de ações, pois servirão de *backup* caso alguma atividade realizada no objeto original não corra pelo melhor. A existência de um *backup* é particularmente importante na fase de testes

uma vez que o processo de migração de dados entre plataformas é propenso à existência de danos e perdas de informação (DGARQ, 2011, p. 26).

A partir do momento que os resultados dos testes à solução proposta se mostrem regulares, então as ações de preservação podem ser estendidas a todos os objetos digitais selecionados.

Depois das estratégias serem implementadas, os objetos digitais devem ser alvo de nova avaliação com o objetivo de verificar se mantiveram a sua integridade e funcionalidades. Também a metainformação desses objetos deve ser reavaliada e atualizada: devem ser mencionadas as alterações ocorridas de modo a formar um registo com as ações sofridas.

6 – Ameaças

Como já foi referido anteriormente, a DGARQ conclui que, por muitas vezes, os documentos são criados para servirem um propósito imediato, acabando as instituições por esquecer um eventual acesso futuro que possa vir a ocorrer. Mas a esta afirmação é possível acrescentar que muitas organizações se preocupam meramente com o conteúdo existente num documento digital, não dando valor ao meio tecnológico que permite a criação e a utilização dessa informação.

No entanto, por vezes, esse contexto intermediário entre o utilizador e a informação sofre mudanças e evoluções e a organização pode ver “informação probatória e indispensável para assegurar o seu posicionamento no ambiente externo, dissipar-se antes do seu período de tempo de vida operacional terminar” (Barbedo, 2005, p. 7).

Podemos então considerar vários tipos de ameaças: a obsolescência tecnológica – que pode afetar tanto os suportes como os formatos –, os danos físicos que o material eletrónico pode sofrer e também a ação humana.

No Anexo A é apresentado um conjunto de boas práticas para uma eficaz manutenção da informação em suportes digitais.

6.1 – Obsolescência tecnológica

Hoje em dia é criado um número abundante de documentos com recurso a plataformas eletrónicas que tendem a ser cada vez mais rápidas, com capacidade de armazenamento cada vez maior e com custos mais baixos. Embora as funcionalidades e capacidades de *software* e *hardware* tenham vindo a aumentar, os seus ciclos de obsolescência têm vindo a tornar-se cada vez mais curtos, devido a uma indústria informática extremamente competitiva que luta para se manter equilibrada face aos seus concorrentes, o que faz com que os produtos se tornem ultrapassados mais rapidamente.

Os elementos lógicos (*software*) são mais complicados de gerir do que os elementos físicos (*hardware*): o *software* depende de um sistema operativo composto por várias camadas, e pode ser parametrizado com módulos ou *plug-ins*. É o conjunto destes elementos que faz com que a aplicação possa ser executada e a informação lida (Figuerola *et al.*, 2009, p. 267).

6.1.1 – Obsolescência de formatos e *software*

Para que um objeto digital possa ser interpretado pelo conjunto *software-hardware* deve obedecer às especificações do seu formato. A DGARQ (2011, p. 29-30) apresenta algumas das razões que justificam a falha na interpretação pela plataforma tecnológica, nomeadamente:

- O aparecimento de um novo formato de ficheiro mais complexo que torna um formato mais antigo ultrapassado;
- Deixa de ser criado *software* compatível para determinado formato.

Durante muito tempo, um dado formato de um documento estava intrinsecamente ligado ao *software* que lhe deu origem, não sendo possível abrir esse ficheiro noutra aplicação, uma vez que as suas especificações não eram divulgadas pelas empresas que os detinham. Se esse *software* deixasse de ser produzido ou se os custos para a sua aquisição fossem inabarcáveis para a organização, então iria-se perder o acesso a essa informação. Para evitar que este tipo de problemas aconteça, é necessário fazer a escolha acertada de formatos e *software* – aquando da elaboração do plano de preservação digital.

Idealmente devem ser escolhidos formatos com especificações abertas e *software* não proprietário com provas dadas da sua eficácia. Por outras palavras, recomenda-se a utilização de formatos cujas especificações sejam divulgadas na totalidade e estejam bem documentados, pois desse modo existe a possibilidade de outros programadores desenvolverem produtos capazes de interpretar documentos armazenados nesses formatos. A seleção dos formatos deve recair sobre aqueles que não possuam restrições a nível de direitos de autor ou de licenças, com o objetivo de assegurar a independência de *software* e *hardware* e a durabilidade da informação.

A escolha do *software* deve também ter em conta a garantia de retrocompatibilidade com versões anteriores e dos ciclos de atualização; o conjunto de funcionalidades existentes; a panóplia de formatos que podem ser interpretados; e, por último mas não menos importante, a possibilidade de suportar metainformação, que é um aspeto tão importante para o sucesso da preservação digital numa organização (DGARQ, 2011, p. 29-30). Embora os fabricantes de produtos eletrónicos procurem garantir que as novas versões dos seus artigos sejam compatíveis com as anteriores (retrocompatibilidade), isso apenas é possível durante um período de tempo não muito longo pelo que se torna necessário uma posterior atualização.

A divulgação do código-fonte e a utilização de especificações abertas garantem a interoperabilidade entre aplicações e plataformas que tenham como origem o mesmo código.

Um exemplo de formato aberto é o PDF (*Portable Document Format*), que inicialmente era um formato proprietário da Adobe, e cujas especificações acabaram por ser reveladas na sua totalidade. Deu origem ao PDF/A definido pela norma ISO 19005-1:2005 e é o padrão para a preservação a longo prazo de documentos digitais.

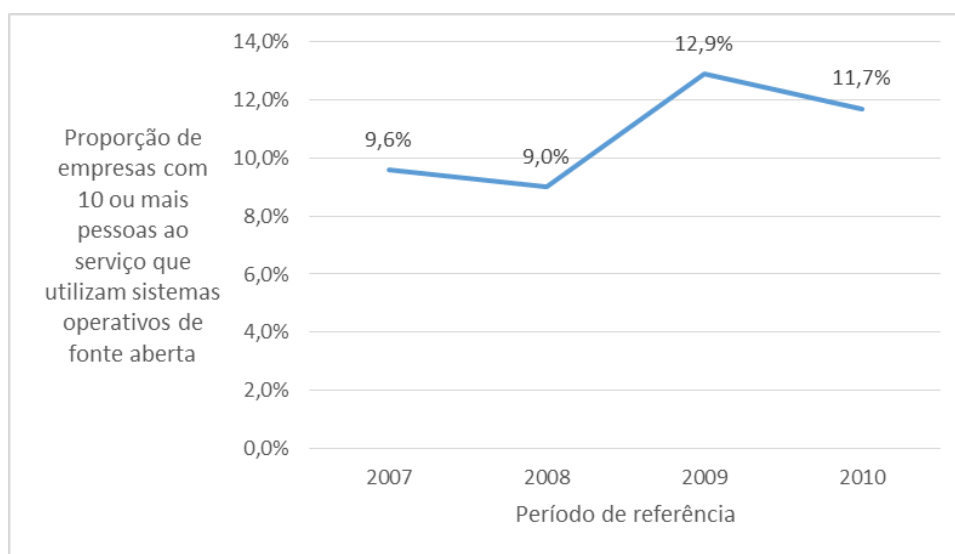
Outro exemplo de formato aberto é o *OpenDocument Format* (ODF)¹³ baseado na linguagem XML (*eXtensible Markup Language*)¹⁴ e desenvolvido pela OASIS (*Organization for the Advancement of Structured Information Standards*). É também definido pela norma ISO/IEC 26300 e pretende ser o padrão para a criação e edição de documentos de aplicações de escritório.

¹³ Disponível em [www: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office).

¹⁴ Disponível em [www: http://www.w3.org/XML/](http://www.w3.org/XML/). Trata-se de uma linguagem aberta e desenvolvida pelo *World Wide Web Consortium* (W3C).

São exemplos de *software* aberto e livre o Apache OpenOffice¹⁵ e o LibreOffice¹⁶ que permitem interpretar quer formatos ODF como também formatos proprietários da Microsoft.

Parece importante referir o estudo feito pelo INE (Instituto Nacional de Estatística), representado no Gráfico 1, que revela a percentagem de empresas em Portugal com 10 ou mais pessoas ao serviço que utilizam sistemas operativos de fonte aberta.



Fonte: INE [Em linha]

Gráfico 1 – Proporção de empresas com 10 ou mais pessoas ao serviço que utilizam sistemas operativos de fonte aberta – 2007 a 2010

Através da análise do gráfico anterior é possível concluir que nos quatro anos visados a tendência de utilização de sistemas operativos *open source* é inconstante. Em 2009 existiu uma subida de quase 4% em relação ao ano transato mas, em 2010, voltou a descer novamente para perto de 12%. Tendo em conta o leque de vantagens que o uso de *software* de código aberto traz para o acesso a longo termo da informação e também a conjuntura económico-financeira vivida em Portugal seria de esperar que a taxa de empresas que já tivessem aderido a este tipo de aplicações fosse superior.

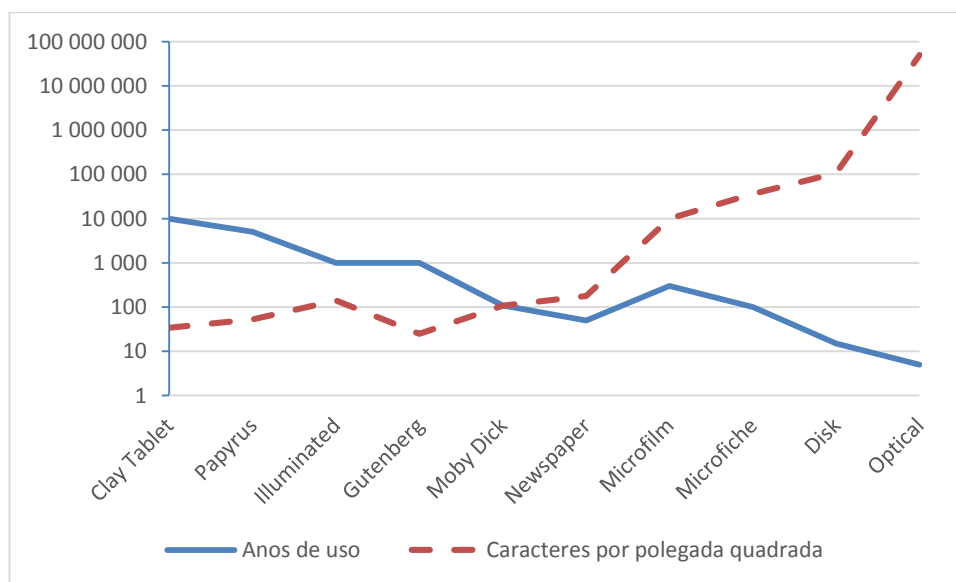
¹⁵ Disponível em [www: http://www.openoffice.org/pt/](http://www.openoffice.org/pt/).

¹⁶ Disponível em [www: https://pt.libreoffice.org/](https://pt.libreoffice.org/).

6.1.2 – Obsolescência de suporte e *hardware*

No caso dos suportes e *hardware*, a obsolescência destes dispositivos é marcada pela sua constante evolução no que concerne às características técnicas que são disponibilizadas. Isto é, têm existido inúmeras melhorias no que toca aos componentes dos computadores, nomeadamente o aumento da velocidade dos processadores, as constantes melhorias nos aspetos gráficos ou as taxas de transmissão entre equipamentos.

Curiosamente, embora a capacidade de armazenamento dos dispositivos tenha vindo a crescer exponencialmente, o espaço físico ocupado tem vindo a ser cada vez menor. Com a constante evolução tecnológica a que temos assistido, cada vez vemos dispositivos de armazenamento mais pequenos mas com características melhores. No Gráfico 2 é possível verificar a evolução dos suportes, ou seja, podemos observar a densidade da informação de cada um dos suportes (por polegada quadrada) em comparação com a sua esperança de vida. Atualmente, os suportes óticos, apesar de terem capacidade para armazenar um elevado número de informação em pouco espaço, têm um horizonte de obsolescência de cerca de cinco anos.



Fonte: Conway Cit. por Keefer e Gallart, 2007, p. 61

Gráfico 2 – Densidade de informação vs. Esperança de vida

O habitual exemplo da obsolescência de suportes/*hardware* foi a descontinuação da produção de computadores com leitores de disquetes de 5¼ ou 3,5 polegadas em detrimento dos CD/DVD e dos dispositivos de armazenamento USB. Aqueles que não transferiram os seus documentos para outros suportes, acabaram, mais tarde ou mais cedo, por deixar de lhes poder aceder.

No entanto, tem vindo a ocorrer uma mudança de paradigma relativamente aos dispositivos de armazenamento. Hoje em dia quase todos os dispositivos tecnológicos (computadores, *smartphones*, *tablets*) estão ligados à Internet, o que originou o aparecimento de plataformas na nuvem¹⁷, geridas por terceiros (Ferreira, Saraiva e Rodrigues, 2012, p. 11). A maioria destes servidores são independentes do sistema operativo e basta apenas que os dispositivos tenham ligação à Internet para que o serviço esteja disponível a qualquer altura e em qualquer lugar.

A omnipresença da informação garantida pela *cloud* vai fazer com que a utilização de dispositivos de armazenamento físicos diminua pouco a pouco. Contudo, há também cuidados a ter com este tipo de ferramentas. Apesar de se tornarem um aliado para o acesso fácil à informação, muitas destas plataformas mais conhecidas (Dropbox¹⁸, Google Drive¹⁹) permitem a partilha de documentos e são também colaborativas. Ou seja, quem faz o *upload* da informação deve assegurar que apenas determinado grupo de colaboradores pode aceder a certos documentos e também deve tomar as medidas necessárias para evitar a edição ou eliminação (propositada ou acidental) dessa informação. Devem ser garantidas as condições necessárias para asseverar a autenticidade, fidedignidade e privacidade dos documentos.

6.2 – Danos físicos

Tal como a informação em papel, também a informação em meio digital está sujeita a danos nos seus suportes.

Desde o momento de fabricação e durante o processo de aquisição, instalação e utilização na organização deve ser dada a devida importância ao meio ambiente que rodeia os suportes digitais uma vez que os níveis de humidade,

¹⁷ Nuvem (ou em inglês *cloud*) consiste num “conjunto de servidores remotos alojados na Internet, utilizados para armazenar, gerir e processar dados em vez dos servidores locais ou de computadores pessoais” (Infopédia, 2003-2015°).

¹⁸ Disponível em [www: https://www.dropbox.com/](https://www.dropbox.com/).

¹⁹ Disponível em [www: https://drive.google.com/](https://drive.google.com/).

temperatura e luz e a proximidade de campos magnéticos podem resultar num efeito nefasto. Embora neste processo existam alguns momentos em que a organização adquirente é alheia, após a sua aquisição devem ser garantidas as condições ótimas para um correto funcionamento e manutenção dos suportes digitais (Keefer e Gallart, 2007, p. 61-62). Devem também ser previstos meios de combate a incêndios ou inundações.

A falta de cuidado no manuseamento dos suportes de armazenamento pode também ser um fator crítico dado que basta uma dedada ou arranhão num suporte ótico para que a informação deixe de estar acessível na totalidade. Também o uso excessivo de determinado dispositivo pode provocar o seu desgaste.

Basta existir um pequeno descuido para que o suporte não esteja totalmente intacto e, conseqüentemente, se torne ilegível. Assim, deixa de existir integridade dos dados e a informação deixa de poder ser interpretada.

6.3 – Ação humana

Nem sempre o ser humano age da melhor forma para com os documentos, quer estejam em suporte papel ou em suporte digital. A perda de dados em objetos digitais, para além das causas já referidas anteriormente, pode ter também origem humana.

Por um lado, as pessoas são negligentes com a informação digital e podem eliminá-la, sendo que esta ação pode ocorrer de forma acidental ou propositada. Por outra perspectiva, a organização pode ser alvo de atos de vandalismo ou roubo, existindo a possibilidade de ficarem privados da sua estrutura informática.

Para evitar males maiores, a organização deve implementar um controlo da informação, onde se pretende que apenas as pessoas/departamentos que criem/utilizem determinada informação lhe possa aceder. A criação de *backups* e a sua disponibilização em diferentes locais e, também, o controlo e armazenamento de versões dos documentos devem ser uma das preocupações dos administradores, para que, na eventualidade de existir perda de informação, se possa recuperar toda ou parte dela.

7 – Estratégias de preservação digital

Como já foi referido anteriormente, quer o *hardware* como o *software* possuem ciclos de atualização relativamente curtos, podendo-se tornar num obstáculo para a preservação digital. Para diminuir ou, quando possível, evitar a perda de informação, as organizações devem definir estratégias de preservação para cada um dos diferentes tipos de objetos digitais – durante a elaboração do plano de preservação digital – que melhor se adaptem à sua realidade. Ferreira (2006, p. 69) define *estratégia de preservação* como toda a “abordagem técnica que garante o acesso continuado à informação existente em formatos digitais”. A implementação de estratégias de preservação digital tem como objetivo possibilitar o acesso e a interpretação de informação que foi criada num ambiente tecnológico possivelmente diferente do atual. Esta tarefa requer, por vezes, mão-de-obra especializada.

Segundo Thibodeau (2002, p. 15-16), a escolha das estratégias de preservação a aplicar nos objetos digitais devem-se basear em quatro critérios:

- Viabilidade: implica existência de *hardware* e *software* capazes de suportar uma determinada estratégia;
- Sustentabilidade: indica a capacidade de uma técnica poder ser usada indefinidamente ao longo do tempo ou a crença de um eventual aparecimento de outra técnica, permitindo, dessa forma, assegurar uma sequência lógica de preservação;
 - Uma estratégia deve dar garantias de imunidade/isolamento aos efeitos da obsolescência tecnológica;
 - Uma estratégia deve permitir a interação com outros métodos;
- Praticabilidade: requer uma análise ao grau de dificuldade e custos de implementação e manutenção de uma estratégia, verificando se os valores estão dentro de limites razoáveis;
- Adequabilidade: depende dos tipos de objetos a preservar e dos objetivos de preservação por parte da organização. Ou seja, a aplicação de uma determinada técnica varia consoante o tipo de objeto digital e se a organização procura preservar a sua dimensão física/lógica ou conceptual.

Num processo de preservação digital a longo prazo não se pode esperar que um objeto digital seja alvo de um único tipo de estratégia de preservação. Para Lee *et al.* (2002, p. 103), o sucesso de uma política de preservação digital reside na aplicação combinada de diferentes estratégias, ou seja, uma abordagem pode incluir aspetos baseados em diferentes técnicas.

Ao longo dos anos são várias as classificações e estratégias propostas por diferentes autores. De seguida são apresentadas três das várias classificações existentes, elaboradas por Beagrie e Jones, Thomaz e Soares e, por último, Thibodeau.

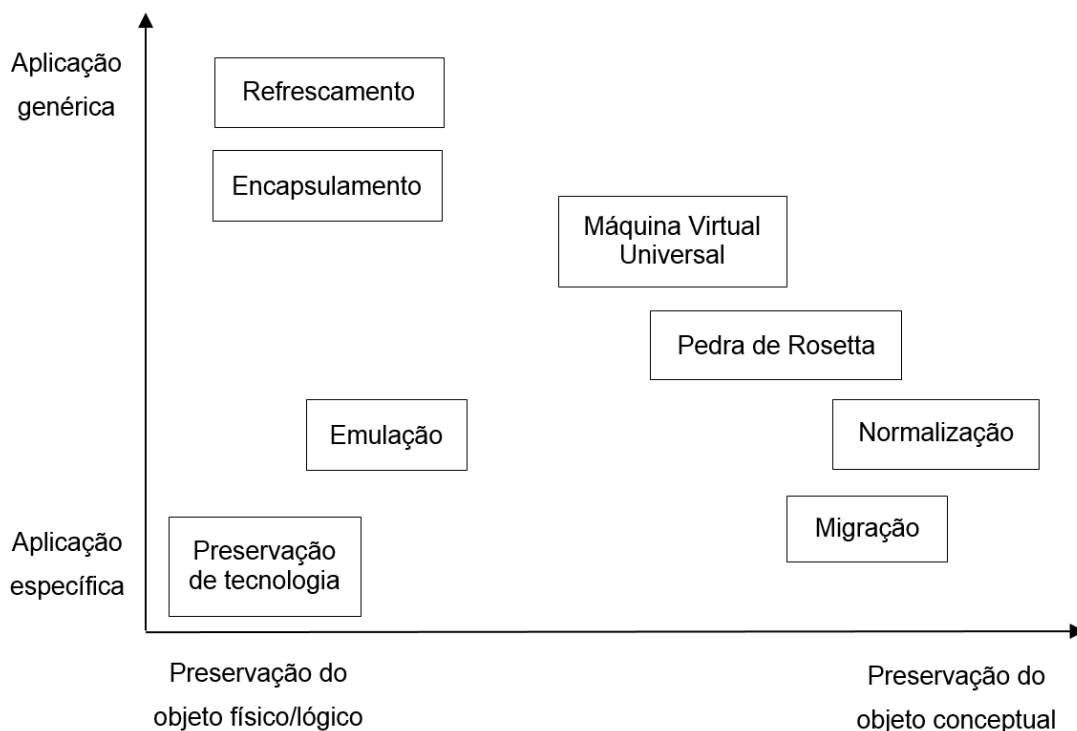
Beagrie e Jones (2008, p. 111) consideram a existência de dois tipos de estratégias de preservação: as primárias e as secundárias. Definem estratégias primárias como sendo aquelas que podem ser escolhidas por um repositório para a preservação de materiais digitais a médio e longo prazo. Já as estratégias secundárias podem ser aplicadas a curto e médio prazo, seja pelos repositórios responsáveis pela preservação como por todos aqueles que tenham interesse nos materiais digitais (como os criadores). Estas estratégias podem anteceder as primárias e a sua aplicação pode modificar de tal ordem o objeto digital que seja necessária uma reavaliação do seu estado, podendo resultar no adiamento da aplicação de estratégias primárias. Por isto, Beagrie e Jones consideram que a denominação de *estratégias secundárias* não significa necessariamente a sua inferioridade em relação às estratégias primárias. Não obstante, estes autores consideram a migração e a emulação como estratégias primárias e referem a existência de prós e contras em cada uma delas. No entanto, a necessidade de se aplicar alguma destas estratégias pode ser minimizada, ou até mesmo evitada, se se aplicarem procedimentos preventivos de preservação. Entre os quais podemos encontrar um conjunto de boas práticas ao nível do armazenamento e manutenção dos dispositivos tecnológicos e uma escolha criteriosa de estratégias de preservação secundárias (preservação de tecnologia, normalização, encapsulamento, entre outras).

Para Thomaz e Soares (2004, p. 4) existem dois tipos de estratégias: as estruturais e as operacionais. As estruturais “dizem respeito aos investimentos ou esforços iniciais por parte da instituição, preparando seu ambiente para o processo da preservação digital”, enquanto as operacionais se referem às “atividades ou medidas concretas de preservação digital”. Relativamente às estratégias estruturais referem: a adoção de padrões; a elaboração de manuais; os metadados para preservação digital;

a montagem de infraestrutura para preservação digital; e a formação de uma rede de relações. Quanto às estratégias operacionais, consideram: a escolha do meio de armazenamento; a migração; a emulação; a impressão em papel ou microfilme; e a conservação da tecnologia.

Já Thibodeau (2002, p. 17-18) considera que a seleção de estratégias de preservação envolve uma gama de aplicabilidade desses métodos tendo em conta a quantidade e variedade de objetos a preservar. Propõe uma classificação bidimensional, em que, por um lado, se pode encontrar o tipo de aplicação da estratégia (específica ou genérica) e, por outro, o seu objetivo (preservação da tecnologia ou do objeto). Esta classificação não reflete todas as estratégias de preservação que têm vindo a ser propostas ou implementadas ao longo dos anos. De facto, não inclui métodos focados nos metadados, inclui sim uma gama de métodos que permitem ultrapassar a questão da obsolescência tecnológica.

Ferreira adaptou a proposta de Thibodeau e apresentou uma classificação com um número mais reduzido de estratégias em relação à original (Figura 3).



Fonte: Ferreira, 2006, p. 32

Figura 3 – Proposta de classificação de Thibodeau adaptada por Ferreira

Através da análise do eixo horizontal da Figura 3, podemos encontrar à esquerda as estratégias centradas na conservação do objeto físico ou lógico, enquanto à direita se dispõem as estratégias com vista à conservação das características essenciais do objeto.

A classificação das estratégias de preservação digital não é consensual: algumas das técnicas referidas neste trabalho não são consideradas como estratégias por alguns dos autores consultados. Isto deve-se ao facto de se tratarem de soluções cuja implementação deve ser feita a curto e médio prazo e, por isso, são tidas como um conjunto de boas práticas. Tais práticas podem ser aplicadas em conjunto com as estratégias efetivas de preservação que, por sua vez, se refletem a médio e longo prazo. Tendo em conta estes aspetos, Lee *et al.* (2002, p. 94, 102-103) consideram como estratégias efetivas para a preservação a longo termo: a migração, a emulação e o encapsulamento. No Anexo B é possível encontrar um esquema que auxilia na seleção da estratégia mais adequada dentro das três referidas.

Neste trabalho será seguida a linha orientadora proposta por Ferreira em *Introdução à Preservação Digital – Conceitos, estratégias e actuais consensos*. Importa também referir que no Anexo C podem ser encontradas vantagens e desvantagens de algumas das estratégias que serão apresentadas.

7.1 – Preservação de tecnologia

A *Preservação de tecnologia* implica a conservação, em boas condições, de todo o sistema informático (*hardware* e *software*) existente no momento de criação ou captura dos objetos digitais que se quer ver preservados – mesmo que já se tenha tornado obsoleto. Isto é, pretende garantir a manutenção do sistema intermediário capaz de tornar a informação digital interpretável para o utilizador sempre que haja necessidade de ser recuperada. A preservação de tecnologia incide sob a conservação do objeto digital original em todas as suas dimensões (física, lógica e conceptual), mantendo assim as suas características comportamentais e o seu aspeto visual.

Ao adotar esta estratégia, a organização deve estar ciente de que toda a tecnologia se torna inevitavelmente obsoleta, o que faz com que a informação deixe de estar acessível. A existência de museus tecnológicos requer a alocação de

recursos financeiros para a manutenção do espaço e suporte técnico necessários, o que se torna impraticável a longo prazo²⁰.

Esta estratégia apenas deve ser aplicada, a curto e médio prazo, em casos específicos em que o objeto digital não possa ser convertido para outros formatos, seja porque foi criado num formato proprietário ou porque o *software* já se tenha tornado obsoleto. Deve então ser considerada como uma medida temporária que permite prolongar o período de tempo até que seja possível implementar outras estratégias de preservação mais eficazes.

7.2 – Refrescamento

A técnica de *Refrescamento* trata-se de uma ação preventiva e consiste na transferência periódica dos objetos digitais existentes num dado suporte de armazenamento para outro mais recente, antes que o primeiro se torne obsoleto ou sofra deterioração. Esta técnica não resolve por si só os problemas da obsolescência tecnológica, apenas os atrasa. A frequência com que se realiza as ações de refrescamento depende das políticas instituídas pela organização e tendo em consideração a evolução tecnológica. Contudo, para que isto seja possível é essencial que o suporte seja frequentemente avaliado e a sua integridade seja assegurada para que a informação possa ser devidamente interpretada pelo sistema informático (Ferreira, 2006, p. 33). É também essencial que os dispositivos periféricos capazes de fazer a leitura dos suportes de armazenamento demonstrem fiabilidade, mesmo após a sua obsolescência. Apesar das verificações periódicas à qualidade dos dispositivos e dos dados existentes, o refrescamento não implica a existência de elevados conhecimentos técnicos especializados.

Alguns autores abordam o refrescamento de diferentes formas. Por exemplo, Borghoff *et al.* (2005, p. 38-39) fazem distinção entre dois tipos de refrescamento: aquele em que os objetos são transferidos entre suportes do mesmo tipo, e aquele em que são transferidos entre suportes de diferentes tipos, atribuindo-lhe o nome de replicação. No primeiro caso não ocorre alteração ao nível da sequência de *bits* que

²⁰ Note-se, por exemplo, que por maiores cuidados que se tenham na preservação de *hardware*, é certo que este virá a degradar-se ao ponto de deixar de funcionar. E, infelizmente, com o avanço tecnológico, o acesso a material antigo necessário para se proceder às reparações torna-se impossível.

forma o objeto digital, no entanto, na replicação, a estrutura interna do objeto pode variar de forma considerável. A forma como a informação está armazenada em suportes de diferentes tipos passa despercebida ao utilizador uma vez que a sua leitura é responsabilidade do sistema informático. Contudo, os utilizadores devem estar conscientes de que o refrescamento entre suportes de diferentes tipos pode provocar problemas no acesso dos objetos, se a eles estiverem associadas funcionalidades que estejam dependentes das características físicas do suporte ou da sequência de *bytes* do objeto²¹ (por exemplo, funcionalidades associadas à garantia de autenticidade ou proteção contra cópias).

Já Keefer e Gallart (2007, p. 82) consideram o refrescamento como a transferência de dados entre suportes sem que exista alteração de *software* ou do formato. Para estas autoras a distinção da transferência de informação entre suportes do mesmo tipo ou da transferência entre tipos diferentes não se aplica. No entanto, alertam para as possíveis semelhanças que o refrescamento possa ter com a realização de cópias de segurança, embora se tratem de processos com diferentes objetivos e que podem seguir abordagens distintas. Enquanto o refrescamento tem como objetivo a preservação, as cópias de segurança estão previstas para a recuperação de dados em caso de desastre e, por isso, os dispositivos considerados para esse efeito devem também ser alvos de ações de preservação ao longo do tempo para que a informação existente possa continuar a ser acedida.

Para Pinto (2009^b, p. 146) o refrescamento de suporte não constitui uma solução de preservação, mas sim um dos pontos de partida para a aplicação de um plano de preservação na sua totalidade. Trata-se de uma medida que vai ocorrer várias vezes ao longo do ciclo de vida dos objetos digitais e que deve ser utilizada em paralelo com outras estratégias porque, devido à rápida obsolescência de *hardware*, torna-se fundamental a transferência do objeto digital para suportes mais atuais.

7.3 – Emulação

A *Emulação* faz uso das tecnologias atuais para simular o comportamento do conjunto de *hardware* e/ou *software* que existia no momento da criação dos objetos digitais e que, entretanto, se tornou ultrapassado. Para a criação de um emulador,

²¹ Por exemplo, a conversão de objetos entre arquiteturas *Little Endian* e *Big Endian*.

capaz de simular o ambiente tecnológico existente anteriormente, é necessário que todo o contexto do ambiente de criação esteja devidamente documentado, desde a arquitetura de *hardware*, o funcionamento do sistema operativo e das aplicações de *software* que lidavam com o objeto digital. Além disso, são necessários recursos humanos qualificados. Segundo Lee *et al.* (2002, p. 102), a emulação é a escolha mais adequada quando existem poucas informações acerca do formato interno do objeto digital.

Borghoff *et al.* (2010, p. 57-58) consideram que a forma mais segura de preservar os documentos digitais de forma autêntica para as gerações vindouras, é proceder ao seu armazenamento no formato inicial e assegurar a disponibilidade de uma versão operacional do sistema informático original. Como já foi referido anteriormente, a preservação de toda a infraestrutura tecnológica original é inviável a longo prazo e, por isso, a emulação surge como solução para reproduzir o seu comportamento nas gerações tecnológicas futuras.

Thibodeau (2002, p. 19-20) considera a existência de dois tipos de emuladores: aquele que simula uma aplicação para ser executada nas plataformas tecnológicas atuais ou aquele que simula comandos do sistema operativo antigo para poder ser executado nos recentes. Na primeira abordagem pode existir a necessidade de desenvolver um emulador para cada versão da aplicação, enquanto na segunda todas as aplicações executadas sob aquela versão do sistema operativo podem ser processadas usando o mesmo emulador. Por isso, a segunda abordagem torna-se mais simples e eficaz. Deixa assim de ser necessária uma verificação constante dos formatos dos objetos digitais uma vez que, mesmo sendo através de emulação, a plataforma digital original está a ser conservada.

Esta solução permite que o objeto digital original a preservar se mantenha fidedigno e autêntico, conservando as suas características e funcionalidades através da preservação da sua dimensão lógica.

Contudo, é uma estratégia difícil de implementar dada a exigência de conhecimentos técnicos especializados necessários para a implementação de um emulador fiável, mas muito importante para a preservação de objetos providos de características dinâmicas, como é o caso dos jogos de computador²². A necessidade de mão-de-obra especializada, capaz de implementar uma solução de emulação,

²² Alguns dos exemplos mais comuns da emulação para a preservação digital, é o seu uso na preservação de jogos (digitais) de arquiteturas obsoletas, como, por exemplo, ZX Spectrum ou Timex.

pode tornar esta solução financeiramente inviável para a organização. Outra grande desvantagem é o facto de o emulador também estar sujeito à obsolescência tecnológica necessitando também ele próprio de ser alvo de ações de preservação.

7.4 – Migração

A técnica de *Migração* consiste em “transportar periodicamente os recursos digitais de uma plataforma tecnológica para outra, adaptando-os aos ambientes de chegada” (Pinto, 2009^b, p. 146). Trata-se de um processo recorrente de transferência de documentos de um suporte ou formato para outros mais recentes, quando esses mesmos documentos são atingidos de forma efetiva pela obsolescência de *software* ou *hardware*, ou então como forma preventiva. É então essencial que a organização estabeleça procedimentos que permitam proceder a inspeções periódicas frequentes dos objetos digitais. Essas inspeções podem indicar a necessidade de novas ações de migração que devem ser desde logo planeadas e implementadas.

É importante não confundir a estratégia de migração com refrescamento, uma vez que se tratam de técnicas distintas. Enquanto o refrescamento procura manter a sequência de *bytes* de um objeto digital através da sua cópia para diferentes suportes, a migração procura transformar a representação lógica do objeto de acordo com as codificações necessárias para que a interpretação e apresentação possam ser feitas pelo novo *hardware* ou *software* (National Library of Australia, 2003, p. 134). A migração inclui o refrescamento, já que são feitas cópias entre diferentes suportes, mas difere dele uma vez que nem sempre é possível criar uma cópia digital exata ou replicar as funcionalidades e aparência e, mesmo assim, garantir a sua compatibilidade numa nova configuração tecnológica (Beagrie e Jones, 2008, p. 11). Assim, podemos dizer que o refrescamento procura contrariar a obsolescência dos suportes de armazenamento, ao passo que a migração visa minimizar não só a obsolescência de *hardware* mas também a de *software*.

A migração tem como objetivo deixar a informação acessível recorrendo às tecnologias futuras, sem necessidade de utilizar outros instrumentos, como é o caso dos emuladores. Enquanto a emulação preserva o objeto lógico, ou seja, mantém as codificações originais, a migração modifica a estrutura dos elementos de informação de um objeto digital de acordo com a nova plataforma. Isto é, a migração altera a dimensão lógica do objeto digital mas conserva a dimensão conceptual.

A reescrita do objeto digital que está a ser alvo de migração pode acrescentar detalhes que poderão ser considerados como corrupções ao seu estado inicial (Pinto, 2009^b, p. 146). A ocorrência destas falhas – que pode originar modificações a nível da estrutura, aparência e funcionalidades – está relacionada com incompatibilidades entre os formatos de origem e os de destino e/ou também com a utilização de conversores de baixa qualidade (Lawrence *et al.*; Rauber e Aschenbrenner Cit. por Ferreira, 2006, p. 36). É por isso que na elaboração do plano de preservação digital a escolha do formato de destino e do conversor a utilizar deve ser considerada de extrema importância. Deve-se garantir a melhor combinação entre formato de destino e conversor para que o resultado da estratégia seja o idealizado.

A aplicação da estratégia de migração está indicada para grandes quantidades de objetos digitais que possuam o mesmo formato. Os resultados de uma ação de migração podem ser potencializados se forem adotados formatos normalizados, não proprietários e cuja documentação se encontre divulgada publicamente e de forma gratuita²³. No entanto, esta estratégia não resolve de forma permanente os problemas da preservação digital, até porque o formato de destino vai, eventualmente, também se tornar obsoleto, o que irá originar a necessidade de um novo processo de migração (Ferreira, 2006, p. 36-37). O nível de corrupção referido no parágrafo anterior agrava-se com a realização consecutiva de migrações sob objetos já migrados anteriormente²⁴. Para evitar tal situação, devem ser realizados testes de controlo de qualidade tendo como base os objetos originais e os objetos migrados de modo a verificar se o conteúdo do objeto migrado mantém as características iniciais e, portanto, não sofreu alterações inesperadas durante o processo de transferência.

Neste processo podem também perder-se informações relativamente à estrutura, metainformação ou conteúdo. No entanto, para que esta situação de perda seja minimizada, todas as ações efetuadas no âmbito do processo de migração devem ser bem documentadas para que, no final da ação de migração, seja possível fazer um levantamento da informação eventualmente perdida e situar em que momento do processo tal aconteceu e quais as consequências para a manutenção da

²³ Ver mais considerações acerca da seleção dos formatos dos objetos digitais em Normalização.

²⁴ Esta situação pode ser contrariada através da adoção da estratégia Migração a-pedido.

autenticidade e fidedignidade do objeto digital. Assim, será possível aprender com os erros e evitar determinada ação no futuro.

Não obstante, segundo Thomaz e Soares (2004, p. 7), a migração é a estratégia de preservação digital mais utilizada pelas organizações e para a DGARQ (2011, p. 54) é a “única que tem vindo a dar provas da sua eficácia”.

Para alguns autores, existem estratégias que derivam da migração e, por isso, serão apresentadas de seguida.

7.4.1 – Migração para suportes analógicos

A *Migração para suportes analógicos* consiste na conversão de objetos existentes em meio eletrónico para suportes não digitais, isto é, trata-se da reprodução de objetos digitais para papel, microfilme ou outro tipo de suporte analógico. Este tipo de solução permite preservar os documentos por vários anos, desde que os suportes físicos, também estes, sejam alvo de ações de conservação ao longo do tempo (Thomaz e Soares, 2004, p. 8). No entanto, o novo objeto analógico deixa de estar acessível através das plataformas digitais, o que implica a necessidade de mecanismos que permitam a sua localização. Nesse seguimento, deixa de ser possível a recuperação e acesso imediatos à informação, que agora já nos é habitual graças às novas tecnologias.

Contudo, a utilização de suportes analógicos está limitada: cada vez mais os objetos digitais são compostos por elementos variados, sendo muitos deles interativos ou dinâmicos, o que dificulta ou até mesmo impossibilita a sua reprodução para este tipo de suportes (Pinto, 2009^b, p. 145-146).

Rothenberg (Cit. por Sant’Anna, 2001, p. 130) afirma que a estratégia mais utilizada para evitar a perda de informações, numa eventual ausência do meio tecnológico, é a reprodução em papel. Porém, este tipo de estratégia não se mostra adequada para objetos digitais complexos que façam uso de todas as funcionalidades permitidas pelo mundo tecnológico.

7.4.2 – Normalização

A *Normalização* pretende facilitar o processo de preservação digital, através da definição de formatos de preservação para cada tipo de objeto digital, e deve ser

encarada não como uma estratégia de preservação mas como uma boa prática que auxilia na implementação de estratégias efetivas. A redução e controlo do número de formatos existentes auxilia na preservação, uma vez que permite que a mesma estratégia possa ser aplicada a mais objetos digitais, traduzindo-se assim numa diminuição dos custos e da complexidade de todo o processo. Para que tal seja possível, é fulcral que no momento de criação do objeto digital o formato já esteja definido pois, dessa forma, irão evitar-se conversões posteriores.

A equipa multidisciplinar deve conhecer os formatos disponíveis para os diferentes tipos de objetos digitais existentes na sua organização, uma vez que uma correta definição implica de forma benéfica todo o processo de preservação digital. A escolha dos formatos deve recair sobre aqueles que estão definidos por normas estáveis – cuja responsabilidade cabe a instituições conceituadas a nível internacional –, não proprietários e cujas especificações estejam divulgadas publicamente de forma bem documentada, evitando assim problemas ao nível dos direitos de autor e pagamento de licenças. Devem ser escolhidos formatos com a capacidade de albergar as características essenciais atuais do objeto digital e que permitam uma sustentabilidade a longo prazo.

Para que a organização possa garantir essa sustentabilidade é importante que as normas adotadas sejam constantemente acompanhadas, pois, mesmo se tratando de normas estáveis e de larga aplicação, podem estar sujeitas a modificações que devem ser prontamente observadas para que não prejudiquem o objetivo principal: a preservação a longo prazo dos objetos digitais.

A recomendação de utilização de formatos abertos permite que a normalização difunda também a interoperabilidade entre plataformas tecnológicas. Lee *et al.* (2002, p. 97) consideram que a utilização de formatos baseados em XML seria útil para a resolução do problema da preservação digital.

Sabendo a relevância de uma correta definição de formatos para a preservação a longo prazo de objetos digitais, Adrian Brown (*The National Archives*) elaborou o relatório *Selecting File Formats for Long-Term Preservation* no qual define um conjunto de doze critérios que devem ser considerados na tomada de decisão sobre os formatos a utilizar no momento de criação desses objetos. Esta lista de fatores pode ser analisada de forma mais pormenorizada no Anexo D.

Não obstante, o *Arquivo da Web Portuguesa*²⁵ divulgou uma lista de formatos categorizados de acordo com diferentes tipos de objetos digitais. Esta lista pode ser considerada pela equipa multidisciplinar no momento da decisão dos formatos, uma vez que nela estão refletidos os formatos mais e menos adequados para a preservação. Esta lista pode ser consultada no Anexo E.

7.4.3 – Atualização de versões de formato

A maior parte das aplicações de *software* garantem a retrocompatibilidade dos objetos digitais criados pelas suas versões antecedentes. O facto de a aplicação interpretar um documento num formato mais antigo, faz com que seja possível gravar esse mesmo documento no novo formato produzido por essa aplicação (Ferreira, 2006, p. 37-38). No entanto, não há garantias que não existam perdas nessa atualização de formato.

Thibodeau (2002, p. 23) refere possíveis problemas advindos desta forma de migração: apesar de se utilizar a codificação original do objeto digital, a atualização para a versão mais recente de um formato pode originar características que o objeto não possuía. Esta situação é agravada se forem feitas sucessivas transformações, pois as alterações ao objeto inicial irão se acumular acabando por poder deturpar a sua identidade. Este autor acrescenta ainda outro eventual problema que é a descontinuação definitiva da aplicação e do formato por parte do fabricante, o que evidencia que esta estratégia não deve ser considerada a longo prazo e, conseqüentemente, não deve ser tida como estratégia de preservação por si só.

Contudo, para Pinto (2009^b, p. 146) esta é a ação aconselhada pelo fabricante do *software* e a mais frequentemente usada pelo utilizador comum. Trata-se de uma ação que não requer conhecimentos técnicos especializados e o facto de fazer uso das ferramentas já disponíveis permite reduzir os custos que o plano de preservação digital traz à organização e também atrasar a necessidade de uma estratégia efetiva.

Como simples exemplo, podemos encontrar o Microsoft Word capaz de interpretar objetos com formato DOC e que podem ser gravados para o formato mais recente (DOCX).

²⁵ Disponível em www: <http://www.arquivo.pt/>.

7.4.4 – Conversão para formatos concorrentes

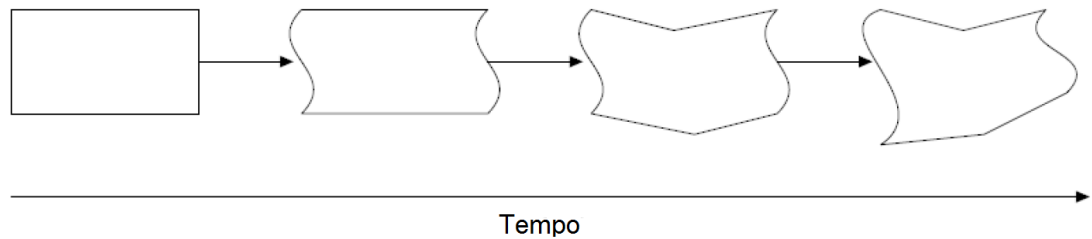
Os fabricantes deveriam assegurar que as características de um dado formato estariam disponíveis nas versões de *software* subsequentes, contudo isso nem sempre acontece. De modo a ultrapassar algumas ruturas tecnológicas – como uma eventual descontinuidade dos formatos, lançamento de novas versões do formato ou do *software* –, os objetos digitais devem ser convertidos para formatos concorrentes, independentemente da aplicação utilizada na sua criação (DGARQ, 2011, p. 55; Ferreira, 2006, p. 38).

A escolha dos formatos de destino deve recair sob formatos normalizados com especificações abertas, não proprietários e devidamente documentados.

O habitual exemplo deste tipo de estratégia é a conversão de ficheiros do Microsoft Word para o formato PDF. Note-se, no entanto, que alguns dos formatos concorrentes podem não manter as mesmas características dos documentos originais. Neste exemplo, um documento em PDF não permite que seja facilmente editado, o que é simples de realizar num documento Word.

7.4.5 – Migração a-pedido

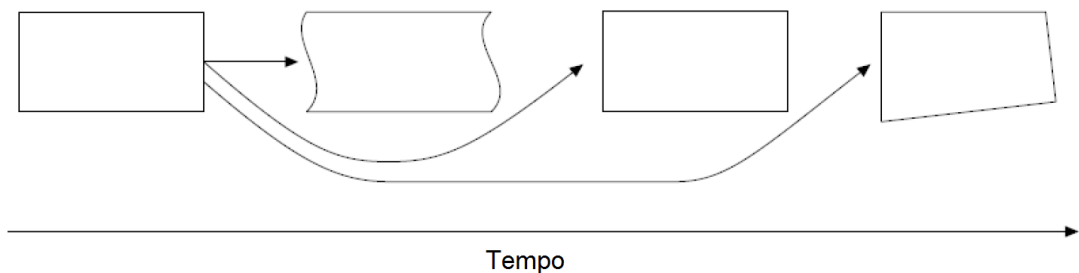
Na *Migração a-pedido* as conversões são feitas com base no objeto digital original, evitando assim a degradação que ocorre quando se realizam ações de migração consecutivas. Sempre que se realiza uma migração ocorrem mudanças ao nível do objeto lógico, o que faz com que o ficheiro transportado possua propriedades diferentes do original. Após sucessivas ações de migração essas diferenças podem ser consideráveis, como é observável na Figura 4 (Mellor, Wheatley e Sergeant Cit. por Ferreira, 2006, p. 40).



Fonte: Ferreira, 2006, p. 40

Figura 4 – Degradação do objeto digital ao longo de sucessivas ações de migração

Na migração a-pedido é possível minimizar as diferenças entre o objeto original e os objetos transportados (Figura 5). Se após uma ação de migração, o resultado tenha sido insatisfatório, a situação pode ser resolvida com a realização de novas conversões fazendo uso de melhores conversores ou de formatos mais adequados.



Fonte: Ferreira, 2006, p. 41

Figura 5 – Migração a-pedido

Esta técnica procura preservar os objetos digitais, enquanto sequências binárias, apenas os transformando no momento em que o seu acesso é solicitado, evitando-se, desta forma, as constantes migrações realizadas a cada mudança tecnológica. A migração a-pedido teve origem nos projetos CEDARS²⁶ e CAMiLEON²⁷

²⁶ “Projecto de preservação digital desenvolvido entre Abril de 1998 e Março de 2002, sob a direcção geral do CURL (*Consortium of University Research Libraries*) e financeiramente apoiado pelo JISC/CEI através do *eLib Programme* tinha por finalidade tratar dos aspectos estratégicos, metodológicos e práticos da preservação digital” (Pinto, 2009^b, p. 275).

²⁷ “Projecto de investigação apoiado pelo JISC, desenvolvido pelas Universidades de Michigan e de Leeds, no qual são desenvolvidas e avaliadas estratégias de preservação digital visando garantir o acesso a longo prazo. Nele são estudados diversos aspectos da emulação de tecnologias como estratégia de preservação digital, pretendendo-se desenvolver ferramentas, directrizes e estudos

e tinha como propósito a redução do número de ações de migração aplicadas aos objetos digitais.

O funcionamento desta estratégia consiste na preservação do objeto digital original e também da sua metainformação associada essencial para descrever a forma como o objeto deve ser transferido para os ambientes digitais contemporâneos (Keefer e Gallart, 2007, p. 86). A *National Library of Australia* (2003, p. 139) acrescenta ainda que o objeto digital e a respetiva metainformação devem estar relacionados, no entanto devem permanecer armazenados separadamente para que a metainformação possa ser acedida e modificada de forma centralizada.

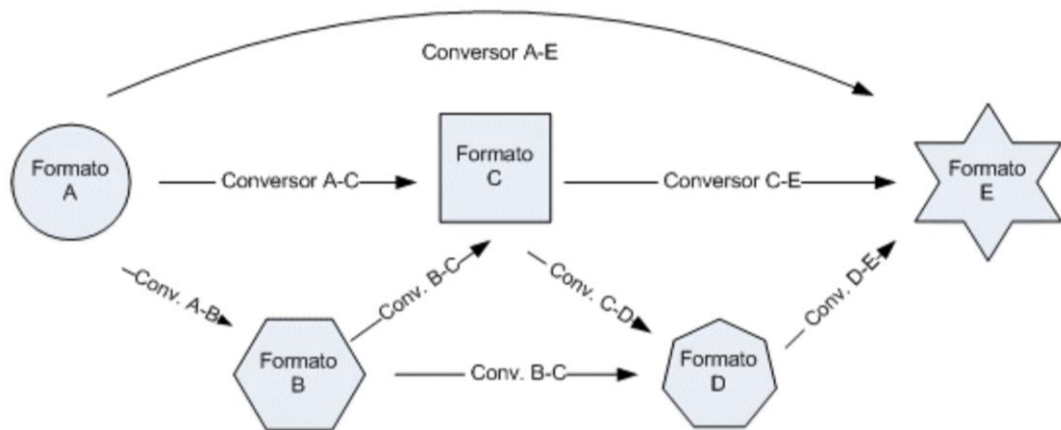
Trata-se então de uma estratégia de preservação orientada para objetos digitais cujo acesso é pouco requisitado. Esta estratégia tem mais probabilidade de sucesso, uma vez que a transformação é feita com base no objeto original evitando assim as distorções resultantes das constantes ações de migração e, conseqüentemente, reduzindo os custos associados a essas ações.

Por outro lado, a migração a-pedido não será uma hipótese se não houver mecanismos para aceder aos meios em que os objetos digitais foram originalmente armazenados.

7.4.6 – Migração distribuída

Graças a estudos desenvolvidos na área da migração surgiram arquiteturas distribuídas de conversores. Esta estratégia consiste num conjunto de serviços de conversão disponíveis na Internet e que podem ser utilizados de forma remota, através do recurso a uma aplicação-cliente (Figura 6) (Ferreira, 2006, p. 41).

de custos de aplicação da emulação comparativamente com outras estratégias de preservação digital. No seu sítio são disponibilizadas ligações e relatórios, nomeadamente, em 2003, o «*Migration on Request – a practical technique for preservation*» (uma estratégia de preservação que visa ultrapassar os problemas da migração) [...]” (Pinto, 2009^b, p. 343).



Fonte: Ferreira, 2006, p. 42

Figura 6 – Migração distribuída baseada em serviços Web

Ferreira (2006, p. 41-42) faz referência a várias iniciativas para o desenvolvimento deste tipo de conversores, nomeadamente:

- O *Typed Objects Model* (TOM)²⁸ é um sistema distribuído de conversores usado na migração de objetos digitais para outros formatos;
- No *Lister Hill National Center for Biomedical Communications* desenvolveu-se um serviço *Web*²⁹ que permite a conversão de objetos digitais de um de cinquenta formatos diferentes suportados para PDF;
- A Universidade do Minho tem desenvolvido uma Arquitetura Orientada ao Serviço³⁰, apelidada de CRiB³¹ (*Conversion and Recommendation of Digital Object Formats*), que disponibiliza serviços de conversão, avaliação e recomendação de estratégias de migração.

Ferreira (2006, p. 42-43) apresenta algumas das vantagens da migração distribuída comparativamente com outras estratégias de migração, sendo elas:

- O facto de existirem vários caminhos de migração reduz o risco de falha do serviço devido a um eventual desaparecimento dos conversores;

²⁸ Até à data da redação desta dissertação não foi possível encontrar uma página *Web* disponível relativa a este projeto.

²⁹ Disponível em www: <http://docmorph.nlm.nih.gov/docmorph/>.

³⁰ Do inglês *Service-Oriented Architecture* (SOA).

³¹ Até à data da redação desta dissertação não foi possível encontrar uma página *Web* disponível relativa a este projeto. Ver mais considerações acerca desta iniciativa em CRiB.

- É compatível com outros tipos de migração, como a normalização e a migração a-pedido;
- Permite reduzir custos com a preservação se existir uma rede global de conversores.

Contudo, este autor refere ainda que o facto de estar em causa uma enorme quantidade de dados digitais pode ser um impedimento para grande parte das empresas devido a não possuírem as condições tecnológicas suficientes para transferir, armazenar e processar tal quantidade.

7.5 – Encapsulamento

O *Encapsulamento* consiste em preservar, conjuntamente com o objeto digital, as informações necessárias (descrição do objeto e das suas propriedades)³² para que, mais tarde, seja possível a construção de conversores, emuladores ou visualizadores (Pinto, 2009^b, p. 147).

Esta técnica de preservação atua apenas no momento em que é solicitado o acesso aos objetos digitais, ao contrário do que acontece na migração onde todos os documentos são alvos de ação. É adequada para objetos cuja previsão de acesso é num futuro longínquo, pois podem passar vários anos até que a comunidade demonstre interesse por um dado conjunto de documentos e se iniciem os esforços necessários para a construção de conversores, emuladores ou visualizadores.

Uma iniciativa no âmbito desta técnica de preservação é o Formato Universal de Preservação³³ (UPF) em que se pretende criar um formato normalizado, independente de plataformas tecnológicas, capaz de incorporar metainformação de preservação junto do objeto digital (Ferreira, 2006, p. 43-44).

Raymond Lorie propôs uma variante da emulação, a Máquina Virtual Universal, que consiste num *software* aplicado a um recurso que gera outro programa executável onde se podem encontrar todos os dados relativos a esse recurso. Tanto o recurso e o executável são preservados em conjunto e para que o executável possa

³² A “cápsula” composta pelo objeto digital e pela respetiva metainformação é considerada como *pacote de informação* no Modelo de referência OAIS.

³³ Do inglês *Universal Preservation Format*.

ser interpretado basta abri-lo recorrendo ao *Universal Virtual Computer* (Pinto, 2009^b, p. 147).

7.6 – Pedra de Rosetta digital

A *Pedra de Rosetta*, encontrada no século XIX, permitiu a Champollion descodificar vestígios deixados pelo povo egípcio. A Pedra de Rosetta consiste num bloco de granito com três línguas gravadas (egípcio hieroglífico, egípcio cursivo e grego clássico). Graças aos seus conhecimentos em grego, Champollion conseguiu decifrar a versão egípcia, o que levou à descodificação de outros artefactos (Ferreira, 2006, p. 44).

Heminger e Robertson sugerem uma estratégia idêntica para recuperar objetos digitais com poucas informações sobre o seu formato. Estes autores propõem que não se preservem as regras que auxiliam na descodificação da informação, mas sim amostras representativas do formato desde que sejam perceptíveis pelo ser humano, tal como as inscrições existentes na Pedra de Rosetta original. Com tais informações seria possível deduzir as regras necessárias para converter o objeto para um formato contemporâneo (Pinto, 2009^b, p. 147).

Um exemplo dado por Thibodeau (2002, p. 24-25) é a existência de uma impressão em papel de um documento de texto e a respetiva representação em código binário. Para fazer a conversão do objeto para um formato contemporâneo será necessário comparar o texto com a sua representação binária.

Esta estratégia³⁴ não é considerada uma estratégia de base mas sim uma medida de arqueologia digital usada em último recurso quando as demais estratégias falham (Heminger e Robertson Cit. por Ferreira, 2006, p. 45).

8 – Metainformação

Como já foi referido neste trabalho, as TIC estão em constante evolução e alguns recursos digitais acabam por se tornar ultrapassados muito rapidamente. Para

³⁴ Esta abordagem pode ser comparada à engenharia reversa de *software*. Consiste em partir do sistema final para as suas especificações, procurando entender e descrever o seu funcionamento e o comportamento dos seus componentes num nível mais alto de abstração. Através desta análise é possível criar novas representações desse sistema (Costa e Sanches, 1996, p. 16).

evitar que se perca o acesso aos objetos digitais devem-se aplicar estratégias de preservação (que foram mencionadas na secção anterior). No entanto, é necessário aprofundar a importância que a metainformação tem para o sucesso de qualquer programa de preservação de informação digital.

Méndez (Cit. por Romero, 2006, p. 129) descreve os metadados como sendo elementos ou estruturas de organização de informação associados a um objeto de informação digital e que permitem a sua classificação, categorização ou descrição. Este autor define a metainformação, de forma sucinta, como a informação estruturada sobre a informação distribuída. Trata-se então de um sistema de informação paralelo, indissociável do objeto digital que deve ser atualizado ao longo do seu ciclo de vida. Graças aos metadados é possível, mais tarde, obter a contextualização desse recurso. Para a *National Information Standards Organization* (NISO) (2004, p. 1), a metainformação é fundamental para garantir que os recursos sobrevivam e estarão acessíveis no futuro.

Algumas aplicações³⁵ já adicionam metadados de forma automática a um recurso digital no momento da sua criação, no entanto, não é suficiente. Por isso, cabe ao criador desse recurso descrever o contexto organizacional e tecnológico, uma vez que se pressupõe que seja a pessoa com mais conhecimento para realizar tal atividade.

Os metadados podem estar embutidos no próprio recurso (no caso de ficheiros em HTML são utilizadas, no cabeçalho, as etiquetas <META>) ou armazenados num outro ficheiro e estarem relacionados com o recurso a que são afetos. No entanto, existem vantagens em mantê-los incluídos ou separados, como referem Borghoff *et al.* (2010, p. 9, tradução nossa):

Os documentos digitais e os seus metadados são fortemente ligados entre si: porque um sem o outro é inútil, a consistência é de extrema importância. Por outro lado, existem boas razões para os armazenar separadamente: entre outras coisas, os metadados são usados para recuperação de documentos. Armazenar a informação catalogada num lugar único pode acelerar o acesso de forma significativa. Mais importante ainda, o documento apenas pode ser

³⁵ Consideremos um documento de texto criado pelo Microsoft Word. Elementos como título, autor, data de criação, data de modificação ou a própria versão da aplicação são automaticamente gravadas no momento da criação do documento. Já outros elementos como etiquetas, comentários ou assunto podem ser editados pelo utilizador.

Mesmo que este documento seja convertido para PDF, os metadados continuam associados e podem ser verificados através da ferramenta de propriedades do documento disponibilizada pela aplicação.

lido após codificação e o formato do documento foi determinado com base nos metadados disponíveis.

Isto é, o facto de os metadados estarem inclusos no objeto evita que haja perda de informação e que exista quebra de ligação entre o recurso e metadados e permite que ambos sejam atualizados em simultâneo. Por outro lado, o facto de os metadados se encontrarem separados do objeto digital permite que sejam administrados e conservados por sistemas de preservação, assegurando assim uma gestão mais simples e eficaz. Ao guardar apenas a metainformação faz com que os ficheiros sejam de tamanho mais reduzido e o seu processamento mais eficiente e, por isso, o processo de pesquisa e recuperação dos objetos digitais é agilizado (Dappert e Enders, 2010, p. 11).

Embora a classificação dos diferentes tipos de metainformação não seja consensual, alguns autores que realizam trabalhos nesta área sugerem três categorias: descritiva, administrativa e estrutural.

Os metadados descritivos permitem procurar, identificar e recuperar um objeto digital através da descrição do seu conteúdo intelectual. Podemos considerar como metadados descritivos os elementos título, resumo, autor ou palavras-chave.

Os metadados administrativos auxiliam na gestão do ciclo de vida dos objetos digitais, prestando informação acerca dos requisitos necessários para manutenção, armazenamento e disseminação. Neste tipo de metainformação podemos encontrar elementos como a data e o propósito da criação, data de modificação, permissões de acesso e informação técnica como o formato do objeto e a sua localização.

Por último, encontramos os metadados estruturais responsáveis por descrever a organização de objetos digitais complexos, isto é, permitem relacionar os diversos elementos existentes que os compõem. São fulcrais para descrever os requisitos tecnológicos necessários para que um objeto digital seja apresentado corretamente em vários sistemas, uma vez que armazenam informação acerca dos seus elementos como tamanho, cor das fontes ou localização de imagens/tabelas. O habitual exemplo dado para este tipo de metadados é a descrição que deve ser feita para um livro, ou seja, é necessário numerar os capítulos e as respetivas páginas também devem ser ordenadas.

Não obstante, emerge uma outra categoria de metainformação: a metainformação de preservação.

8.1 – Metainformação de preservação

A metainformação de preservação tem como objectivo descrever e documentar os processos e actividades relacionadas com a preservação de materiais digitais. Ou seja, a metainformação de preservação é responsável por reunir, junto do material custodiado, informação detalhada sobre a sua proveniência, autenticidade, actividades de preservação, ambiente tecnológico e condicionantes legais.

(Lavoie e Gartner Cit. por Ferreira, 2006, p. 54)

Para Lavoie e Gartner (Cit. por Keefer e Gallart, 2007, p. 140) os metadados de preservação não existem de forma isolada e podem ser considerados como metadados descritivos, estruturais e administrativos (Figura 7). Devem ser vistos como o conjunto de metadados capazes de influenciar a preservação a longo prazo, permitindo a pesquisa, recuperação e utilização de objetos digitais de forma íntegra e autêntica.

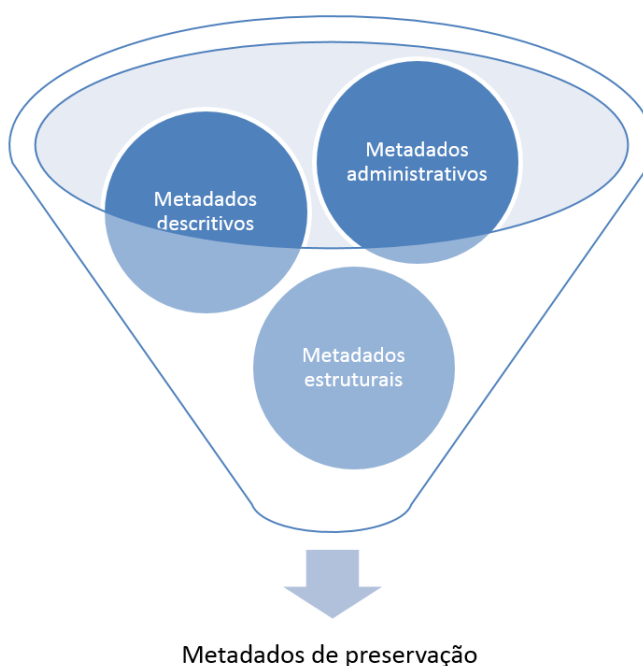


Figura 7 – Metadados de preservação

Os metadados de preservação são um dos pontos fulcrais para o sucesso de qualquer estratégia de preservação aplicada a objetos digitais. Deegan e Turner (2006, p. 22) referem que sejam quais forem as estratégias ou técnicas adotadas para preservar recursos em formato digital, apenas serão bem-sucedidas se esses recursos forem bem documentados ao longo do seu ciclo de vida. Para que uma

estratégia atinja os objetivos para os quais foi delineada necessita de vários elementos de informação que descrevam o processo organizacional para qual esse objeto foi criado, o contexto tecnológico existente, dependências perante outros objetos e modificações ocorridas. Na definição de metadados de preservação supramencionada é possível encontrar as categorias de informação que Lavoie e Gartner (Cit. por Sayão, 2010, p. 11-12) consideram necessárias para a preservação digital:

- Proveniência: devem ser documentadas as informações relativas ao histórico do objeto digital desde o momento da sua criação;
- Autenticidade: devem ser incluídas as informações necessárias para assegurar que o objeto se mantém íntegro e livre de alterações não documentadas;
- Atividades de preservação: devem ser registadas as estratégias de preservação aplicadas ao objeto digital e também os efeitos da sua implementação;
- Ambiente tecnológico: devem estar descritos os requisitos que permitem à plataforma tecnológica – *hardware* e *software* – interpretar o objeto digital;
- Condicionantes legais: devem estar registadas as informações relativas à propriedade intelectual que podem ter influência – no presente e no futuro – nas ações de utilização, disseminação e preservação.

O uso de metadados permite valorizar o objeto digital, mas para o seu uso devem ser cumpridas as recomendações elaboradas por organismos internacionais competentes, uma vez que existem vários esquemas de metainformação orientados para diversas áreas de atuação e para diferentes tipos de recursos. Na fase de elaboração do plano de preservação digital, a equipa multidisciplinar responsável deve ter em conta os vários esquemas existentes e deve ponderar a necessidade de adotar um ou mais esquemas e quais os que melhor se adequam à comunidade de interesse em que a organização está inserida. Dessa deliberação deve resultar a descrição da terminologia a utilizar aquando da criação ou atualização dos metadados dos objetos digitais (Pinto, 2009^b, p. 154).

Ao escolher um esquema de metainformação deve-se ter em conta se esse esquema é suficientemente amplo para que possa suportar requisitos de preservação, no presente e no futuro; se é suficientemente prático de ser implementado; se pode ser integrado com sistemas de gestão de metadados; e se permite interoperabilidade entre diferentes plataformas (Lavoie e Gartner Cit. por Sayão, 2010, p. 12).

Importa ainda referir que, enquanto os esquemas de metadados orientados para a procura de objetos podem ser de imediato testados e os seus conceitos refinados, os esquemas de metadados de preservação são desenvolvidos, implementados e só mais tarde podem ser avaliados. Só quando for necessária, no futuro, a consulta desses mesmos objetos, é que se fica a saber se o esquema de metadados adotado tem (ou teve) condições suficientes para salvaguardar a preservação a longo prazo (Sayão, 2010, p. 13).

8.2 – Esquemas de metainformação

Embora não seja pretensão deste trabalho relatar de forma pormenorizada os esquemas de metainformação mais utilizados, será feita uma breve apresentação dos esquemas referidos pela maior parte dos autores consultados. No Anexo F poderão ser encontrados exemplos de cada um deles.

8.2.1 – Dicionário de Dados PREMIS³⁶

Em 2003 a OCLC (*Online Computer Library Center*) e o RLG (*Research Libraries Group*) formaram o grupo de trabalho PREMIS (*PREservation Metadata: Implementation Strategies*)³⁷ para se debruçar na criação de modelos de metadados orientados para a preservação digital.

Deste grupo de trabalho resultou, mais tarde, um relatório no qual se inclui o Dicionário de Dados PREMIS que define um conjunto de orientações que vão de encontro às necessidades da metainformação de preservação. Trata-se de um vasto guia que define uma coleção de metadados necessários para a preservação digital a longo termo e oferece uma análise das opções existentes para a sua codificação, gestão, armazenamento e permuta.

³⁶ Do inglês *PREMIS Data Dictionary*.

³⁷ Disponível em www: <http://oclc.org/research/activities/pmwg.html>.

O Dicionário de Dados PREMIS foca-se num conjunto básico de elementos que a maior parte dos repositórios deve compreender para poder apoiar a preservação por um longo período. Este esquema não particulariza a forma como a metainformação deve ser representada no sistema, apenas define o que o sistema deve interpretar e ser capaz de exportar. Isto inclui metainformação administrativa, mas também metainformação técnica genérica partilhada por todos os tipos de conteúdo.

O uso dos esquemas XML que acompanham o Dicionário de Dados é opcional, no entanto podem simplificar a implementação e a troca de informação de preservação (Deegan e Tanner, 2006, p. 73).

Este modelo está organizado em torno de cinco entidades (PREMIS Editorial Committee, 2012, p. 5-7):

- **Objetos (*objects*):** unidade discreta de informação digital. Podem ser subdivididos em:
 - **Ficheiros (*files*):** sequência binária reconhecida pelo *software*. Possui determinadas características como formato, permissões de acesso ou informações técnicas (tamanho, data de criação ou data de modificação);
 - **Fluxo de *bits* (*bitstream*):** conjunto de dados existentes dentro de um ficheiro com propriedades significativas para preservação;
 - **Representação (*representation*):** conjunto de ficheiros e metadados estruturais necessários para apresentar, na íntegra, uma entidade intelectual;
- **Entidades intelectuais (*intellectual entity*):** conjunto de conteúdos definidos como uma unidade (por exemplo, um livro);
- **Agentes (*agents*):** entidades (pessoas, organizações ou programas de computador) que podem estar associadas a eventos ou direitos;
- **Eventos (*events*):** estruturas que armazenam metainformação relacionada com ações realizadas sob os objetos ao longo do seu ciclo de vida;
- **Direitos (*rights*):** permissões legais relacionadas com os objetos ou agentes.

Devido à grande quantidade de informação existente num repositório, o grupo de trabalho PREMIS aconselha a criação e utilização de metainformação gerada automaticamente e, também, o uso de vocábulos controlados para que se possa garantir a interoperabilidade entre sistemas (Campos e Saramago, 2007, p. 5).

8.2.2 – Dublin Core

Trata-se de um esquema recomendado pela *Dublin Core Metadata Initiative* (DCMI)³⁸, cuja ambição inicial consistia em desenvolver um conjunto extensível e básico de elementos de metainformação capazes de descrever um documento baseado na *Web*.

Já em 1995, num *workshop* organizado pela OCLC e o NCSA (*National Center for Supercomputing Applications*) em Dublin (Ohio), se discutia a necessidade de desenvolver métodos que permitissem a descrição, organização, recuperação e acesso de recursos digitais localizados na *Web*. É como resultado deste *workshop* que surge a DCMI e é no ano de 1999 que surge a publicação do conjunto de quinze elementos propostos por este grupo de trabalho. Em 2001, o esquema de metainformação Dublin Core tornou-se um padrão internacional – ISO 15386.

Os principais objetivos deste consórcio são a formalização e simplificação da descrição dos elementos face à proliferação de recursos digitais e à frágil catalogação dos recursos; a extensibilidade; a independência da plataforma tecnológica; e, ainda, a reutilização dos padrões de metainformação já existentes (Borghoff *et al.*, 2010, p. 136).

Este esquema de metainformação possui quinze elementos amplos e genéricos, capazes de descrever um vasto leque de recursos. Cada elemento é opcional, flexível e pode ser repetido, são eles (NISO, 2013, p. 2-5):

- Título (*title*): o nome dado ao recurso;
- Criador (*creator*): entidade responsável pela criação do recurso (pode ser uma pessoa, uma organização ou um serviço);
- Assunto (*subject*): assunto relativo ao conteúdo do recurso (geralmente representado por palavras-chave e recomenda-se o uso de vocabulário controlado);

³⁸ Disponível em www: <http://dublincore.org/>.

- Descrição (*description*): descrição detalhada do conteúdo do recurso (pode incluir um resumo, uma tabela de conteúdos ou uma representação gráfica, entre outros);
- Editor (*publisher*): entidade responsável por disponibilizar o recurso (pode ser uma pessoa, uma organização ou um serviço);
- Outros responsáveis (*contributor*): entidade responsável por realizar contribuições ao conteúdo do recurso (pode ser uma pessoa, uma organização ou um serviço);
- Data (*date*): data associada a um evento no ciclo de vida do recurso (pode ser a data de criação ou a data em que foi aplicada uma estratégia de preservação);
- Tipo (*type*): natureza ou género do recurso (deve ser utilizado vocabulário controlado);
- Formato (*format*): informação relativa ao formato do ficheiro, meio físico ou dimensões do recurso (deve ser utilizado vocabulário controlado);
- Identificador (*identifier*): referência não ambígua do recurso (recomenda-se a identificação do recurso através de uma *string* em conformidade com um sistema de identificação formal);
- Fonte (*source*): referência a um recurso do qual deriva o atual (pode derivar na totalidade ou em parte);
- Língua (*language*): língua em que se encontra o conteúdo intelectual do recurso (deve ser utilizado vocabulário controlado);
- Relação (*relation*): um recurso relacionado;
- Cobertura (*coverage*): âmbito espacial, temporal ou administrativo do recurso;
- Direitos (*rights*): informação relativa aos direitos legais do recurso (geralmente inclui declarações associadas aos direitos de propriedade do recurso, nomeadamente direitos relativos à propriedade intelectual).

O esquema de metainformação Dublin Core pode ser codificado em RDF (*Resource Description Framework*)³⁹: um dialeto XML recomendado pelo consórcio W3C para codificar, permutar e reutilizar metadados normalizados. O RDF permite ao

³⁹ Disponível em www: <http://www.w3.org/RDF/>.

software descobrir informação acerca de um objeto digital, ou seja, propicia a descoberta de recursos, a catalogação de conteúdos e, também, os direitos legais associados. O facto de se utilizar a linguagem XML facilita a interoperabilidade entre sistemas.

Embora este esquema tenha sido desenvolvido com o objetivo de descrever recursos da *Web*, o facto de ser simples e conciso levou a que a sua utilização se tenha alargado a outras comunidades, não estando limitado a um tipo de objeto em particular ou domínio (NISO, 2004, p. 3).

8.2.3 – METS

O esquema de metainformação METS (*Metadata Encoding and Transmission Standards*)⁴⁰ foi criado pela *Digital Library Foundation* e tem sido mantido pela *Network Development and MARC Standards Office of the Library of the Congress*.

Este esquema surgiu da necessidade existente em garantir uma gestão eficaz de objetos digitais complexos. Vejamos os exemplos: do processo de digitalização de um livro resulta um determinado número de páginas; uma dissertação, para além do seu texto, pode também conter imagens ou tabelas. Para que todos os objetos possam ser representados corretamente nas suas respetivas entidades intelectuais, devem ser geridos de forma consistente ao longo do seu ciclo de vida. Para que se possa garantir a coerência entre elemento e objeto digital (na sua totalidade) deve ser utilizada metainformação específica. O METS auxilia na gestão de objetos digitais que pertençam a uma determinada entidade intelectual (Borghoff *et al.*, 2010, p. 142).

Trata-se de um esquema XML que permite expressar e agrupar metadados descritivos, administrativos e estruturais essenciais para os processos de gestão e intercâmbio de objetos digitais entre plataformas, e, também, descrever a ligação entre estes metadados dentro de um repositório. Wendler (2006, p. 74) considera o METS como uma espécie de mapa para o objeto digital, dada a variedade de metainformação suportada e ainda a capacidade de encapsular o conteúdo e os metadados, bem como direcionar para conteúdo e metainformação externos.

⁴⁰ Disponível em www: <http://www.loc.gov/standards/mets/>.

Um documento METS, dependendo do seu âmbito, pode ser utilizado para estruturar pacotes de informação do modelo de referência OAIS⁴¹. Um documento METS contém sete secções (NISO, 2004, p. 5):

- Cabeçalho METS (*METS header*): contém metainformação que descreve o próprio documento, nomeadamente o nome do autor, data de criação ou de modificação;
- Metainformação descritiva (*descriptive metadata*): pode apontar para metainformação descritiva incluída no próprio documento ou localizada no exterior;
- Metainformação administrativa (*administrative metadata*): fornece informação acerca da criação, do armazenamento, da propriedade intelectual dos objetos digitais e da origem dos elementos que o compõe;
- Secção de ficheiros (*file section*): refere todos os ficheiros que façam parte das diversas versões de um objeto digital;
- Mapa estrutural (*structural map*): define a estrutura hierárquica de um objeto digital, permitindo navegação, e relaciona os elementos dessa estrutura com os respetivos ficheiros e metadados;
- Ligações estruturais (*structural links*): permite hiperligações entre diferentes nós na hierarquia representada pelo mapa estrutural;
- Comportamento (*behavior*): documenta comportamentos executáveis associados ao conteúdo de um documento METS.

O METS pode ser adaptado a várias necessidades e plataformas tecnológicas, e demonstra também ser uma solução para a estabilidade requerida pela preservação digital. O facto de um documento METS estar descrito em XML garante a interoperabilidade e não será afetado pela obsolescência tecnológica.

Em suma, é possível afirmar que o METS visa garantir a coerência lógica e física de entidades digitais (Borghoff *et al.*, 2010, p. 143).

⁴¹ Ver mais considerações acerca desta especificação em Modelo de referência OAIS.

9 – Iniciativas

Algumas organizações têm desenvolvido estudos no que concerne à temática da preservação digital, tentando estabelecer normas e modelos para uma gestão eficaz da informação procurando assim melhorar a sua recuperação e acesso por parte das comunidades-alvo.

Muitas destas iniciativas relacionam-se com a criação e definição de repositórios institucionais, no entanto o facto de uma organização implementar um repositório não é suficiente para assegurar o acesso dos recursos ao longo do tempo: estes recursos encontram-se armazenados mas continuam sujeitos aos efeitos nefastos que surgem como consequência da evolução da tecnologia.

O modelo de referência OAIS (*Open Archival Information System*) é a especificação mais comumente seguida por diversas instituições internacionais no âmbito da criação e gestão de um sistema de arquivo com fins de preservação. Nesse sentido, será feita uma abordagem a este modelo e também a alguns dos muitos projetos que nele se baseiam.

Contudo, também podemos considerar o MoReq2010 (*Modular Requirements for Records Systems*) como uma especificação mas, por sua vez, orientada para a criação de sistemas de gestão de documentos de arquivo ao longo de todo o seu ciclo de vida. Ao contrário do que acontece com o modelo de referência OAIS – onde a informação de arquivo se encontra em repositórios centrais geridos por sistemas especializados –, no MoReq2010 procura-se manter e administrar a informação de arquivo nos sistemas de negócio, não sendo, por isso, necessário a duplicação dos registos (APDSI, 2014, p. 74).

9.1 – Modelo de referência OAIS

O modelo de referência OAIS (*Open Archival Information System*) foi desenvolvido pelo *Consultative Committee for Space Data Systems* (CCSDS) em parceria com a *International Organization for Standardization* (ISO). A parceria entre estas duas instituições surge, em 1990, com o objetivo de desenvolver normas que permitissem gerir o armazenamento, a longo prazo, de objetos digitais. Apenas em 2003 é que o OAIS foi aprovado como norma internacional (ISO 14721:2003). Em 2012 foi feita a sua revisão dando origem à norma ISO 14721:2012.

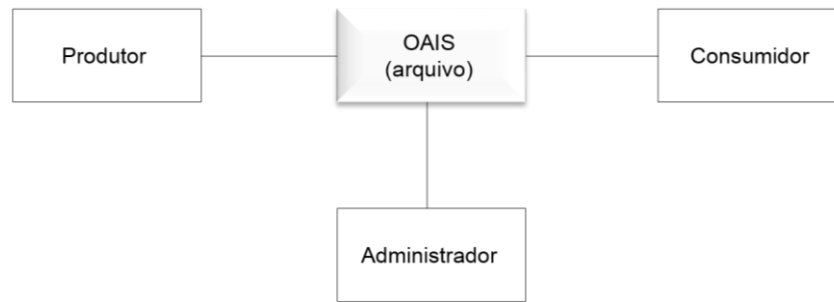
Segundo o CCSDS (2012, 1-1), um arquivo que siga o modelo OAIS consiste num conjunto de pessoas ou sistemas cuja responsabilidade é preservar informação, tornando-a disponível à sua comunidade de interesse.

O modelo OAIS pode abranger quer o armazenamento de objetos físicos como digitais, embora seja mais focado para o mundo digital. Tem sido um modelo bem acolhido por diversas comunidades cujo interesse passa também pela preservação de material digital a longo termo.

Ainda de acordo com o CCSDS, este modelo tem como objetivos: aumentar a compreensão e sensibilização de conceitos necessários para a preservação de informação digital, sobretudo em instituições não arquivísticas, para que também possam estar habilitadas a participar no processo de preservação; definir léxico capaz de descrever e comparar arquiteturas, estratégias e técnicas de preservação; aumentar o consenso relativamente ao processo de preservação; e orientar a identificação e desenvolvimento de normas relacionadas com o OAIS.

Trata-se de um esquema conceptual que procura identificar os componentes necessários para a manutenção de um sistema orientado para a gestão e preservação do acesso continuado à informação. O modelo de referência OAIS define duas estruturas: um modelo funcional e um modelo de informação. O modelo funcional expressa as atividades que um OAIS⁴² deve ser capaz de realizar, enquanto o modelo de informação define as informações, representadas por metadados, necessárias para gerir de forma eficaz as ações de preservação (CCSDS Cit. por Sayão, 2010, p.14). Na Figura 8 é possível verificar as entidades participantes no modelo de referência OAIS:

⁴² Por abuso de linguagem, por vezes será usado “um OAIS” ou “um arquivo OAIS” como forma de se referir um arquivo que siga as normas/sugestões propostas pelo modelo de referência OAIS.



Fonte: CCSDS, 2012, p. 2-2

Figura 8 – Interação entre entidades externas e o arquivo OAIS

Podemos encontrar quatro entidades no modelo de referência OAIS: Produtor, Administrador, Consumidor e o Arquivo. O Produtor é a entidade externa (pessoas ou organizações) que faculta a informação a preservar. O Administrador é responsável por definir e monitorizar as políticas gerais do arquivo. O Consumidor consulta o arquivo e acede à informação que nele se encontra armazenada. Existe uma subclasse de Consumidores relevante, que deve ser destacada: a comunidade-alvo, que deve ser capaz de interpretar a informação de forma independente, ou seja, presume-se que tenha conhecimentos suficientes para compreender um dado recurso sem necessitar de recorrer ao Produtor.

É de salientar ainda que o CCSDS (2012, p. 3-1) considera que um arquivo, segundo o modelo OAIS, tem também responsabilidades a seu cargo, nomeadamente: negociar e aceitar informações dos produtores; manter o controlo da informação de modo a assegurar a preservação a longo termo; determinar as comunidades-alvo que serão capazes de compreender os objetos armazenados; garantir que a informação a preservar seja compreendida pela comunidade-alvo sem precisar da assistência dos produtores; cumprir as políticas e procedimentos de forma a garantir que a informação seja preservada independentemente de qualquer eventual situação que possa ocorrer; e disponibilizar a informação preservada à comunidade de interesse, permitindo que seja difundida através de cópias autênticas.

Toda a informação submetida por um Produtor a um arquivo OAIS, bem como as ações de gestão, armazenamento e preservação dentro do próprio arquivo e a difusão por parte do Consumidor são feitas através de *Pacotes de Informação*. O CCSDS (2012, p. 2-5, 2-6, 4-29, 4-30) define Pacote de Informação como uma entidade lógica composta por dois tipos de informação:

- Informação de Conteúdo (*Content Information*): informação que o repositório deve preservar (objeto digital) em conjunto com a informação de representação (informações necessárias para a correta interpretação da sequência de *bytes* que compõe o objeto depositado junto da comunidade de interesse);
- Informação de Descrição de Preservação (*Preservation Description Information*): informação necessária para gerir a preservação do objeto depositado a que está associada. Este tipo de informação foca-se na descrição do passado e presente da Informação de Conteúdo, procura garantir a sua correta identificação e assegurar que esta não tenha sido alterada indevidamente. Pode ser subdividida em cinco tipos:
 - Informação de referência (*reference information*): identifica e descreve os mecanismos usados para referenciar a Informação de Conteúdo, de modo a que possa ser localizada dentro e fora do repositório de forma unívoca;
 - Informação de contexto (*context information*): documenta as relações entre o objeto depositado e o seu ambiente, nomeadamente a razão da sua criação, relação com outros conteúdos ou dependências técnicas;
 - Informação de proveniência (*provenance information*): documenta o ciclo de vida do objeto armazenado por forma a garantir a sua integridade e autenticidade aos seus utilizadores futuros. Descreve a sua fonte, alterações ocorridas, historial de custódia⁴³ e ações de preservação aplicadas bem como as respetivas consequências;
 - Informação de fixidade (*fixity information*): documenta os mecanismos de autenticação disponibilizados que permitem garantir que não existiram modificações não relatadas;
 - Informação sobre os direitos de acesso (*access rights information*): identifica as restrições de acesso aos objetos

⁴³ Custódia refere-se à “responsabilidade jurídica de guarda e proteção de arquivos, independentemente de vínculo de propriedade” (Arquivo Nacional, 2005, p. 62).

armazenados definidas no momento da sua submissão no arquivo.

Para que a preservação de um Pacote de Informação seja possível são, ainda, necessários mais dois tipos de informações:

- Informação de empacotamento (*packaging information*): agrupa e identifica o objeto de informação e a respetiva metainformação numa única unidade lógica;
- Descrição do pacote (*package description*): metainformação que identifica e documenta os objetos de informação inseridos no depósito, e que facilita os processos de pesquisa e recuperação. Geralmente deriva da Informação de Conteúdo ou da Informação de Descrição de Preservação.

Os Pacotes de Informação podem assumir três diferentes tipos: Pacote de Informação de Submissão⁴⁴ (PIS), Pacote de Informação de Arquivo⁴⁵ (PIA) e Pacote de Informação de Disseminação⁴⁶ (PID).

O PIS consiste no par composto pelo objeto digital e a respetiva metainformação, tal como o Produtor o envia para depósito. O PIA é o recurso depositado, armazenado sob forma de uma sequência de *bytes* e com as alterações necessárias para a sua gestão e armazenamento. É o pacote de informação fulcral para a preservação a longo prazo, uma vez que é nele que se encontram os metadados essenciais para que essa sequência de *bytes* se possa, mais tarde, converter num ficheiro compreensível e utilizável. Por fim, o PID é o recurso tal como é entregue ao Consumidor. Note-se que o PID pode consistir numa versão diferente da contida no PIS por já ter sido submetida a ações de preservação. Tal como ocorre no PIS, também o PID contém metainformação a par do recurso propriamente dito, para que este possa ser devidamente interpretado.

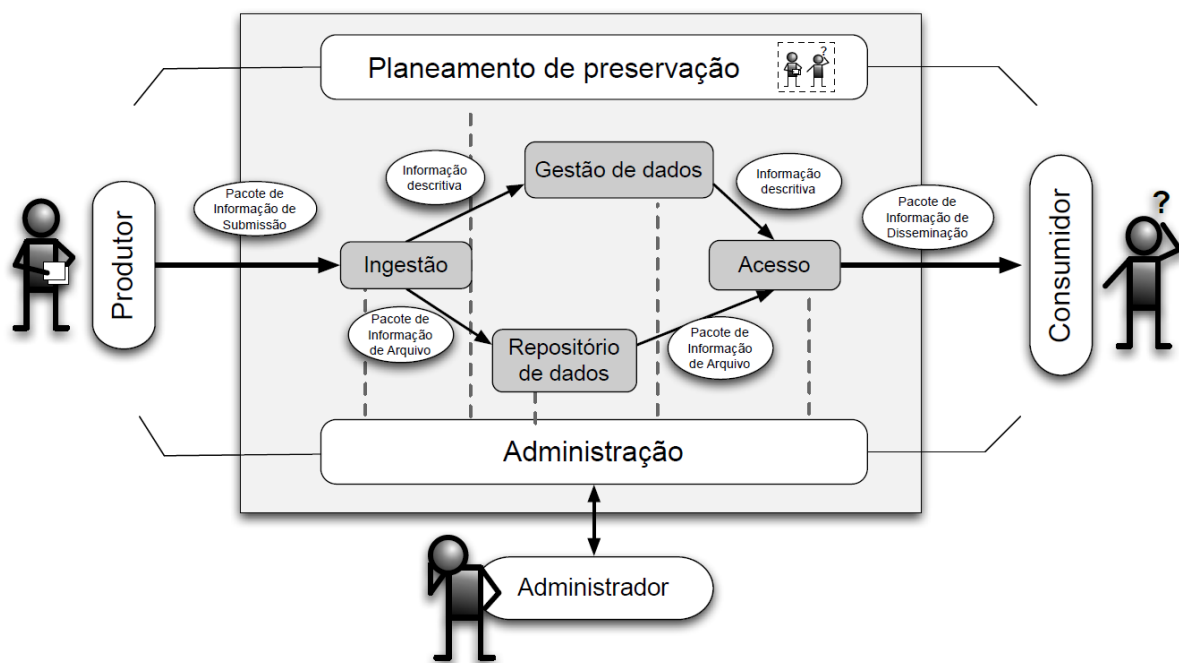
Na Figura 9 é possível observar as entidades internas que compõem um OAIS e os Pacotes de Informação, referidos nos parágrafos anteriores, que são administrados no seu interior. Então, um arquivo OAIS é composto por seis

⁴⁴ Do inglês *Submission Information Package*.

⁴⁵ Do inglês *Archival Information Package*.

⁴⁶ Do inglês *Dissemination Information Package*.

componentes funcionais, sendo que quatro deles são relativos às atividades a que o material depositado se submete (Ingestão, Gestão de dados, Repositório de dados e Acesso), enquanto os outros dois refletem o funcionamento do repositório (Administração e Planeamento de preservação).



Fonte: Ferreira, Saraiva e Rodrigues, 2006, p. 28

Figura 9 – Modelo funcional do OAIS

O processo de submissão da informação tem início na Ingestão, que fornece os serviços necessários para a aceitação do PIS submetido pelo Produtor e prepara os conteúdos para serem geridos e armazenados no arquivo. Caso a Informação descritiva, que acompanha o objeto, não tenha sido inserida pelo Produtor, cabe ao sistema gerá-la internamente. A Ingestão é responsável, também, por transformar o PIS em PIA para poder ser armazenado no Repositório de dados. Extrai, ainda, Informação descritiva dos PIA enviando-a para a Gestão de dados.

O Repositório de dados fornece os serviços necessários para armazenar, gerir e recuperar os PIA. Esta entidade recebe os PIA vindos da Ingestão e adiciona-os para armazenamento permanente; realiza ações de preservação através da atualização de formatos ou suportes; executa procedimentos de rotina e verificação de erros; e fornece meios de recuperação de desastres. Este componente é ainda

capaz de facultar as funcionalidades essenciais ao processo de recuperação dos PIA, de modo a cumprir os pedidos dos Consumidores.

A Gestão de dados preenche, gere e permite o acesso quer à informação descritiva como a dados administrativos utilizados para gerir o arquivo. Estes dados são armazenados numa base de dados cuja responsabilidade de administração é, também, desta entidade, nomeadamente: a realização de atualizações sempre que exista adição ou alteração a nível dos metadados, a realização de consultas e a produção de respostas a essas solicitações.

A Administração é responsável pelo funcionamento geral do sistema de arquivo, o que inclui a negociação dos acordos de submissão com os Produtores; a realização de auditorias para garantir que os objetos depositados cumprem os requisitos exigidos; a gestão da plataforma tecnológica do sistema; e a disponibilização de serviços de apoio aos Consumidores. Cabe-lhe ainda a monitorização e análise das operações realizadas no sistema de modo a melhorar o seu desempenho bem como a decisão de atualizar ou migrar o conteúdo do depósito. Por fim, resta referir que é também responsável por estabelecer e manter as políticas do arquivo.

Quanto ao Planeamento de preservação, é responsável por todas as tarefas necessárias para garantir que os objetos depositados no arquivo se mantêm acessíveis pela comunidade-alvo, mesmo que o ambiente digital original já se tenha, entretanto, tornado obsoleto. Esta entidade analisa o conteúdo do arquivo, efetua recomendações de estratégias de preservação, com base na sua análise das alterações no ambiente tecnológico e das modificações no contexto da comunidade-alvo, e elabora relatórios periódicos com análise de riscos. Para além de efetuar as recomendações de estratégias é também responsável por as implementar.

A última entidade funcional é o Acesso: trata-se da interface de comunicação entre o OAIS e o Consumidor que auxilia este último na pesquisa e recuperação de informação depositada no arquivo, graças à informação descritiva e aos PIA. Recebe os pedidos por parte dos Consumidores; verifica se as informações solicitadas não são protegidas e se podem ser acedidas por aquele determinado Consumidor; processa os pedidos efetuados; e responde às solicitações através da entrega do respetivo PID ao Consumidor.

Falta mencionar que o modelo de referência OAIS consiste numa descrição de alto nível, graças à representação do fluxo de informação dentro destas seis

entidades funcionais, desde o momento de submissão do objeto até à sua posterior utilização, não referindo limitações a nível de recursos que podem ser submetidos a depósito, nem das especificações tecnológicas que devem ser utilizadas para garantir a preservação a longo prazo.

9.1.1 – DSpace

O DSpace⁴⁷ é um *software* de código aberto que tem como objetivo a implementação de repositórios institucionais e está disponível gratuitamente. Foi lançado em 2002 como resultado de uma parceria entre o *Massachusetts Institute of Technology* (MIT) *Libraries* e a *Hewlett-Packard Labs*. O seu objetivo inicial era a implementação de repositórios orientados a receber recursos produzidos pela investigação científica ou destinados ao ensino. Neste momento, a organização que providencia o suporte deste sistema é a *DuraSpace Foundation*.

O DSpace encontra-se em conformidade com o Modelo de referência OAIS e com o protocolo OAI-PMH⁴⁸. O esquema de metainformação predefinido é o Dublin Core.

Esta plataforma permite recolher, preservar e disseminar qualquer tipo de ficheiro de uma organização: embora o DSpace seja reconhecido por acolher recursos académicos baseados em formato de texto (como comunicações, artigos, dissertações ou teses), há também quem o utilize para ficheiros multimédia. O DSpace fornece um amplo conjunto de ferramentas para a gestão do repositório e dos objetos digitais armazenados.

Para o Repositório Científico de Acesso Aberto de Portugal (RCAAP), o sistema DSpace tem como vantagens: a publicação rápida de objetos digitais; a indexação dos recursos, sempre que possível, para que as pesquisas nos motores de busca sejam otimizadas, permitindo que os recursos submetidos cheguem a mais indivíduos; o armazenamento dos objetos digitais num único local; a gestão dos recursos submetidos, isto é, permite saber quantos documentos foram publicados e a quantidade de consultas efetuadas; e a identificação única de cada documento.

⁴⁷ Disponível em www: <http://www.dspace.org/>.

⁴⁸ Do inglês *Open Archives Initiative Protocol for Metadata Harvesting*. Consiste num protocolo que permite a interoperabilidade entre repositórios, através da recolha da metainformação existente.

Contudo, o DSpace (DuraSpace, 2014^a) não garante, por si só, uma estratégia de preservação digital. É importante não esquecer que uma estratégia de preservação deve implicar a criação de cópias de segurança dos documentos ou a definição de meios para prevenção em caso de desastre. É neste sentido que esta plataforma está apta a auxiliar as organizações na questão da preservação.

O DSpace identifica dois tipos de preservação: a preservação da sequência de *bytes* e a preservação funcional. Na primeira, o ficheiro submetido permanece inalterado ao longo do tempo (a sequência binária permanece igual), apesar das mudanças tecnológicas ocorridas. Já na preservação funcional o documento é modificado continuamente ao longo do tempo, mantendo-se acessível tal como no momento de submissão apesar da evolução tecnológica. Podemos assumir o TIFF (*Tagged Image File Format*) e o XML como exemplos de formatos que podem ser funcionalmente preservados através de ações de migração.

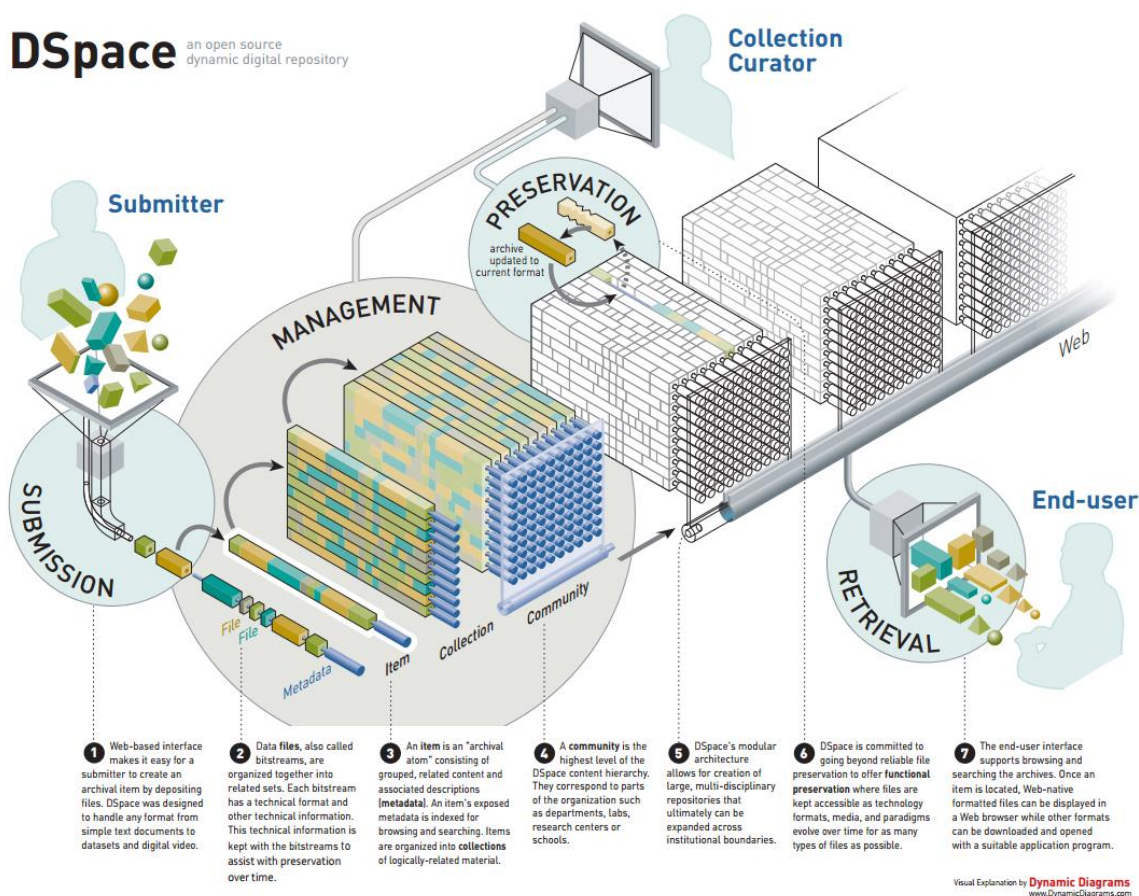
No entanto, nem todos os formatos podem usufruir da preservação funcional, uma vez que cada organização faz uso das aplicações que mais se adequam ao seu propósito, o que leva a que se utilizem vários tipos de formatos, proprietários ou não proprietários, mais ou menos conhecidos. Tendo isso em consideração, o DSpace permite que uma organização defina a sua capacidade de preservar o conteúdo desse formato através da escolha entre três níveis de preservação:

- Formato suportado: formato que pode ser funcionalmente preservado através de ações de migração ou emulação, como o TIFF, XML ou PDF;
- Formato conhecido: tipo de formato em que não se pode dar garantias de que seja preservado, como é o caso dos formatos proprietários (por exemplo, DOC ou PPT). No entanto, o facto de ser amplamente utilizado aumenta a probabilidade de que existam ferramentas de migração desenvolvidas por terceiros;
- Formato não suportado: formato do qual não existe informação suficiente, o que inviabiliza a realização da preservação funcional.

Na Figura 10 é possível observar a estrutura funcional do DSpace. Esta é composta, fundamentalmente, por comunidades, coleções, itens e ficheiros.

Uma comunidade é a entidade mais abrangente na hierarquia do DSpace. Representa a divisão de uma instituição em diferentes partes, como escolas,

departamentos ou grupos de investigação. Uma comunidade é composta por coleções. Uma coleção consiste na agregação de conteúdos relacionados e pode estar presente em diferentes comunidades. Por sua vez, uma coleção reúne um conjunto de itens. Já um item consiste no conjunto da representação do ficheiro e respetiva metainformação, sendo que a última se encontra indexada, de forma a permitir a identificação do objeto. Um item pode existir em diferentes coleções, sendo apenas uma delas a sua proprietária, e pode conter vários pacotes de ficheiros. Por fim, um ficheiro tem associado um formato e também informação técnica.



Fonte: DSpace [Em linha]

Figura 10 – Funcionamento da plataforma DSpace

Na plataforma DSpace os utilizadores apenas podem aceder, consultar e adicionar recursos às coleções a que tenham acesso. Contudo, se um utilizador estiver envolvido em várias coleções, as suas permissões em cada uma delas pode diferir. No DSpace consideram-se três tipos de utilizadores: *Submitter*, *Collection*

Curator e End-user. O *Submitter* é aquele que contribui, adicionando documentos às coleções a que está associado. O *Collection Curator* é responsável por administrar uma ou mais coleções, assegurando que os itens nelas armazenados sejam preservados e possam ser acedidos no futuro. Já o *End-user*, ou utilizador final, pode procurar e recuperar documentos inseridos numa determinada coleção, desde que tenha permissões de acesso para tal efeito (Pinho, 2012, p. 16).

O DSpace baseia-se numa estrutura modular, o que permite a criação de repositórios com mais funcionalidades e com melhores apetências para realizar as funções de armazenamento e difusão. O facto de ser completamente personalizável faz com que possa atender às necessidades informacionais de qualquer organização. Este sistema é amplamente utilizado, nomeadamente, por instituições académicas, governamentais, privadas e comerciais.

9.1.2 – Fedora

O Fedora (*Flexible Extensible Digital Object Repository Architecture*)⁴⁹ foi desenvolvido pela *University of Virginia Library* e pela *Cornell University* e, neste momento, encontra-se, também, sob a alçada da *DuraSpace Foundation*. Tal como o DSpace, trata-se de um sistema de código aberto, modular, e que está em conformidade quer com o Modelo de referência OAIS quer com o protocolo OAI-PMH.

Originalmente, o objetivo desta plataforma era o desenvolvimento de uma arquitetura com capacidade de armazenar, gerir e aceder a conteúdo digital. O Fedora pode ser utilizado como um sistema de repositório autónomo, no entanto foi concebido com o objetivo de ser integrado como um componente de um outro qualquer sistema. Dada a sua flexibilidade, esta plataforma oferece escalabilidade mas, também, durabilidade da informação uma vez que não está dependente de *software* específico. Na maioria dos casos, o Fedora é parte de uma solução que pode incluir outras funcionalidades como motores de busca, gestão de fluxos e mecanismos de segurança.

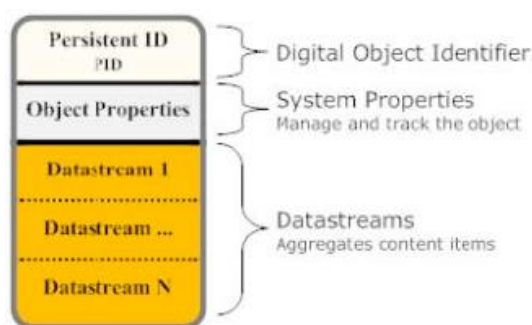
O Fedora fornece ferramentas para a inserção, gestão e disseminação de objetos digitais dentro de um repositório. Esta plataforma assume a existência de objetos digitais complexos e permite efetuar agregações de objetos de diversas fontes

⁴⁹ Disponível em www: <http://www.fedora-commons.org/>.

(que podem estar dentro do repositório ou localizadas na *Web*), por exemplo: um objeto digital – no seu todo – pode ser um conjunto de imagens e texto. Permite, ainda, associar determinados serviços a objetos digitais.

Um objeto num repositório Fedora descreve o conteúdo (dados e metadados) e um conjunto de serviços ou comportamentos associados a esse conteúdo. Na Figura 11 é possível observar a estrutura de um objeto digital segundo a plataforma Fedora e que é codificada por um esquema XML:

- *Digital Object Identifier* (DOI): consiste num identificador único e persistente para o objeto digital;
- *System Properties*: propriedades definidas pelo sistema e necessárias para gerir e localizar o objeto dentro do repositório;
- *Datastreams*: representa o conteúdo de um objeto digital (imagens, textos, gravações áudio), incluindo metainformação (exceto metainformação de sistema).



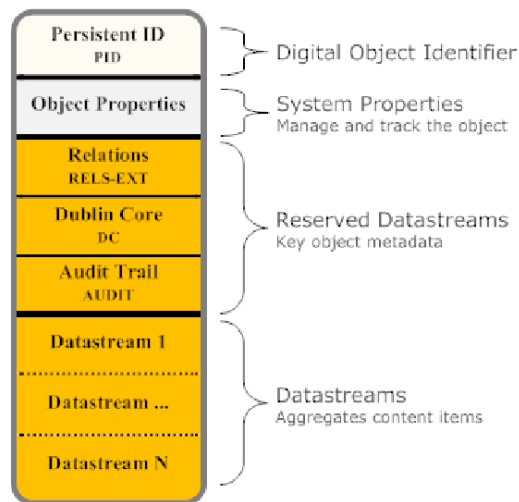
Fonte: DuraSpace, 2011 [Em linha]

Figura 11 – Modelo de um objeto digital de acordo com o Fedora

Um objeto digital no Fedora pode ser composto por vários *datastreams*, aos quais são atribuídos identificadores únicos. No entanto, o Fedora reserva três *datastreams* utilizados para caracterizar o objeto digital (Figura 12):

- RELS-EXT: armazena a descrição dos relacionamentos com outros objetos digitais;
- DC: armazena metadados sobre aquele objeto, segundo o esquema definido pelo Dublin Core. No caso de não serem inseridos metadados, estes serão gerados automaticamente;

- **AUDIT:** privado ao sistema, armazena todos os registos das alterações efetuadas aos objetos.



Fonte: DuraSpace, 2011 [Em linha]

Figura 12 – Especificação dos *datastreams* reservados

Esta plataforma permite realizar a gestão de versões dos objetos digitais. Ou seja, sempre que exista uma nova versão de um objeto, a informação acerca dessa alteração e do respetivo responsável é registada no *datastream* de *Audit Trail* que documenta cronologicamente as modificações ocorridas. Essa nova versão é adicionada sob a forma de um novo *datastream* do objeto em causa.

O Fedora visa garantir a longevidade dos recursos armazenados e fornece funcionalidades capazes de auxiliar na preservação digital. Este sistema providencia um serviço de repositório nuclear, contudo o facto de ser um sistema modular faz com que seja possível acrescentar novos serviços – através de *Web Services*⁵⁰ – de modo a melhorar a sua capacidade de funcionamento. O Fedora está indicado para bibliotecas e arquivos. Está apto a fornecer acesso a coleções grandes e complexas, quer de âmbito académico, histórico ou cultural, independentemente do tipo de formato dos objetos digitais.

⁵⁰ Um *Web Service* é o meio de comunicação utilizado na interação entre dispositivos eletrónicos através de uma rede. Permite que novas aplicações ou aplicações desenvolvidas em diferentes plataformas sejam compatíveis (interoperabilidade) (W3C Working Group, 2004).

9.1.3 – LOCKSS

A iniciativa LOCKSS (*Lots Of Copies Keep Stuff Safe*)⁵¹ teve origem na *Stanford University* em 1998. Esta iniciativa tinha como objetivo apoiar a comunidade bibliotecária numa era marcada pelo aumento crescente da utilização de plataformas digitais, através da criação de uma plataforma que permitisse preservar os documentos divulgados na Internet, de modo a que o seu acesso fosse garantido ao longo do tempo. Segundo Schneider (Cit. por Souza *et al.*, 2012, p. 70-71), inicialmente, o LOCKSS apenas tinha como objetivo preservar revistas científicas em formato digital em bibliotecas académicas, para que os bibliotecários pudessem obter, conservar e disponibilizar cópias de artigos dessas revistas. No entanto, com o passar do tempo, o LOCKSS ficou apto a preservar qualquer tipo de objeto digital, desde dissertações/teses, jornais ou coleções de fotografias.

Consiste num sistema de preservação de código aberto, baseado no princípio de que a existência de várias cópias do conteúdo que se pretende preservar, distribuídas por diversos locais geograficamente afastados, permite garantir a sua recuperação e acesso por parte das entidades participantes. Trata-se de um sistema *peer-to-peer*⁵², o que permite que o mesmo recurso digital seja armazenado e preservado localmente em várias instituições, tendo as últimas que efetuar as respetivas ações de custódia dos artigos nos seus servidores locais (Márdero Arellano, 2012, p. 86). O sistema LOCKSS deve monitorizar periodicamente as cópias das publicações existentes, de modo a assegurar que a integridade do conteúdo se mantém em cada nó.

Para que uma biblioteca possa partilhar um determinado objeto digital disponível na Internet, presume-se que adquiriu os direitos juntos da editora e/ou autores responsáveis, facilitando, desta forma, os processos legais de partilha entre os diversos servidores existentes. No caso dos periódicos científicos, para que determinada publicação possa ser armazenada, preservada e acedida pelo sistema LOCKSS, é necessário que a respetiva editora conceda permissão legal para tal. Esta permissão é inserida e preservada em conjunto com o conteúdo, para que quer o

⁵¹ Disponível em [www: http://www.lockss.org/](http://www.lockss.org/).

⁵² Arquitetura de redes de computadores onde cada nó da rede pode funcionar em simultâneo como cliente e servidor. Permite partilhar dados entre os nós sem existir um servidor central que controle o fluxo de dados.

recurso quer os respetivos direitos jurídicos não se percam ao longo do tempo (Reich, 2012, p. 78).

O ideal é que o sistema LOCKSS necessite de uma infraestrutura tecnológica de baixo custo e uma necessidade reduzida de gestão técnica. Por isso, uma biblioteca pode, com o auxílio do *software* LOCKSS, transformar um mero computador com acesso à Internet numa aplicação de preservação digital denominada por *caixa LOCKSS*. Uma caixa LOCKSS permite adquirir, preservar e providenciar o acesso a conteúdo importante para a comunidade daquela biblioteca, e fá-lo graças à sua própria rede tecnológica.

Uma caixa LOCKSS possui determinadas funções, designadamente (Reich, 2012, p. 79):

- A recolha automática de conteúdos previamente selecionados das páginas de Internet das publicações científicas;
- A preservação do conteúdo, o que implica a comparação com o mesmo conteúdo recolhido por outras caixas LOCKSS, reparando as eventuais diferenças;
- A disponibilização de páginas de Internet onde se podem encontrar os objetos preservados, caso a página da editora não esteja disponível;
- A verificação periódica da integridade do conteúdo armazenado;
- A existência de uma interface que permita ao administrador adicionar novos conteúdos, monitorizar e controlar o acesso aos recursos já existentes.

Resta referir que o sistema LOCKSS está em conformidade com o Modelo de referência OAIS, tal como descreve o documento *Formal Statement of Conformance to ISO 14721:2003*. Por exemplo, no LOCKSS os PIS são criados pelos editores que publicam uma “página de manifesto” que permite a captura do conteúdo e respetivos metadados. Outro exemplo, é que todos os nós no LOCKSS preservam o mesmo PIS e colaboram na sua auditoria e reparação, por forma a assegurar que a informação permanece autêntica e passível de ser preservada contra todas as adversidades que possam ocorrer (LOCKSS, 2004, p. 1-2).

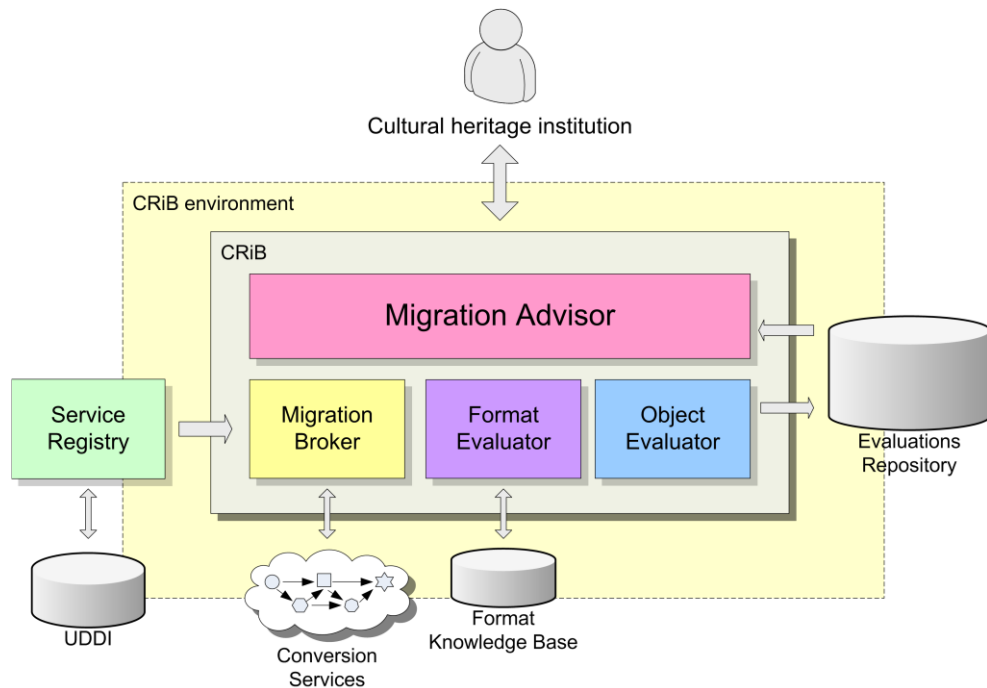
Em suma, Reich (2012, p. 76) considera que o LOCKSS engloba as melhores práticas para uma preservação confiável e de acesso persistente a objetos digitais

através da replicação da informação, distribuição geográfica e diversidade organizacional.

9.1.4 – CRiB

O sistema CRiB consiste numa arquitetura orientada a serviços desenvolvida pela Universidade do Minho. Tem como objetivo auxiliar organizações possuidoras de património cultural na implementação de estratégias de preservação baseadas na migração de formatos. O CRiB faz uso de *Web Services*, o que permite às instituições recorrerem a terceiros (*outsourcing*) para implementar as funcionalidades necessárias para levar o seu plano de preservação digital a bom porto. Este sistema oferece um conjunto de serviços que podem ser implementados em plataformas de repositórios – como DSpace ou Fedora – e que permitem: a identificação do formato do recurso digital; a monitorização dos formatos obsoletos e comparação com as soluções existentes; a elaboração de recomendações de estratégias de migração, consoante os critérios definidos pela instituição; a execução de migrações de formato complexas; a avaliação dessas ações tendo em conta as exigências definidas previamente; a elaboração de relatórios detalhados com a estratégia aplicada (Ferreira, Saraiva e Rodrigues, 2012, p. 34-35).

Através da Figura 13 é possível observar os diversos componentes deste serviço, que contribuem para a elaboração da melhor recomendação de preservação para a instituição visada. De seguida será feita uma breve descrição destes componentes com o objetivo de explicar as funcionalidades de cada um, tendo como base o artigo *CRiB: a service oriented architecture for digital preservation outsourcing*.



Fonte: Ferreira, Baptista e Ramalho, 2007, p. 2

Figura 13 – Arquitetura do sistema CRiB

O *Service Registry* é responsável pela gestão da metainformação necessária para auxiliar a descoberta dos serviços de conversão que, por sua vez, se baseia no padrão UDDI⁵³. Este padrão delinea três elementos para definir os *Web Services*: o primeiro faz a descrição do serviço, o segundo descreve o produtor e o terceiro é composto por exemplos de como a informação pode ser invocada. No entanto, de modo a facilitar a identificação de migrações complexas, foram introduzidos três novos elementos de metadados: o formato original, o formato de destino e o custo de execução. Os valores utilizados nos elementos de metainformação relativos aos formatos dos ficheiros devem ter origem em vocabulário controlado e, neste caso, utiliza-se o registo PRONOM⁵⁴.

O componente *Migration Broker* executa a migração de formatos, assegurando que uma migração composta é manipulada da mesma forma que uma migração singular. É também responsável por medir o desempenho de cada serviço

⁵³ Do inglês *Universal Description, Discovery and Integration*. Trata-se de uma estrutura independente de plataforma que permite localizar, descrever e integrar *Web Services* (OASIS, 2004, p. 5).

⁵⁴ Trata-se de uma iniciativa da *The National Archives* e consiste numa base de dados que visa armazenar informação sobre aplicações de *software* e sobre os formatos de ficheiros existentes que a elas podem ser associados (Brown, 2003, p. 4).

de migração (descoberto através da interação com o *Service Registry*), de acordo com critérios relacionados com disponibilidade, estabilidade, escalabilidade e taxa de transferência. Posteriormente, estas avaliações serão transferidas para o *Migration Advisor*, uma vez que servirão de base para o *ranking* das opções de migração.

O *Object Evaluator* é responsável por avaliar a qualidade de uma ação de migração, através da comparação do objeto original com o objeto convertido. Esta avaliação é feita de acordo com vários critérios, isto é, tem em conta o conjunto de atributos que o objeto digital deve manter durante a intervenção e que o caracterizam como sendo uma única entidade intelectual seja qual for o formato que agora o representa. Estas avaliações são armazenadas no *Evaluations Repository* e devolvidas ao cliente seguindo a estrutura da entidade Evento do Dicionário de Dados PREMIS.

O componente *Format Evaluator* fornece informação acerca dos formatos reconhecidos e suportados pelo sistema, o que permite ao *Migration Advisor* decidir qual o formato mais adequado para receber as características do objeto original. As avaliações geradas pelo *Format Evaluator* são diferentes das do *Migration Broker*: enquanto as avaliações do *Migration Broker* se focam em critérios objetivos medidos durante o processo de conversão, o *Format Evaluator* é suportado pelo *Format Knowledge Base* que possui factos previamente armazenados sobre os formatos suportados. Após a ação de conversão, o potencial ganho de preservação é determinado através da comparação dos valores recolhidos nesta base de dados.

Resta referir aquele que é considerado o núcleo do CRiB uma vez que é ele o componente responsável pela produção das recomendações de migração – o *Migration Advisor*. Este componente cumpre a sua função através da confrontação dos requisitos de preservação definidos pelo cliente com o conhecimento acumulado acerca da qualidade de cada serviço de conversão. É importante referir que o *Migration Advisor* aprende a cada conversão, uma vez que a qualidade de cada serviço de conversão é medida através de vários critérios (perda de dados expectável, estado dos formatos envolvidos e desempenho da migração) e é através desta informação que o *Migration Advisor* é capaz de ordenar as alternativas possíveis e produzir a recomendação apropriada.

Este sistema procura resolver o facto de existirem limitações nas funcionalidades de preservação nos repositórios digitais, ao permitir que um cliente possa invocar *Web Services* para realizar migrações de formatos. Tem como

vantagens a entrega de um relatório detalhado da intervenção de migração efetuada (o que assegura a autenticidade do objeto digital) e a redução dos custos de preservação dado que fornece mecanismos que permitem automatizar e facilitar as atividades de preservação.

Para finalizar, convém mencionar que o sistema CRiB é utilizado pelo projeto RODA (Repositório de Objetos Digitais Autênticos)⁵⁵, inicialmente desenvolvido em 2006 pelo Instituto de Arquivos Nacionais/Torre do Tombo. Tinha como objetivo o desenvolvimento de um repositório de objetos digitais capaz de armazenar, descrever e permitir acesso a toda a documentação digital criada no âmbito da Administração Pública. O RODA baseia-se em tecnologias de código aberto e apoia-se no Modelo de referência OAIS.

O RODA está dividido em diversos componentes e serviços, no entanto o núcleo da sua estrutura centra-se no Fedora que, por sua vez, implementa as funções de repositório, tais como o armazenamento dos objetos digitais e metainformação. A essa estrutura somam-se ainda o RODA *Core Services* (que pode ser acedido programaticamente, e portanto, usado por outras aplicações) e o RODA *Web User Interface* (que permite ao utilizador a realização de pesquisas, o acesso, a gestão dos recursos armazenados e da metainformação e, também, efetuar tarefas de ingestão, preservação e disseminação) (Ramalho *et al.*, 2008, p. 1).

Apesar dos esforços investidos no desenvolvimento do projeto RODA, não existia, na altura, um mecanismo que suportasse atividades de preservação digital e, por isso, os recursos digitais submetidos no repositório continuavam sujeitos à obsolescência tecnológica.

O facto de o RODA e o CRiB seguirem uma arquitetura orientada a serviços – com o objetivo de garantir a interoperabilidade entre serviços, componentes ou plataformas eletrónicas – e o facto deste último oferecer um conjunto de serviços que permitem às instituições manter os seus recursos digitais de forma interpretável e com codificações atuais, fez com que fosse escolhido para ser a entidade responsável pelo planeamento e execução de estratégias de preservação do RODA (Ramalho *et al.*, 2008, p. 2, 5).

⁵⁵ Disponível em www: <http://dglab.gov.pt/roda-repositorio-objects-digitais-autenticos/>.

9.2 – MoReq2010

O MoReq2010⁵⁶ foi desenvolvido pelo *DLM Forum* e publicado em 2011. Trata-se uma especificação de requisitos para sistemas de gestão de documentos de arquivo. Visa estabelecer um conjunto de requisitos – abrangentes, simples e de fácil compreensão – para um sistema de arquivo que possa ser aplicado a vários tipos de realidades informacionais e organizacionais. Por exemplo, esta especificação pode ser utilizada por empresas como um documento auxiliar para a aquisição e configuração de sistemas de gestão de documentos de arquivo de acordo com a sua área de negócio e obrigações legais, até porque teve em consideração normas internacionais para a gestão documental, nomeadamente a ISO 15489:2001.

Esta especificação define um conjunto comum de serviços nucleares, modulares e flexíveis que são partilhados pelos diferentes tipos de sistemas de gestão de documentos de arquivos e que podem ser integrados em outras aplicações. O desenvolvimento do MoReq2010 fica marcado pela existência de duas fases de consulta pública, que permitiram a colaboração de diferentes comunidades de interesse e que acabaram por se mostrar benéficas para o desenrolar do projeto dada a troca de ideias vindas de vários quadrantes da sociedade.

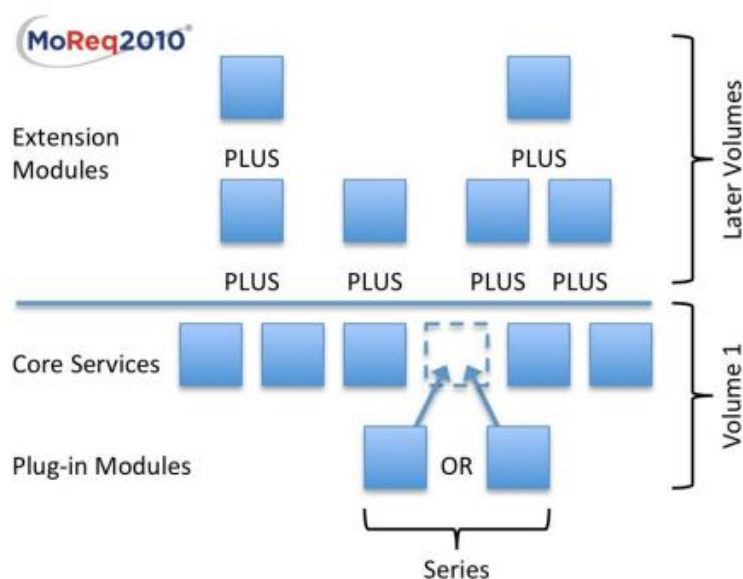
O MoReq2010 tem como antecessores o MoReq1 – publicado em 2001 – e o MoReq2 – publicado em 2008. Desde a sua primeira publicação, o MoReq sempre se procurou manter como uma especificação de âmbito universal (especificação internacional adotada por um elevado número de países incluindo países fora do espaço europeu), disponível em vários idiomas, e pese embora o facto de não ter sido criado para ser considerado como norma, hoje é reconhecido como tal devido à sua ampla aplicação e adoção a nível internacional (DLM Forum, 2011, p. 16).

No entanto, o MoReq2010 também se distingue dos seus predecessores pois apresenta “novos requisitos, nova estrutura, novos conceitos, diferente organização, uma completa nova visão da gestão de documentos de arquivo”. Esta especificação está “organizada por módulos de funcionalidades que [...] podem ser facilmente adicionados ou removidos, de acordo com as necessidades e objectivos da organização” (Vieira e Borbinha, 2011, p. 1). Esta reorganização de módulos permite criar diferentes tipos de sistemas de gestão de documentos de arquivo.

⁵⁶ Disponível em [www: http://moreq2010.eu/](http://moreq2010.eu/).

A implementação modular do MoReq2010, referida no parágrafo anterior, foi de tal ordem distinta em comparação com as características dos seus antecessores, que a DLM Forum alterou o nome da especificação de *Model Requirements Specification for the Management of Electronic Records* para *Modular Requirements for Records Systems*. Enquanto a estrutura dos antecessores se focava nas funções interdependentes de um sistema de gestão de documentos de arquivos, a nova estrutura modular do MoReq permite uma menor dependência, maior flexibilidade e possibilita a personalização da especificação (Vieira e Borbinha, 2011, p. 8).

Na Figura 14 é possível observar a constituição da especificação MoReq2010. É composta por um conjunto de serviços nucleares, módulos de *plug-in*, séries de módulos onde pelo menos um de cada série está ligado ao sistema de gestão de documentos de arquivo e módulos de extensão que correspondem a módulos opcionais e que podem ser adicionados conforme a necessidade da organização (Vieira e Borbinha, 2011, p. 3).

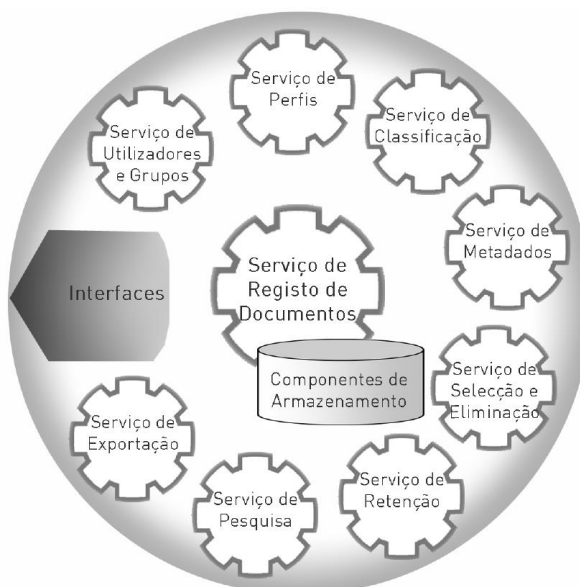


Fonte: DLM Forum, 2011, p. 29

Figura 14 – Estrutura do MoReq2010

Silva e António (2012, p. 3) definem serviço como “a capacidade de realizar tarefas que, no seu conjunto, constituem uma funcionalidade coerente tanto do ponto de vista da entidade que o fornece como da que o solicita”. São nove os serviços

essenciais para o atividade de um sistema de gestão de documentos de arquivo e podem ser vistos na Figura 15.



Fonte: Silva e António, 2012, p. 5

Figura 15 – Serviços do MoReq2010

De seguida serão descritos, de forma sucinta, os requisitos funcionais do MoReq2010 (António, 2014, p. 7-8; Vieira e Borbinha, 2011, p. 4-7):

- Serviço de Utilizadores e Grupos: responsável pela gestão e autenticação de utilizadores ou grupos de utilizadores;
- Serviço de Perfis: permite gerir as permissões dos utilizadores para efetuarem ações no sistema;
- Serviço de Classificação: descreve as funcionalidades capazes de gerir e aplicar planos de classificação;
- Serviço de Registo de Documentos: assegura a gestão documental nos diferentes níveis de agregações. Pelo facto de ser o módulo central não pode ser partilhado com outros sistemas;
- Serviço de Metadados: responsável por gerir os diferentes tipos de objetos e respetivos metadados;
- Serviço de Seleção e Eliminação: permite garantir o cumprimento dos prazos de conservação e destino final, previamente definidos pela tabela de seleção;

- Serviço de Retenção: cabe-lhe a implementação de políticas indispensáveis à salvaguarda e preservação dos documentos de arquivo, conforme aspetos administrativos ou legais associados. Ou seja, impossibilita que um documento de arquivo seja destruído através da retenção do processo da tabela de seleção, precisamente no momento da sua destruição;
- Serviço de Pesquisa: responsável pela recuperação e acesso da informação e metadados associados. O sistema deve garantir que apenas recupera os documentos a que aquele utilizador pode aceder, conforme as suas permissões;
- Serviço de Exportação: permite que o sistema exporte registos para outros sistemas de gestão de documentos, através da sua descrição em XML, mantendo a respetiva metainformação.

“A representação de informação num sistema alinhado com o MoReq2010 é feita através de entidades” (Vieira e Borbinha, 2011, p. 3). Uma entidade possui, obrigatoriamente, metainformação associada, um histórico dos eventos a que esteve sujeita e uma lista com as regras de acesso (observável na Figura 16). Desta forma, os objetos – presentes e futuros – do sistema de arquivo são identificáveis através da análise destas três características. A lista de permissões permite maior controlo e segurança sob um objeto pois é nela que se definem as regras de acesso. No futuro, o MoReq2010 pretende assegurar que uma entidade e respetiva metainformação possam migrar para outros sistemas, sem que haja perda de informação.



Fonte: Silva e António, 2012, p. 6

Figura 16 – Estrutura de uma entidade no MoReq2010

O MoReq2010 define ainda outro conceito: a agregação. Uma agregação “pode ser baseada em qualquer critério ou requisito organizacional ou num conjunto de actividades e/ou documentos relacionados que, quando combinados, podem existir a um nível superior” (Silva e António, 2012, p. 6). Pode então representar uma pasta na qual se agrupam os documentos de arquivo com características semelhantes.

Uma das grandes preocupações do MoReq2010 é a correta e completa exportação de registos para outros sistemas de gestão de documentos, o que também inclui a sua metainformação e todas as propriedades associadas. Para que a exportação, e consequentemente a preservação, seja bem sucedida é necessário que os sistemas sejam compatíveis, o que só é possível através da adoção de boas práticas, nomeadamente, normas/especificações com alta taxa de aceitação a nível internacional.

Em suma, de acordo com Silva e António (2011, p. 7), esta especificação de sistemas de gestão de documentos de arquivo permite: abranger todo o ciclo de vida do documento assegurando as suas características; controlar os prazos de conservação definidos no plano de classificação; desenvolver estratégias de preservação digital; e aceder aos documentos.

Capítulo Dois – Médias e grandes empresas portuguesas

Este capítulo tem como objetivo demonstrar a importância da informação para uma organização, como um auxiliar na criação de valor e na prossecução dos objetivos. Chama-se particular atenção para a informação criada ou recebida no âmbito laboral de uma organização e que reflete as atividades realizadas pelos colaboradores.

De seguida, apresentam-se as várias categorias de empresas existentes, de acordo com recomendações europeias, considerando o seu número de efetivos bem como os seus lucros. Assim, foi possível definir os parâmetros necessários para classificar uma empresa como média ou de grande dimensão.

Por último, faz-se uma breve descrição sobre a realidade das empresas em Portugal através da análise de estudos elaborados pelo Instituto Nacional de Estatística (INE).

1 – O papel da informação no mundo empresarial

Atualmente a informação é um dos maiores ativos de uma organização. Tal como os recursos humanos, financeiros e materiais são essenciais para a concretização da atividade laboral de uma empresa, também a informação possui um papel preponderante para atingir esse propósito. Mas para que possamos compreender tal importância, convém começar por compreender o que é uma organização. Pinto *et al.* (2010, p. 16) consideram uma organização como:

(uma organização é) [...] um grupo estruturado de pessoas – ou seja, um grupo onde cada pessoa é responsável por tarefas bem definidas e onde existe um sistema de articulação entre elas –, que desenvolve um conjunto de actividades visando a definição e prossecução de objectivos comuns (de forma continuada no tempo).

A constituição de uma organização tem como pretensão a criação de valor para os seus *stakeholders*, ou seja, para aquelas entidades que contactam com a realidade empresarial dessa organização, sejam os acionistas, administradores e restantes funcionários, fornecedores, clientes ou, até mesmo, o Estado. Isto mostra que uma organização não se encontra isolada, e ao longo da sua existência sofre diversos estímulos vindos tanto do seu interior como do exterior. Para além das

entidades referidas anteriormente, é necessário também considerar as mensagens e os sinais dados pelos seus concorrentes, uma vez que refletem eventuais tendências ou mudanças no meio ambiente.

Contudo, numa era marcada pela evolução dos dispositivos tecnológicos e o crescimento da Internet, uma organização acede, muitas vezes, em tempo real a esses tais estímulos produzidos por outras entidades. Cabe a cada uma delas ter a capacidade de receber tais informações, por vezes ambíguas, e adaptar-se da forma mais adequada e célere possível, de modo a fazer frente às novas exigências e posicionar-se de modo competitivo face aos demais concorrentes. No entanto, para que tal possa acontecer, é essencial que a informação recebida seja precisa, completa, fiável, relevante e verificável. Caso contrário, se a informação se mostrar de fraca qualidade ou ininterpretável irá afetar a sua compreensão e influenciar de forma negativa os processos de tomada de decisão.

Através da organização e processamento da informação é possível a criação de conhecimento numa empresa. Mas também a troca de experiências e saberes entre colaboradores permite que se desenvolvam novos conhecimentos, se melhorem os processos organizacionais existentes e se produza valor para a instituição. De qualquer das formas, é fulcral que se entenda que o conhecimento reside na mente dos colaboradores, graças às sinergias entre conhecimentos tácito⁵⁷ e explícito⁵⁸ existentes numa organização, e é influenciado pelos seus valores, regras e crenças. Por isso, os colaboradores devem ser encarados como um dos pilares da organização, uma vez que são eles o elo de união entre as estruturas materiais, económicas e informacionais existentes, permitindo, assim, atender à missão e aos objetivos propostos e garantir a inovação e a competitividade no mercado.

Assim sendo, a informação produzida no âmbito organizacional, por parte de um colaborador, é particularmente importante para a instituição uma vez que regista a execução das suas atividades administrativas e reflete o conhecimento e experiência adquiridos ao longo do tempo de trabalho. Para além disso, o seu registo – através de documentos – funciona como prova das atividades ocorridas e daquilo

⁵⁷ “Conhecimento tácito é o conhecimento pessoal, que é difícil de formalizar ou comunicar a outros. É constituído do know-how subjetivo, dos insights e intuições que uma pessoa tem depois de estar imersa numa atividade por um longo período de tempo” (Choo, 2003, p. 37).

⁵⁸ “Conhecimento explícito é o conhecimento formal, que é fácil transmitir entre indivíduos e grupos. É frequentemente codificado em fórmulas matemáticas, regras, especificações, e assim por diante” (Choo, 2003, p. 37).

que ainda se pretende realizar. Tal informação pode ser reutilizada (quer para tomadas de decisão ou planeamento estratégico), o que faz com que os membros da organização sejam produtores mas, também, consumidores da informação (Lousada e Valentim, 2010, p. 364).

Este trabalho tem como foco a preservação da informação orgânica⁵⁹, ou seja, aquela informação produzida ou recebida no contexto laboral de uma organização e que reflete o cumprimento das suas atividades, e que se encontra armazenada (ou pode ser convertida) em meio digital. A manutenção da informação – neste caso, existente em objetos digitais – é fulcral para a sua possível reutilização no futuro e para a conservação da memória institucional, assumindo, assim, um propósito probatório e histórico. Esta visão é defendida por Rousseau e Couture (1998, p. 65):

De facto, a informação orgânica é utilizada pelas unidades do organismo, quer pelo seu valor primário, a fim de decidir, de agir e de controlar as decisões e as ações empreendidas, quer pelo seu valor secundário, a fim de efectuar pesquisas retrospectivas que põem em evidência decisões ou acções passadas [...]. A gestão da informação orgânica, permitindo a pesquisa retrospectiva, reduz a incerteza e melhora a tomada de decisão, aprofundando o conhecimento da cultura institucional e do processo de decisão.

Contudo, apesar de, atualmente, todas as organizações serem capazes de gerar informação em meios digitais, optou-se por apenas tratar as médias e grandes empresas (de Portugal), uma vez que serão estas as que têm maior aptidão para gerar elevadas quantidades de informação. Tendo isso em conta, de seguida, serão apresentados alguns conceitos relacionados com o mundo empresarial e que são essenciais para a compreensão deste trabalho.

2 – Definição de empresas: micro, pequenas, médias e grandes

Atualmente, as empresas atuam num mercado globalizado e altamente competitivo, onde procuram comercializar os seus bens e serviços. Quer seja por opção ou até mesmo por necessidade, cada vez mais empresas contactam com países estrangeiros com o objetivo de expor os seus produtos em montras internacionais.

⁵⁹ Rousseau e Couture (1998, p. 65) apresentam alguns exemplos de informação orgânica: contratos, processos de clientes, processos de pessoal, planos e orçamentos técnicos, atas, relatórios, estudos, pesquisas, etc.

Atendendo ao mercado livre e sem fronteiras, a Comissão Europeia (CE) considerou necessário o estabelecimento de definições comuns para os vários tipos de empresas existentes. O facto de estes conceitos divergirem de país para país, consoante a interpretação do governo instituído, poderiam afetar a concorrência e competitividade.

Assim sendo, optou-se por seguir a linha orientadora proposta pela CE⁶⁰ e cuja aplicação é recomendada aos Estados-Membros mas, também, ao Banco Europeu de Investimento e ao Fundo Europeu de Investimento.

Para a CE (2003, p. 36), uma empresa pode ser considerada como:

(uma empresa é) [...] qualquer entidade que, independentemente, da sua forma jurídica, exerça uma actividade económica, incluindo, designadamente, as entidades que exerçam uma actividade artesanal e outras actividades a título individual ou familiar, as sociedades de pessoas ou as associações que exerçam regularmente uma actividade económica.

A Recomendação da CE tem como principal objetivo a definição de micro, pequenas e médias empresas (PME), uma vez que são elas as grandes responsáveis pelo desenvolvimento económico na Europa. A evidência disso mesmo são os dados que revelam que na União Europeia, aproximadamente, 23 milhões de PME garantem trabalho a 75 milhões de pessoas e representam 99% do número total de empresas (CE, 2006, p. 5). Considerando as, eventuais, dificuldades que estas empresas têm para a obtenção de financiamento, no momento de início de atividade, a CE viu-se na necessidade de as apoiar e o estabelecimento de definições comuns visa garantir que apenas as empresas que se enquadrem nestes termos têm acesso a esses apoios.

A CE (2003, p. 39) define PME da seguinte forma:

1. A categoria das micro, pequenas e médias empresas (PME) é constituída por empresas que empregam menos de 250 pessoas e cujo volume de negócios anual não excede 50 milhões de euros ou cujo balanço total anual não excede 43 milhões de euros.
2. Na categoria das PME, uma pequena empresa é definida como uma empresa que emprega menos de 50 pessoas e cujo volume de negócios anual ou balanço total anual não excede 10 milhões de euros.
3. Na categoria das PME, uma microempresa é definida como uma empresa que emprega menos de 10 pessoas e cujo volume de negócios anual ou balanço total anual não excede 2 milhões de euros.

⁶⁰ Refere-se à Recomendação nº 2003/361/CE de 6 de Maio de 2003.

Nesta definição existem dois tipos de critérios subjacentes: o número de efetivos⁶¹ e um indicador financeiro, que tanto pode ser o volume de negócios⁶² ou o balanço total⁶³. O número de efetivos deve ser obrigatoriamente cumprido; já no caso dos indicadores financeiros, cabe à empresa escolher qual dos dois irá cumprir, podendo exceder um deles sem perder o estatuto. É a comparação entre estes três critérios que permite determinar o tipo de empresa. Na Tabela 1 é apresentada novamente a definição supramencionada mas de forma mais intuitiva.

Tabela 1 – Caracterização de médias, pequenas ou micro empresas

Categoria da empresa	Número de efetivos	Volume de Negócios Anual	ou	Balanço Total Anual
Média	< 250	≤ 50 milhões de euros		≤ 43 milhões de euros
Pequena	< 50	≤ 10 milhões de euros		≤ 10 milhões de euros
Micro	< 10	≤ 2 milhões de euros		≤ 2 milhões de euros

Fonte: CE, 2014 [Em linha]

Não obstante, para que uma empresa possa calcular os seus dados (número de efetivos e volume de negócios/balanço total) – e, dessa forma, possa ser considerada como uma PME – precisa de determinar se é uma empresa autónoma, parceira ou associada. Isto é, tem de definir as relações existentes com outras empresas e, conforme a sua posição, pode ter de juntar os seus dados com os de outras. Caso se trate de uma empresa autónoma, ou seja, totalmente independente, apenas recorre aos seus dados registados nas contas anuais para verificar se respeita os limites referidos na Tabela 1 (CE, 2006, p. 16-17).

⁶¹ Refere-se aos trabalhadores a tempo inteiro, parcial e sazonais, incluindo: assalariados; proprietários-gestores; sócios que exerçam atividade regular na empresa. O número de efetivos é expresso em Unidades de Trabalho-Ano, o que significa que aqueles efetivos que trabalharam a tempo inteiro são considerados como uma unidade; já os trabalhadores a tempo parcial e sazonais são contabilizados com frações de unidade (CE, 2006, p. 15).

⁶² Refere-se ao cálculo das receitas da empresa durante um período de tempo, resultantes das suas vendas e prestações de serviços, depois de pagos quaisquer descontos. (CE, 2006, p. 15).

⁶³ Refere-se à situação económico-financeira da empresa e representa a sua situação patrimonial (ativo, passivo e capital) num determinado período de tempo. Pode ser traduzido na seguinte equação: Ativo = Passivo + Capital Próprio (PACONTAS, 2012).

Importa referir que, no caso de Portugal, a definição de PME – proposta pela CE – foi adotada, tendo sido publicada em Diário de República no Decreto-Lei nº 372/2007.

Uma vez que o conceito de PME se encontra bem definido, podemos deduzir a definição do outro tipo de empresa que resta referir: a grande empresa. Assim sendo, podemos considerar uma grande empresa como aquela que emprega mais de 250 funcionários e cujo volume de negócios anual excede os 50 milhões de euros e cujo balanço total anual excede os 43 milhões de euros.

3 – Empresas em Portugal

Em Portugal, o INE é responsável por produzir estudos para revelar o estado da sociedade, nomeadamente o estado do tecido empresarial português e as suas implicações na economia e na vida das pessoas. Nesses estudos podemos encontrar informação relativa ao número de empresas existentes no nosso país, os seus indicadores económicos, ou comparação desses indicadores entre PME e grandes empresas. É com base nestes estudos que se faz, de seguida, o ponto de situação das empresas que atuam em Portugal.

Assim sendo, os dados acedidos referem-se a 2012 e revelam a existência de 1.062.782 empresas não financeiras⁶⁴ e apenas 23.670 empresas financeiras⁶⁵ (INE, 2014^b, p. 17).

Os mesmos dados publicados pelo INE (e observáveis na Figura 17) revelam ainda que 99,9% das empresas não financeiras em Portugal eram PME (sendo 95,9% microempresas) e eram responsáveis por 57,6% do volume de negócios realizado. Dentro desta categoria, as médias empresas eram 0,5% das empresas e registaram um volume de negócios de 20%. Já as grandes empresas, embora sendo apenas 0,1% das empresas, eram responsáveis por 42,4% do volume de negócios.

⁶⁴ Refere-se às instituições que se dedicam à produção de bens e/ou à prestação de serviços (Infopédia, 2003-2015^a).

⁶⁵ “São as instituições cuja atividade principal é a concessão de crédito, fazendo para isso captação de poupanças junto dos agentes económicos, em especial das famílias. Constituem, como tal, um intermediário de suma importância na transferência de fundos das entidades que possuem poupanças para aquelas que delas necessitam para financiar os seus investimentos ou gastos correntes” (Infopédia, 2003-2015^b).

		Empresas		Pessoal ao Serviço		VVN		VABpm		Gastos com Pessoal	
		Nº	% Total	Nº	% Total	10 ⁶ EUR	% Total	10 ⁶ EUR	% Total	10 ⁶ EUR	% Total
Total empresas não financeiras		1 062 782		3 511 666		325 870		75 969		47 454	
Forma Jurídica	Empresas individuais	709 171	66,7	816 349	23,2	15 190	4,7	5 931	7,8	1 273	2,7
	Sociedades	353 611	33,3	2 695 317	76,8	310 680	95,3	70 038	92,2	46 181	97,3
Dimensão	PME	1 061 767	99,9	2 742 643	78,1	187 660	57,6	45 390	59,7	30 397	64,1
	Micro	1 019 494	95,9	1 574 424	44,8	60 664	18,6	15 583	20,5	9 432	19,9
	Pequenas	36 645	3,4	669 143	19,1	61 953	19,0	14 290	18,8	10 727	22,6
	Médias	5 628	0,5	499 076	14,2	65 044	20,0	15 517	20,4	10 238	21,6
Localização (NUTS II)	Grandes	1 015	0,1	769 023	21,9	138 210	42,4	30 579	40,3	17 056	35,9
	Norte	347 939	32,7	1 161 905	33,1	89 199	27,4	20 851	27,4	13 759	29,0
	Centro	230 274	21,7	653 964	18,6	51 916	15,9	11 915	15,7	7 571	16,0
	Lisboa	309 136	29,1	1 239 802	35,3	154 905	47,5	35 875	47,2	21 195	44,7
Localização (NUTS II)	Alentejo	75 540	7,1	189 033	5,4	14 312	4,4	3 142	4,1	2 063	4,3
	Algarve	54 808	5,2	135 025	3,8	6 342	1,9	1 791	2,4	1 270	2,7
	R.A.Açores	24 559	2,3	64 081	1,8	4 747	1,5	1 040	1,4	701	1,5
	R.A.Madeira	20 526	1,9	67 856	1,9	4 449	1,4	1 356	1,8	895	1,9

Fonte: INE, 2014^b, p. 23

Figura 17 – Principais indicadores económicos das empresas não financeiras de Portugal em 2012

As PME foram responsáveis pela criação de 78,1% dos postos de trabalho nacionais, tendo as microempresas um papel fundamental dado que esta categoria empregou mais de 1,5 milhões de trabalhadores, que correspondem a 44,8% do valor total. As grandes empresas, mesmo sendo somente 0,1% das empresas portuguesas, conseguem empregar mais de 750 mil pessoas, o que equivale a 21,9% do total.

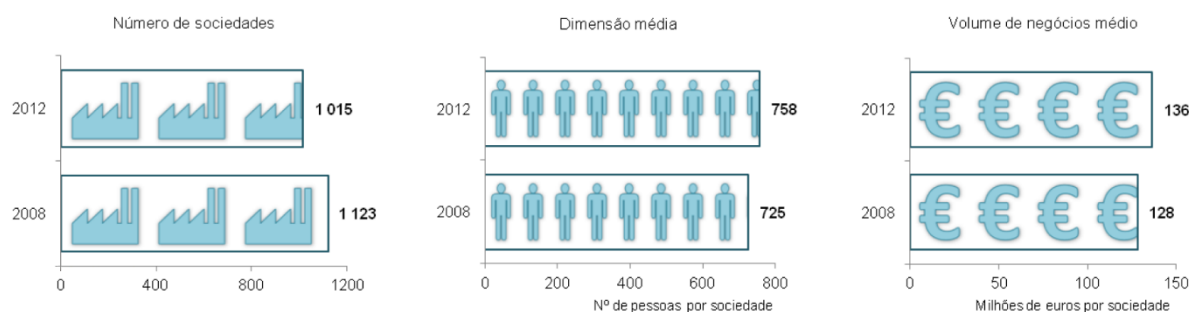
Nesse mesmo estudo, é possível, ainda, notar que grande parte das empresas existentes se concentram nas regiões Norte e Lisboa, e nos sectores das atividades de comércio por grosso e a retalho e das atividades administrativas e dos serviços de apoio (INE, 2014^b, p. 59).

O INE revela ainda um outro estudo onde compara a situação das grandes empresas entre os anos de 2008 e 2012. Isto é, refletem-se as consequências resultantes da crise que teve início em 2008 e que afetou a situação económico-financeira na Europa mas, também, no resto do Mundo.

Neste estudo conclui-se que as grandes empresas conseguiram reagir melhor à crise comparativamente às PME. Neste período de crise, quer as PME e as grandes empresas viram o seu número e desempenho a diminuir, contudo as últimas apresentaram resultados menos negativos face às primeiras, nomeadamente ao nível da redução de pessoal ao serviço e do volume de negócios. As grandes empresas

conseguiram, inclusive, em 2012, apresentar uma dimensão e volume de negócios superiores a 2008 (Figura 18) (INE, 2014^a, p. 1-2).

No que concerne à remuneração mensal, as grandes empresas apresentam valores superiores às PME, mas são estas que registam um maior aumento no período entre 2008 e 2012. No entanto, este valor pode ser explicado com o aumento do salário mínimo nacional (INE, 2014^a, p. 3).



Fonte: INE, 2014^a, p. 5

Figura 18 – Número, dimensão e volume de negócios das grandes empresas (2008-2012)

Apesar de todos os contratempos vividos, é inegável a importância das empresas para os habitantes de um determinado país, uma vez que são elas que permitem a criação de emprego e o aumento do poder de compra, o que, consequentemente, incentiva o crescimento económico.

Contudo, este trabalho não tem como objetivo a realização de uma análise económico-financeira pormenorizada das empresas em Portugal. O que na realidade se pretende é demonstrar a tipologia das empresas existentes, e de que modo constituem um papel ativo e fulcral quer na sociedade quer na economia.

No decorrer das suas atividades organizacionais, os colaboradores de uma empresa produzem e utilizam inúmera informação em formato digital, sendo que grande parte deverá ser conservada por vários anos.

Assim sendo, esta dissertação pretende analisar a importância que a preservação digital tem, também, para as empresas, e não só para as instituições de ensino, arquivos ou bibliotecas. Contudo, optou-se por focalizar o estudo apenas nas médias e grandes empresas, uma vez que se considera que estas empresas estarão mais aptas a produzir maiores quantidades de informação e mais sensíveis à necessidade da sua preservação.

Capítulo Três – Estudo de caso

Este capítulo dedica-se ao estudo da posição atual das empresas portuguesas (em particular, das médias e grandes empresas) em relação à temática da preservação digital.

Seguindo a definição do objeto de estudo, apresentado anteriormente, neste capítulo começa-se por explicar como foi feita a seleção das empresas que foram convidadas para o preenchimento do questionário preparado.

Posteriormente, expõe-se e analisam-se os resultados obtidos.

1 – Metodologia para a seleção das empresas

Esta secção descreve sucintamente a metodologia usada para a seleção das empresas que serão alvo da aplicação do inquérito proposto, seguindo-se de uma análise da lista de empresas compilada.

1.1 – Critérios de seleção

A base de dados Sistema de Análises de Balanços Ibéricos (SABI) – disponível nos laboratórios da Escola Superior de Estudos Industriais e de Gestão (ESEIG) – foi a plataforma escolhida para auxiliar na recuperação dos dados de empresas. Estes dados estarão sujeitos à definição e aplicação de metodologias de seleção de empresas, sobre as quais serão realizados os inquéritos.

Para a aplicação dos critérios de classificação de empresas, e sua posterior divisão em diferentes categorias, selecionaram-se, na plataforma SABI, as empresas sediadas em Portugal e atualmente ativas. Para facilitar a análise das diferentes características das empresas, foram exportados os dados necessários para a elaboração deste estudo para uma tabela, num documento de folha de cálculo. Os campos exportados foram os seguintes:

- Denominação social;
- Último ano disponível⁶⁶;
- Endereço de Internet;

⁶⁶ Data a que se referem os dados da empresa.

- Código da CAE Rev.3⁶⁷ Principal;
- Descrição da CAE Rev.3 Principal;
- Código(s) da CAE Rev.3 Secundária;
- Último número de empregados;
- Número de empregados – último ano disponível⁶⁸;
- Volume de negócios (milhares de euros) – último ano disponível;
- Total do capital próprio e do passivo (milhares de euros)⁶⁹ – último ano disponível.

Embora não exista uma definição estabelecida de grande empresa (ao contrário do que acontece para as PME – ver Tabela 1), considerou-se como pressuposto assumido que uma grande empresa é aquela que possui mais de 250 efetivos e um volume de negócios superior a 50 milhões de euros e um balanço total superior a 43 milhões de euros.

Contudo, para este estudo apenas se contemplam as médias e grandes empresas e, por isso, os valores mínimos para uma empresa ser considerada grande não são relevantes, uma vez que são utilizados os valores minimizantes para as médias empresas.

Também já foi referido que, nos critérios de definição de PME, o critério do número de efetivos tem obrigatoriamente de ser cumprido. Mas, relativamente aos critérios financeiros, uma PME pode cumprir um dos dois (volume de negócios ou balanço anual), podendo exceder um deles sem perder o seu estatuto. O que significa que os limites mínimos de todos os critérios para uma empresa ser considerada média devem ser verificados e devem ser utilizados durante o processo de seleção de empresas. São eles:

- Mais de 50 efetivos;

⁶⁷ Refere-se à Classificação Portuguesa das Atividades Económicas - Revisão 3. Pretende uniformizar as categorias das atividades de acordo com as nomenclaturas propostas pela Comunidade Europeia. O CAE Rev.3 permite consolidar o “sistema estatístico nacional, quer pelo papel que desempenha na recolha, tratamento, publicação e análise da informação, quer pelo sentido de coerência e de unidade que confere ao sistema” (Decreto-Lei nº 16/93, 2007, p. 8440).

⁶⁸ Como é possível observar existem dois campos relativos ao número de trabalhadores de uma empresa. A existência destas duas variáveis deve-se à forma como os dados são obtidos. Enquanto um é obtido através do contacto direto com as empresas, o outro obtém-se graças à submissão de informações de natureza contabilística, fiscal e estatística na plataforma Informação Empresarial Simplificada (http://www.ies.gov.pt/site_IES/site/home.htm).

⁶⁹ Dada a equação fundamental da contabilidade (referida na nota de rodapé 63), considerou-se para como valor do balanço a segunda parte dessa equação: Passivo + Capital Próprio.

- Volume de negócios superior a 10 M€;
- Balanço total superior a 10 M€.

Para além do pressuposto assumido já referido (a definição de grande empresa), assumiu-se também que todas as empresas recolhidas no SABI são autónomas e, por isso, se regem pelos limites supracitados. A justificação para tal pressuposto deriva do facto de ser impossível saber a totalidade das relações organizacionais entre empresas em tempo útil, de modo a poder ajustar o questionário.

Todas as empresas que cumpram os critérios acima referidos serão médias ou grandes empresas e poderão ser contempladas na participação do inquérito por questionário. Para isso, no documento de folha de cálculo devem-se aplicar os seguintes filtros:

- Os números de efetivos considerados devem ser os valores dos campos “Último número de empregados” e “Número de empregados – último ano disponível”, garantindo que um destes campos tem um valor mínimo de 50;
- Os campos “Volume de negócios (milhares de euros) – último ano disponível” e “Total do capital próprio e do passivo (milhares de euros) – último ano disponível” devem ter como valor mínimo um volume de 10 milhões de euros.

De seguida procedeu-se à compilação dos *emails* das empresas selecionadas de modo a efetuar o seu registo na plataforma de questionários.

No final do processo de recolha de informação foram exportados os dados das empresas selecionadas (nome e *email*) para a plataforma LimeSurvey⁷⁰.

Após o fim do prazo disponível para participação, os dados serão processados com recurso à aplicação Microsoft Excel 2013.

⁷⁰ Disponível em www: <https://www.limesurvey.org/pt/>.

1.2 – Processo de seleção das empresas

A exportação dos dados foi feita a 9 de maio de 2014 e nessa data foram encontradas 370.577 empresas ativas em Portugal (ver Figura 19). Esta lista foi filtrada de acordo com os critérios definidos na secção anterior, obtendo-se um número total de 1.921 médias e grandes empresas.

7			
8	1. País/Região: Portugal	491.862	491.862
9	2. Estado: Activa	370.577	370.577
10	Pesquisa Booleana : 1 E 2		
11		Total	370.577

Figura 19 – Número de empresas exportadas pelo SABI

Durante a realização de pesquisas de contacto de *email* para cada empresa surgiu o problema de nem todas as empresas disponibilizarem um mecanismo eficiente de contacto, nomeadamente via *email*. Muitas optam pela disponibilização de um formulário, o que não permitiu a aplicação do questionário. O Gráfico 3 resume o processo de recuperação dos contactos das empresas.

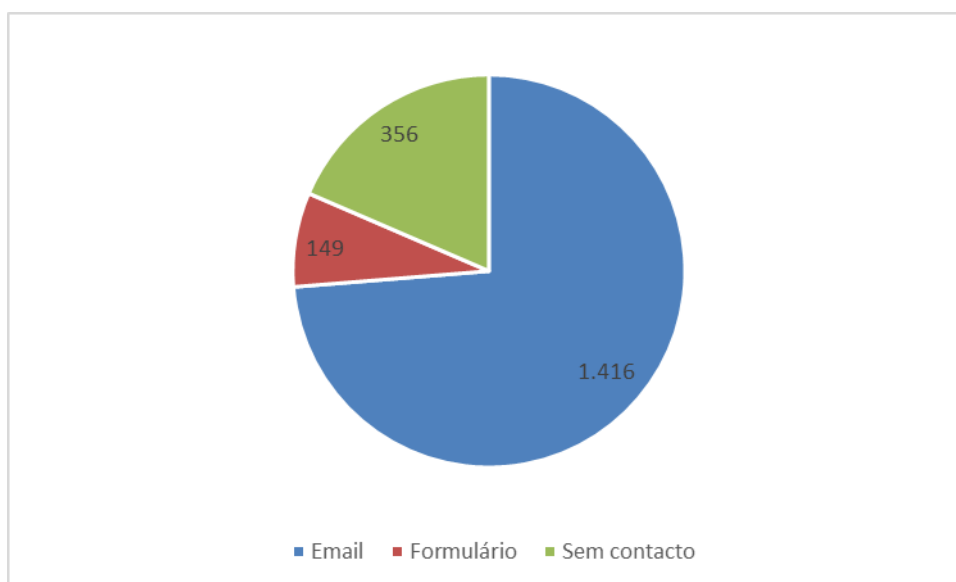


Gráfico 3 – Formas de contacto das empresas seleccionadas

Ao analisar o gráfico podemos observar que se obtiveram 1.416 *emails* de empresas das 1.921 médias e grandes empresas inicialmente seleccionadas. Assim, apenas 73,71% das empresas seleccionadas poderiam ser devidamente contactadas.

Porém, ao importar os contactos das empresas para a plataforma de questionários é possível fazer uso de uma funcionalidade da aplicação que permite verificar a existência de contactos duplicados, nomeadamente através do *email*. Por conseguinte, das 1.416 médias e grandes empresas com *email*, reduzimos esse valor para apenas 1.358 empresas.

2 – Análise ao questionário

Como já foi referido, neste trabalho elaborou-se um inquérito guiado por questionário e utilizou-se uma plataforma de questionários *online* para que as empresas selecionadas tivessem a possibilidade de participar neste estudo. O questionário esteve disponível desde 2 de julho de 2014 até 19 de janeiro de 2015.

O questionário, propriamente dito, é antecedido de uma breve apresentação desta dissertação, nomeadamente: a instituição de ensino e curso frequentados, título e objetivos do trabalho. É pedida a colaboração das empresas e manifesta-se a confidencialidade das respostas dadas. Salienta-se o facto de não serem efetuadas perguntas da esfera privadas/confidenciais e das empresas não serem identificadas no momento da apresentação de resultados.

Esta apresentação (consentimento informado) pode ser encontrada no Anexo G.

O questionário concebido está dividido em quatro grupos:

- Grupo I – Perfil da empresa;
- Grupo II – Gestão documental;
- Grupo III – Preservação digital;
- Grupo IV – Interesse na receção dos resultados e comentários.

Uma vez que algumas das questões apresentadas nos grupos são condicionais, isto é, algumas questões variam de acordo com as respostas dadas a outras questões optou-se por representar o inquérito na forma de fluxogramas, um por cada grupo. Com isso pretende-se que a interpretação do questionário seja mais fácil. Serão ainda apresentadas as motivações que levaram à elaboração de cada pergunta.

O inquérito completo, enviado às empresas selecionadas, é apresentado no Anexo H.

2.1 – Grupo I – Perfil da empresa

Na primeira parte do questionário foram redigidas as questões necessárias para identificar o perfil da empresa.

Logo à partida pretende-se definir a categoria da empresa visada, isto é, se se trata de uma média ou grande empresa; pública ou privada. De seguida procura-se posicionar de forma geográfica a empresa, obtendo a sua localização. Para isso optou-se por solicitar a escolha de um dos dezoito distritos ou de uma das regiões autónomas de Portugal.

No caso de se tratar de uma empresa privada é também questionada a localização da sede social da empresa. Tal questão deve-se ao facto de muitas das empresas que atuam em Portugal estarem sediadas no estrangeiro. Esta questão não se aplica às empresas públicas, uma vez que se encontram sob alçada do Estado Português.

Para finalizar este grupo, é questionado o principal âmbito de atuação da organização. Para auxiliar à elaboração das hipóteses de resposta para esta pergunta, utilizou-se o campo “Descrição da CAE Rev.3 Principal” que havia sido exportado pelo SABI. Assim, para além de obtermos a dimensão da empresa, ficamos também a conhecer a sua principal atividade económica.

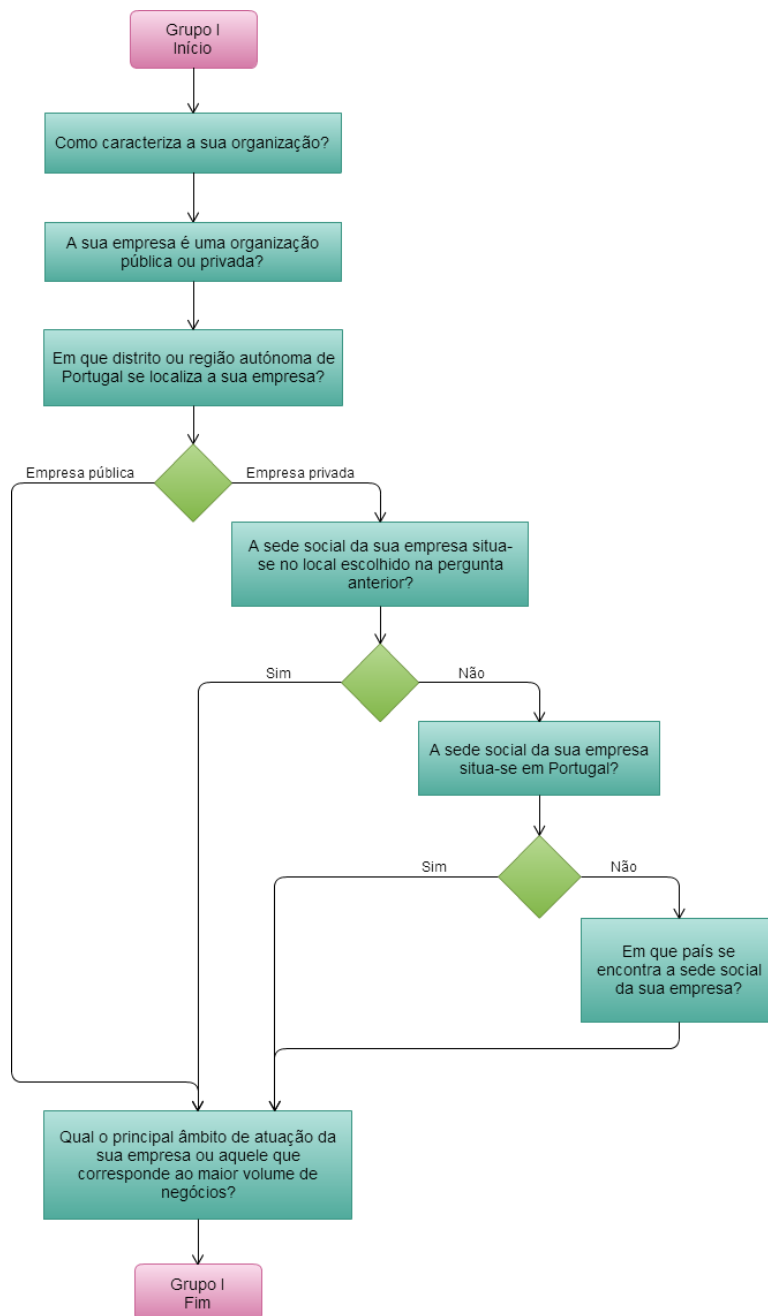


Figura 20 – Fluxograma das questões do Grupo I

2.2 – Grupo II – Gestão documental

Na segunda parte do questionário procura-se entender a importância que a empresa visada dá aos documentos digitais e às tecnologias necessárias para os criar, processar e armazenar.

Inicialmente, a empresa é questionada sobre a utilização de programas que permitem realizar a gestão de documentos e, conseqüentemente, auxiliar os colaboradores nas suas funções laborais do dia-a-dia.

De seguida, questiona-se a existência de formatos recomendados para aqueles objetos digitais que são mais utilizados no seio de uma empresa (documentos de texto, folhas de cálculo, imagens e *emails*). Desta forma pretende-se perceber se as empresas já aderiram a formatos abertos ou se ainda continuam a utilizar formatos proprietários.

Por último, pretende-se saber quais as tecnologias utilizadas para o armazenamento de dados e quais os cuidados tidos com elas. Assim, procura-se perceber se a empresa dá a devida atenção às tecnologias de armazenamento, através da aplicação de boas práticas que permitam uma correta manutenção dos objetos digitais armazenados. Dessa forma, diminui-se as probabilidades de ocorrência de falhas na infraestrutura tecnológica, o que permite garantir a acessibilidade, a longo prazo, dos conteúdos existentes.

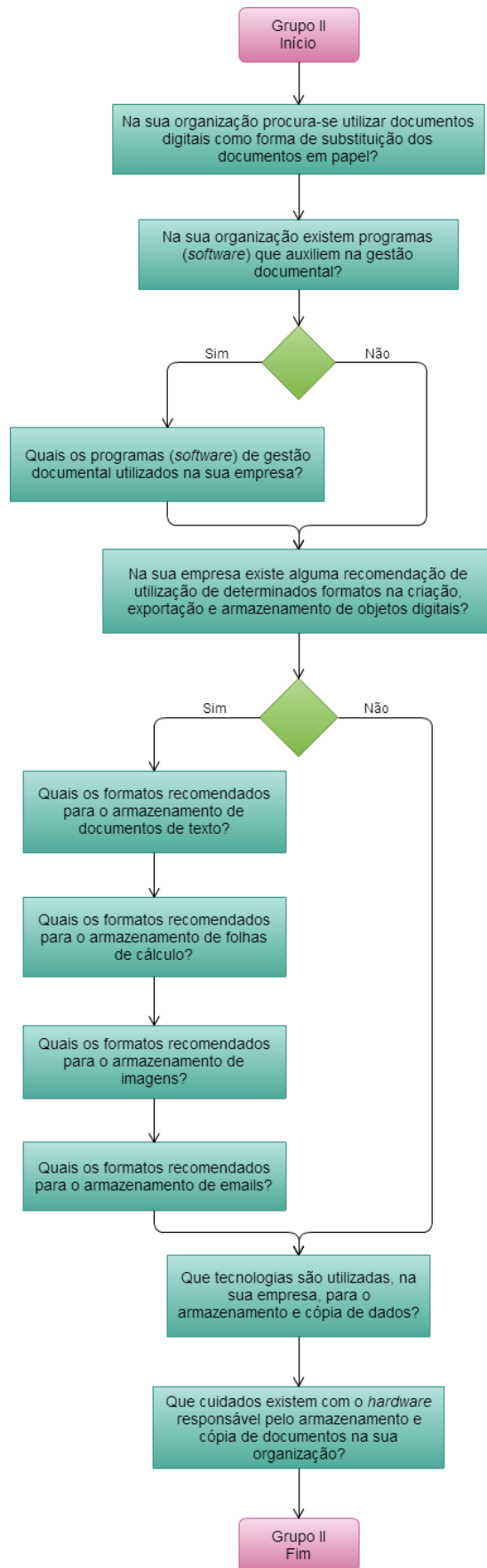


Figura 21 – Fluxograma das questões do Grupo II

2.3 – Grupo III – Preservação digital

A terceira parte do questionário tem como objetivo esclarecer o entendimento da empresa e do inquirido em relação à preservação digital e obter mais informações acerca do processo implementado (caso exista).

A primeira questão pretende clarificar se a empresa, nomeadamente as chefias, se preocupa em manter os objetos digitais acessíveis por um período de tempo prolongado. Nesse sentido, de seguida, questiona-se o inquirido sobre se conhece o conceito de preservação digital.

Logo depois procura-se saber se na organização são aplicadas estratégias de preservação e, no caso de a resposta ser afirmativa, são pedidas mais informações acerca desse assunto. É com o objetivo de descortinar a realidade da preservação digital nas empresas que se questiona a existência de uma equipa multidisciplinar para o planeamento e aplicação das estratégias; o cargo ocupado pelo responsável pelo plano de preservação digital; e quais as estratégias aplicadas.

Pretende-se também saber quais os documentos que são alvo de estratégias e, para isso, são dadas algumas opções de escolha (por exemplo contratos, documentos financeiros ou documentos dos recursos humanos).

Questiona-se ainda qual a motivação para a aplicação de um plano de preservação digital, isto é, se a decisão se deveu à necessidade de armazenar os objetos digitais a longo prazo devido a imposições legais (uma vez que alguns registos funcionam como prova e têm de ser guardados por diversos anos) e/ou para conservar o património da organização.

Por fim, pretende-se obter a opinião do inquirido em relação ao investimento feito pela sua organização para que o plano de preservação digital pudesse ser realizado.

Para terminar este grupo, questiona-se o inquirido se reconhece a importância que a utilização de formatos abertos, normalizados e bem documentados traz para a implementação de um plano de preservação digital.

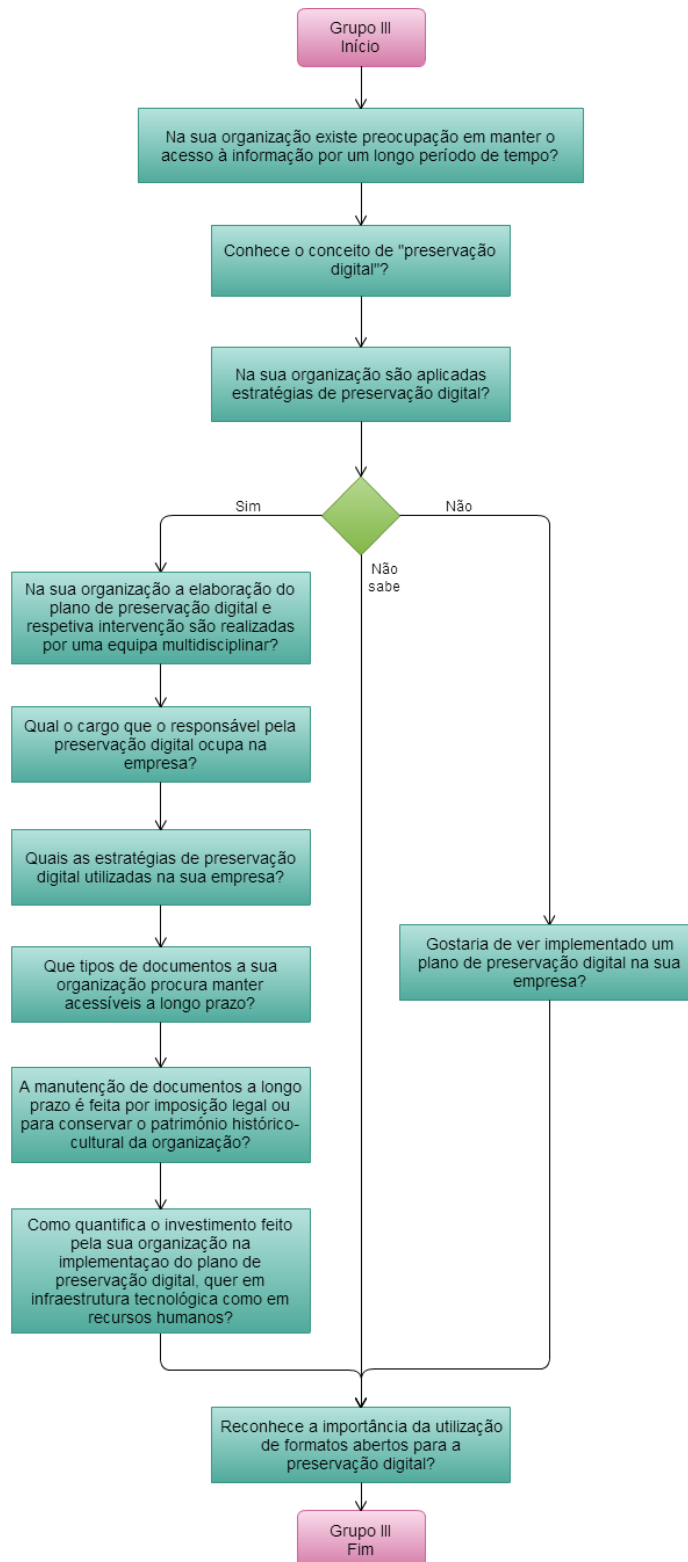


Figura 22 – Fluxograma das questões do Grupo III

2.4 – Grupo IV – Interesse na recepção dos resultados e comentários

Já o último grupo de questões tem como objetivo registar um eventual desejo de a organização receber os resultados do estudo realizado. É, também, dada a possibilidade de registar comentários ou esclarecer alguma informação acerca do questionário.

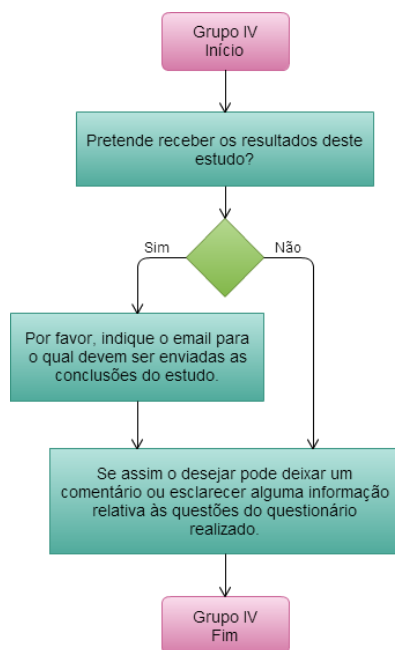


Figura 23 – Fluxograma das questões do Grupo IV

3 – Análise de resultados

Como já foi referido, foram enviados 1.358 pedidos de participação para este estudo. Após o envio dos convites foram enviados vários lembretes periódicos no sentido de recordar as empresas selecionadas de que o questionário ainda não estava preenchido.

Infelizmente, e apesar dos mais de seis meses em que o questionário esteve ativo, apenas 39 empresas completaram o preenchimento, o que equivale a uma taxa de participação de 2,87%. Foi dada a possibilidade às empresas de recusarem a participação, sendo que apenas 20 o fizeram.

Finalmente, registam-se 29 questionários incompletos e 1.270 questionários ignorados.

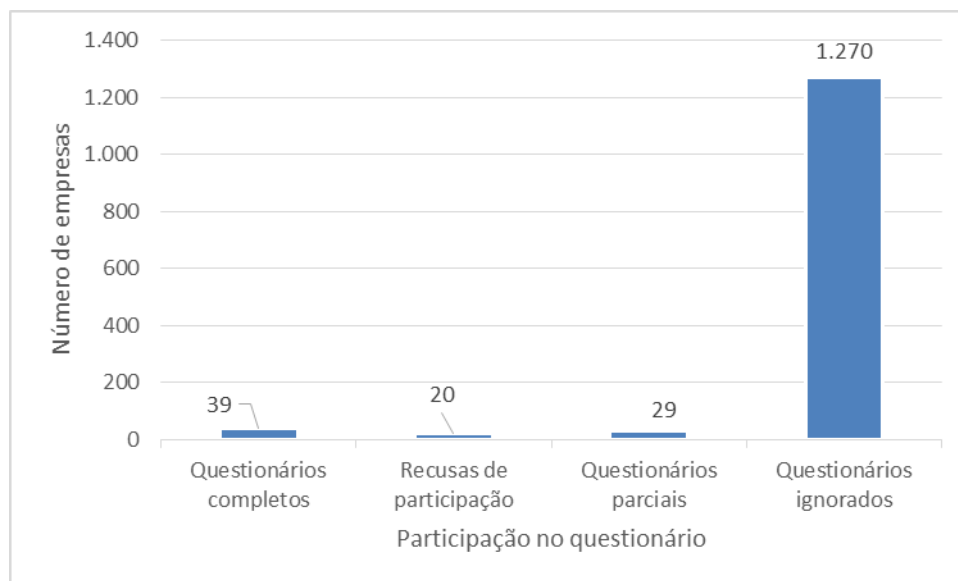


Gráfico 4 – Número de registos do questionário

3.1 – Dimensão da empresa

Das 39 empresas participantes no estudo, 29 delas são médias empresas (o que corresponde a cerca de 75%) e apenas 10 se enquadram na categoria de grandes empresas.

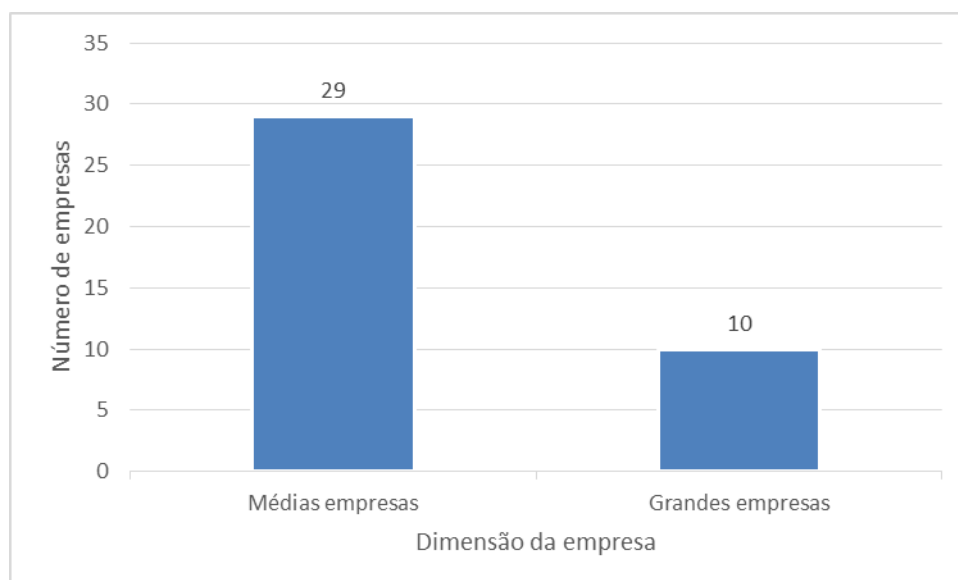


Gráfico 5 – Número de respostas por dimensão da empresa

3.2 – Tipo de empresa

Relativamente à divisão entre empresas públicas e privadas, registaram-se respostas de 7 empresas públicas (18%) e de 32 privadas (82%).

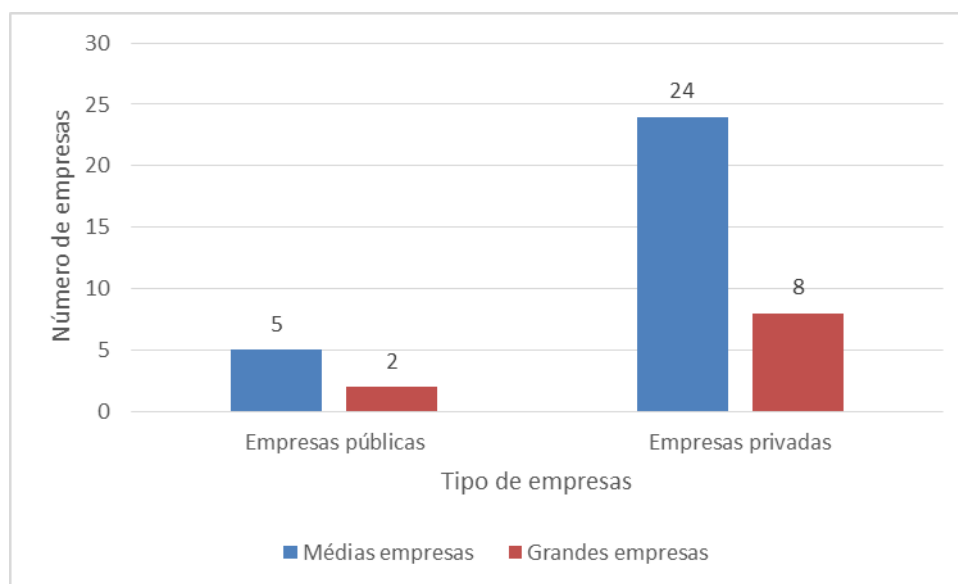


Gráfico 6 – Número de respostas por dimensão e tipologia de empresas

No Gráfico 6 podemos observar a relação entre a dimensão e o tipo de empresa. Quer nas empresas públicas como nas privadas, destaca-se o número de empresas de dimensão média.

3.3 – Localização geográfica e sede social da empresa

No que concerne à distribuição geográfica das empresas participantes, registou-se maior afluência de respostas nos distritos de Lisboa, Aveiro, Braga e Porto. Note-se que, não havendo informação sobre a sede social de todas as empresas, nem se ter garantia de uma amostra não enviesada, estes resultados são os esperados, sendo estes, provavelmente, os quatro maiores distritos de Portugal em termos do sector secundário e terciário.

No entanto, apesar de no conjunto das 1.358 empresas – às quais foram enviados os convites de participação no questionário – estarem englobadas empresas

dos 18 distritos e das 2 regiões autónomas, apenas se obtiveram participações de 14 zonas geográficas.

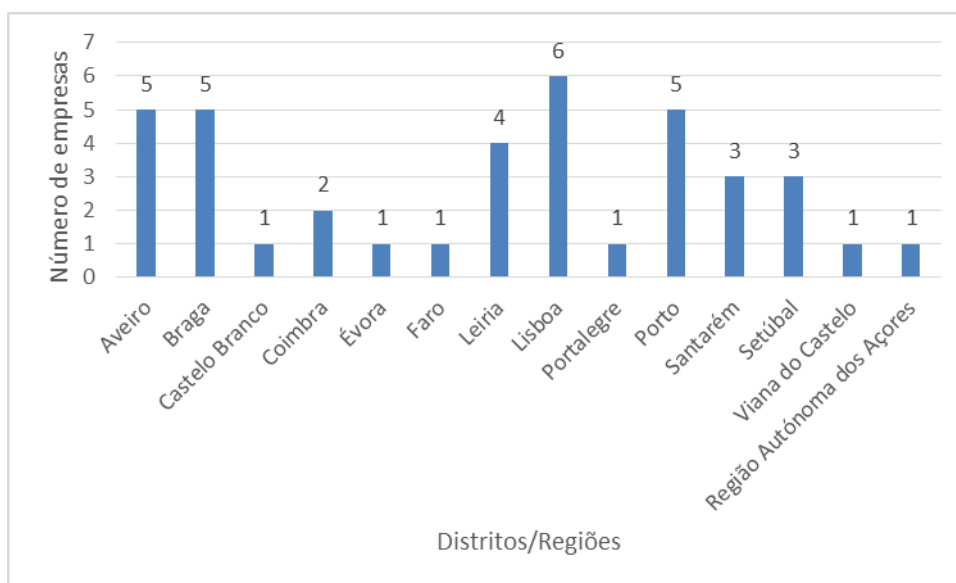


Gráfico 7 – Número de empresas por distritos/regiões

Não obstante, as empresas privadas foram também questionadas se a localização da sua sede social se encontra no distrito/região autónoma previamente escolhido.

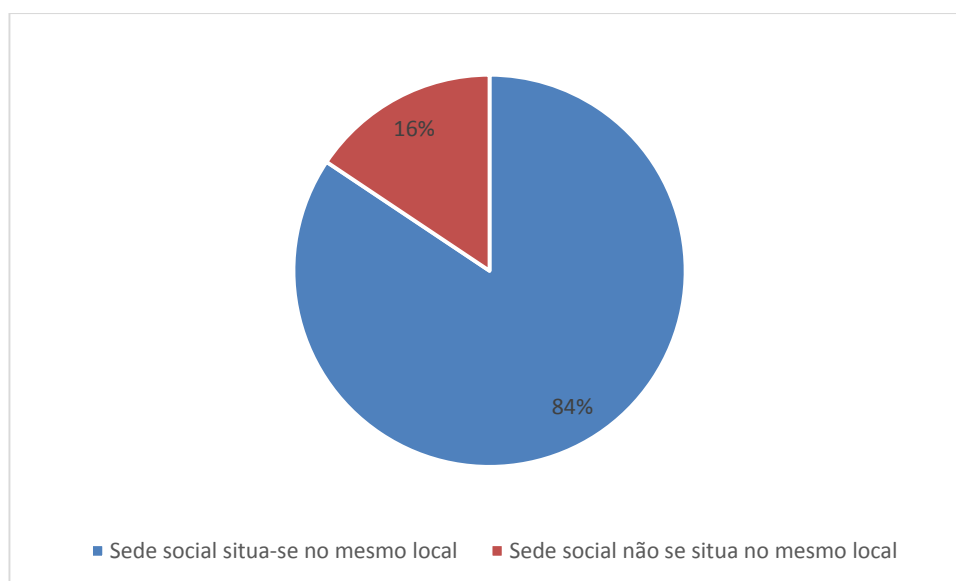


Gráfico 8 – Localização da sede social das empresas privadas

Através do Gráfico 8 é possível observar que para 84% das empresas privadas, a sua sede social se localiza na mesma região geográfica definida anteriormente. As restantes empresas privadas (16%) foram questionadas se a sua sede social se encontra em Portugal ou em outro país, tendo a primeira opção reunido uma taxa de respostas de 100%.

Assim sendo, é possível concluir que todas as empresas privadas que participaram no inquérito têm a sua sede social em Portugal, podendo ou não estar situadas no mesmo local que havia sido selecionado.

3.4 – Âmbito de atuação da empresa

Relativamente à questão sobre o principal âmbito de atuação da empresa ou aquele que é responsável pelo maior volume de negócios obteve-se um maior registo de respostas por parte dos sectores da distribuição de água e da administração local, no caso das empresas públicas.

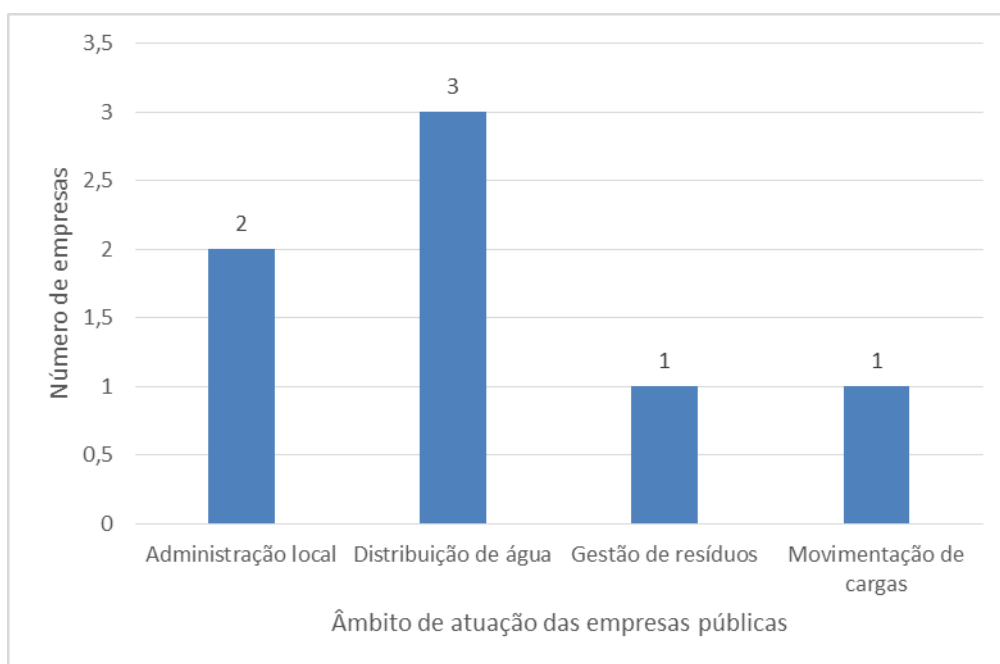


Gráfico 9 – Número de empresas públicas por atividade económica

Já no caso das empresas privadas, registou-se uma maior afluência de respostas nos sectores da indústria de bens alimentares, indústria de produtos químicos, indústria têxtil e tecnologias.

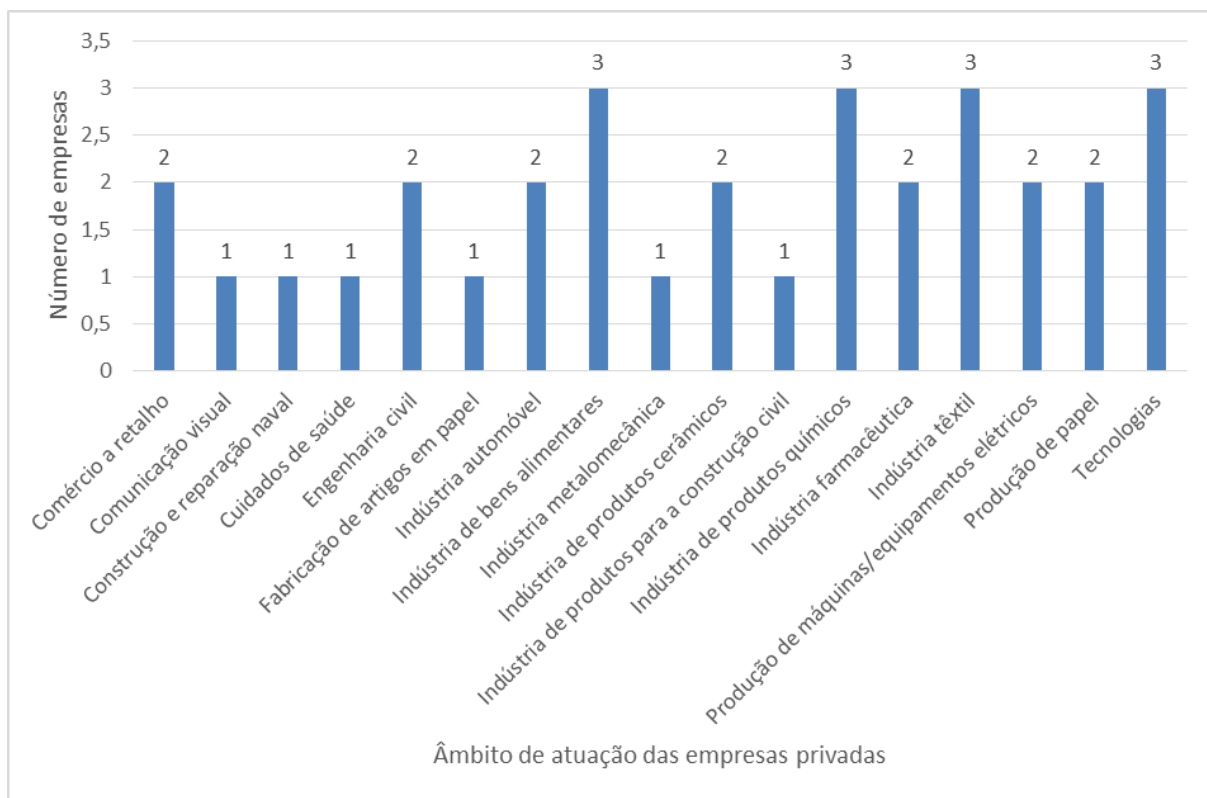


Gráfico 10 – Número de empresas privadas por atividade económica

3.5 – Substituição de documentos em papel por documentos digitais

Na questão sobre a utilização de documentos digitais em detrimento dos documentos em papel, os resultados são esclarecedores: 90% das empresas já têm esse cuidado, ao passo que 10% não assumem essa preocupação, preferindo, ainda, manter os documentos em papel como a principal forma de armazenamento de informação.

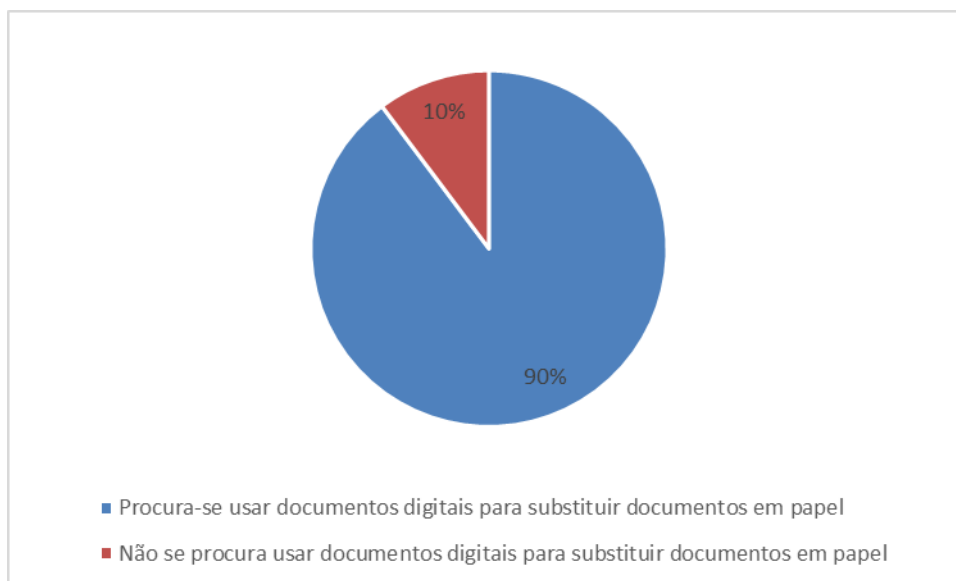


Gráfico 11 – Substituição dos documentos em papel por digitais

De seguida é possível observar uma nova perspetiva sobre esta questão, onde se evidencia a divisão das respostas pelas médias e grandes empresas.

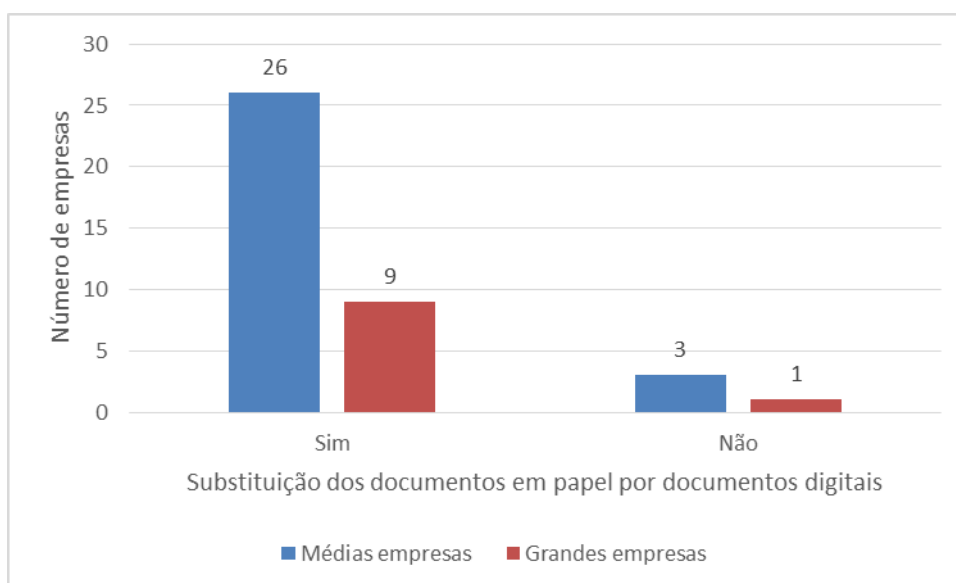


Gráfico 12 – Número de empresas que procuram substituir os documentos em papel por digitais

3.6 – Utilização de *software* de gestão documental

No que concerne à questão relativa à utilização de *software* que permita auxiliar na gestão documental, 59% das empresas afirmam que fazem uso deste tipo de aplicações.

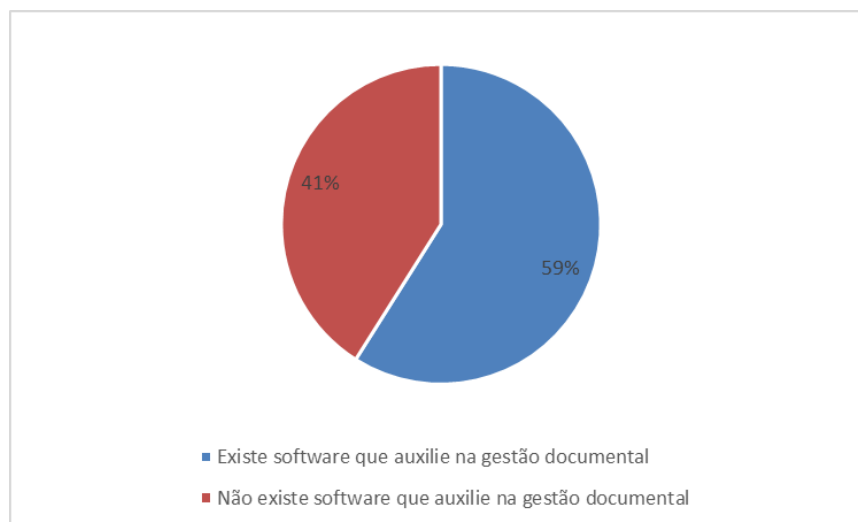


Gráfico 13 – Empresas que utilizam *software* de gestão documental

Para aquelas empresas que afirmaram utilizar *software* de gestão documental, perguntou-se quais os programas utilizados para tal efeito. Para auxiliar esse processo, forneceu-se uma lista com várias opções de escolha e deu-se a hipótese de adicionar programas que não fizessem parte da lista.

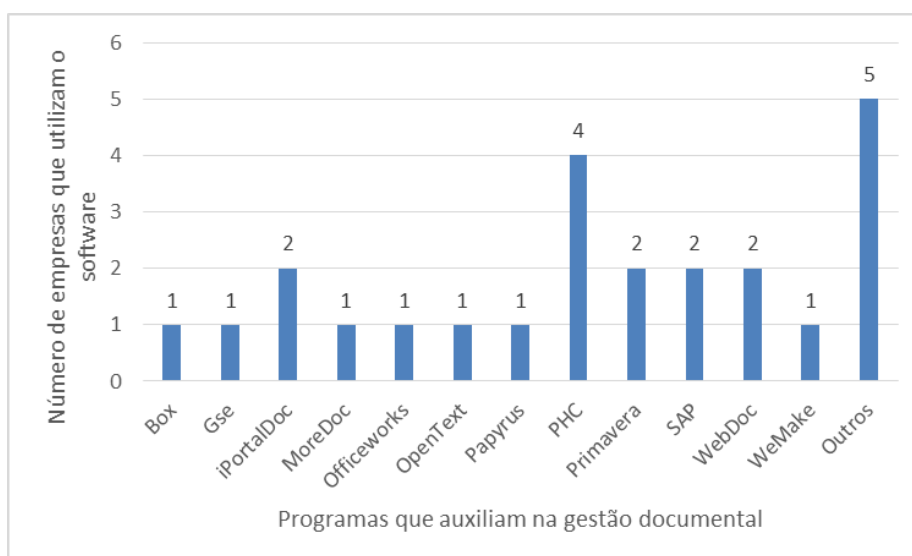


Gráfico 14 – Programas de gestão documental usados pelas empresas

O Gráfico 14 reflete todas as respostas dadas, quer tenham sido selecionadas na lista proposta como submetidas pelas empresas.

Através da sua análise podemos verificar a referência às aplicações Primavera, PHC e SAP. Apesar de se tratarem de aplicações de gestão empresarial, possibilitam a integração com outros módulos que, por sua vez, permitem associar os documentos aos fluxos de negócios existentes na organização.

Relativamente ao *software* de gestão documental, propriamente dito, regista-se o uso do iPortalDoc e do WebDoc em quatro empresas no total.

Contudo, também é possível observar a existência de outros programas. A explicação para a colocação deste campo deve-se a dois fatores: primeiro, algumas das respostas submetidas não se tratam de aplicações de gestão documental; e segundo, não foi possível encontrar informação *online* acerca de algumas respostas submetidas pelas empresas.

3.7 – Recomendação de formatos

Quanto à questão relativa à existência de uma recomendação de utilização de determinados formatos, 67% das empresas responderam positivamente. Ou seja, os objetos digitais devem ser armazenados num certo formato, consoante o tipo de objeto.

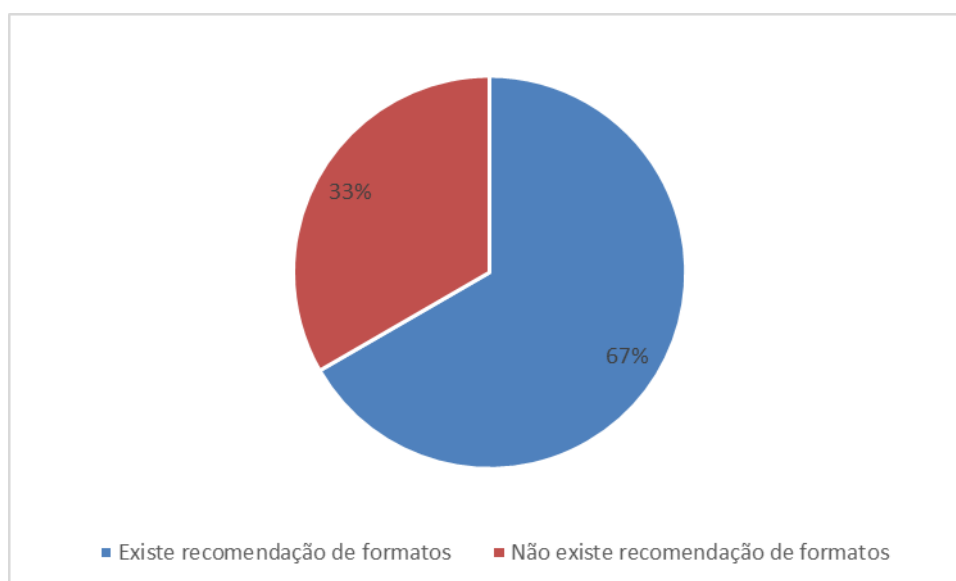


Gráfico 15 – Utilização de formatos recomendados

No próximo gráfico é possível observar a relação entre a existência de formatos recomendados e a dimensão da empresa.

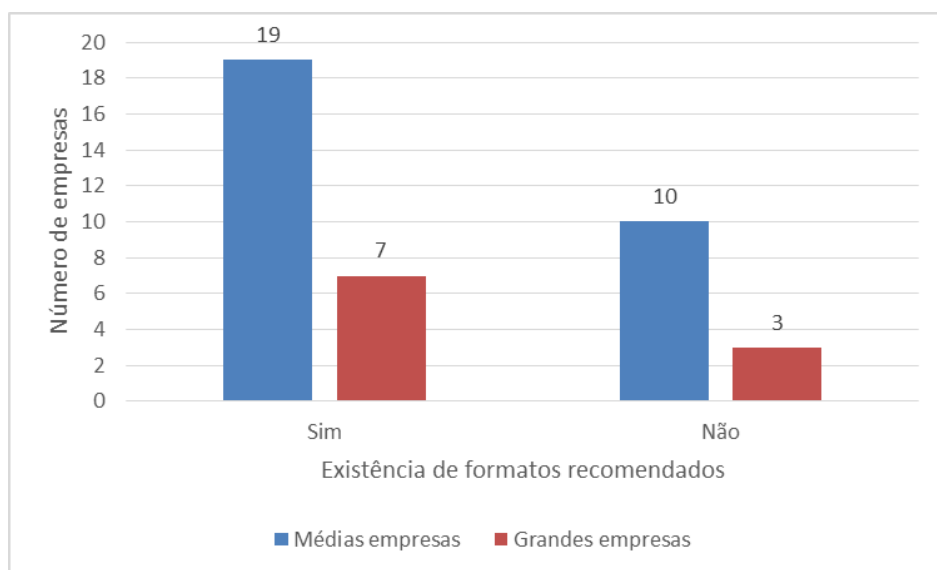


Gráfico 16 – Número de empresas que recomendam o uso de determinados formatos

Ao analisar o Gráfico 16 podemos constatar que 26 empresas sugerem uma lista de formatos para os objetos digitais. A essas empresas foram questionados quais os formatos aconselhados para os quatro tipos de objetos digitais mais habituais: documentos de texto, folhas de cálculo, imagens e *emails*.

3.7.1 – Documentos de texto

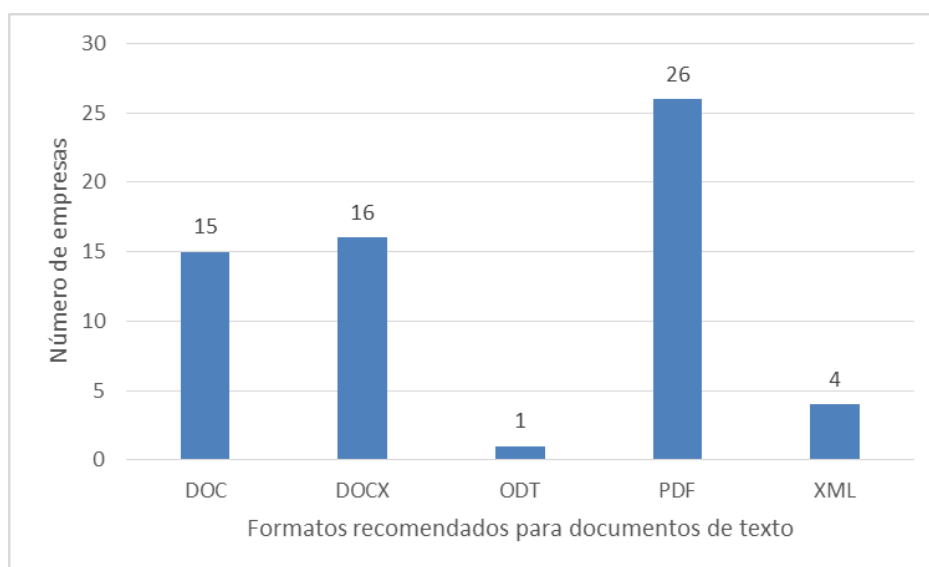


Gráfico 17 – Formatos recomendados para documentos de texto

Relativamente aos formatos aconselhados para documentos de texto, destaca-se o formato normalizado PDF: todas as empresas consideraram este formato como um dos formatos a utilizar para os documentos de texto.

Depois do PDF, os formatos mais utilizados são o DOC e DOCX – pertencentes ao *software* proprietário Microsoft Word.

Já os formatos XML e ODT são os formatos menos referidos pelas empresas.

3.7.2 – Folhas de cálculo

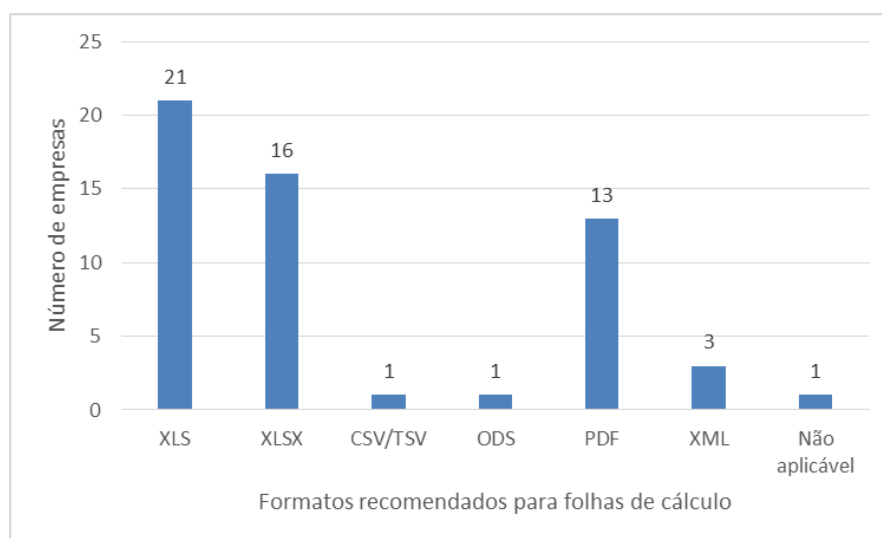


Gráfico 18 – Formatos recomendados para folhas de cálculo

No que concerne às folhas de cálculo, o XLS e XLSX – provenientes do Microsoft Excel – registam um maior número de ocorrências. Quanto ao PDF, é o terceiro formato mais escolhido pelas empresas participantes.

Tal como acontece nos documentos de texto, existe somente 1 registo para os formatos ODS e CSV/TSV. Já o XML é usado por 3 empresas e existe apenas uma que não define formatos para as folhas de cálculo.

3.7.3 – Imagens

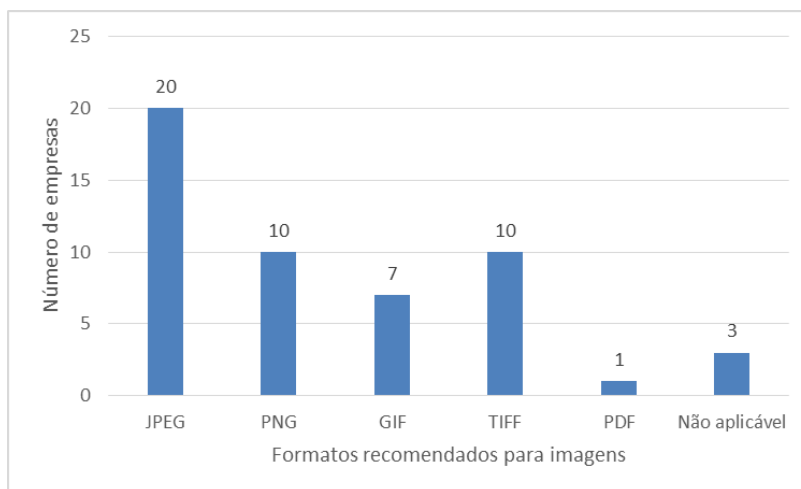


Gráfico 19 – Formatos recomendados para imagens

Para os ficheiros de imagem, o JPEG – definido pela norma ISO/IEC 10918 – surge como principal formato utilizado, seguido do PNG e do TIFF. O PNG trata-se de um formato aberto e recomendado pelo W3C que surgiu como substituto do GIF. Por sua vez, o TIFF, apesar de controlado pela Adobe, possui as suas especificações divulgadas publicamente.

Já o formato GIF é utilizado por 7 empresas, mas – tal como o TIFF – trata-se de um formato proprietário divulgado publicamente.

3.7.4 – Emails

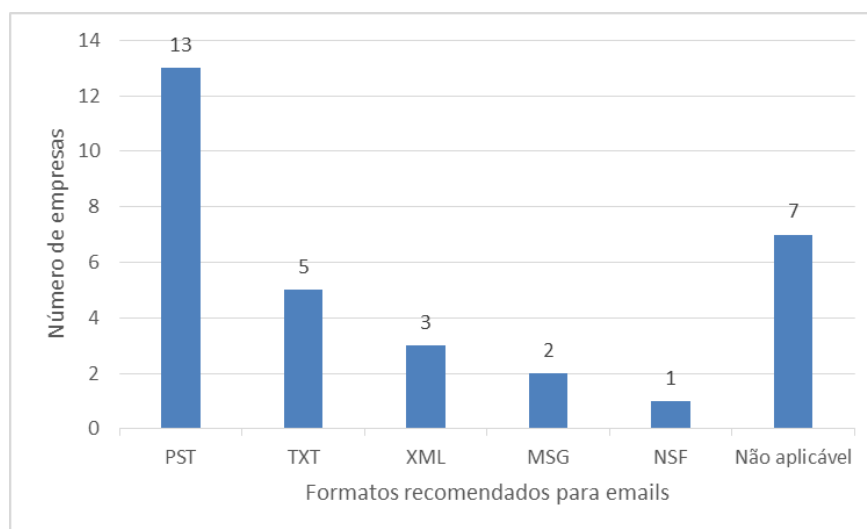


Gráfico 20 – Formatos recomendados para *emails*

Quanto aos formatos utilizados para armazenar *emails*, destaca-se o PST que se trata do ficheiro de dados do Outlook e que permite armazenar *emails* e outras informações (como contactos e compromissos). Contudo o formato MSG, elegível por 2 empresas, é também proveniente desta aplicação, mas apenas armazena os *emails*.

Também é possível verificar que 7 das 26 empresas que possuem recomendação de formatos, não aplicam restrições neste tipo de objeto digital.

3.8 – Dispositivos de armazenamento e cuidados na sua utilização

Relativamente ao *hardware* utilizado para assegurar o armazenamento e cópia dos ficheiros digitais, destaca-se a utilização de discos rígidos por 34 empresas.

Não obstante, existe o sistema RAID (*Redundant Array of Independent Disks*) que permite combinar vários discos rígidos para aumentar a capacidade de armazenamento e a redundância dos dados, o que permite evitar a perda total dos mesmos. Este sistema é utilizado por 16 empresas.

As fitas magnéticas (*tapes*) acedem aos dados de forma sequencial (ao contrário dos discos rígidos), o que reduz a rapidez de acesso aos conteúdos. Contudo, são mais eficientes e duradouras e, por isso, são aconselhadas para sistemas de *backup*. São 13 as empresas que referem utilizar este tipo de tecnologia.

Já os dispositivos óticos (CD e DVD) são usados por um total de 19 empresas.

Mais uma vez se deu a opção de as empresas submeterem diferentes respostas daquelas que estavam previamente definidas. Graças a isso, surgiram duas novas opções: os servidores e a *cloud*.

São 2 as empresas que indicam fazer uso de servidores para armazenar a sua informação digital. No entanto, não especificam o tipo de servidor utilizado.

Curiosamente, e apesar de o armazenamento através de plataformas *online* estar cada vez mais em voga, apenas uma empresa refere a sua utilização.

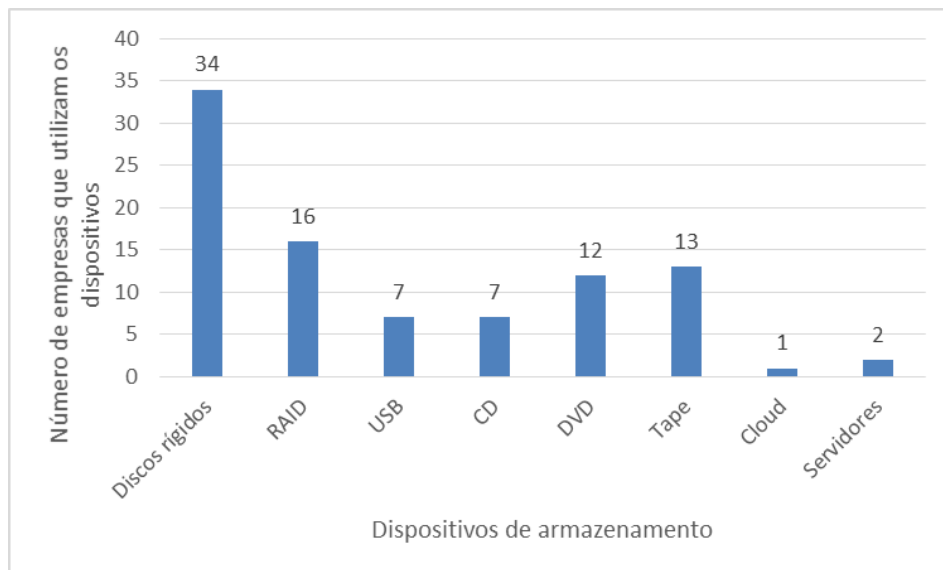


Gráfico 21 – Dispositivos de armazenamento utilizados nas empresas

Como é sabido, o *hardware* responsável por armazenar a informação de uma empresa (e até mesmo a nível pessoal) deve estar sujeito a um conjunto de boas práticas de modo a garantir o seu acesso de forma contínua.

Nesse sentido, as empresas foram questionadas sobre os cuidados tidos com o *hardware* existente. São esses resultados que podemos observar no gráfico seguinte.

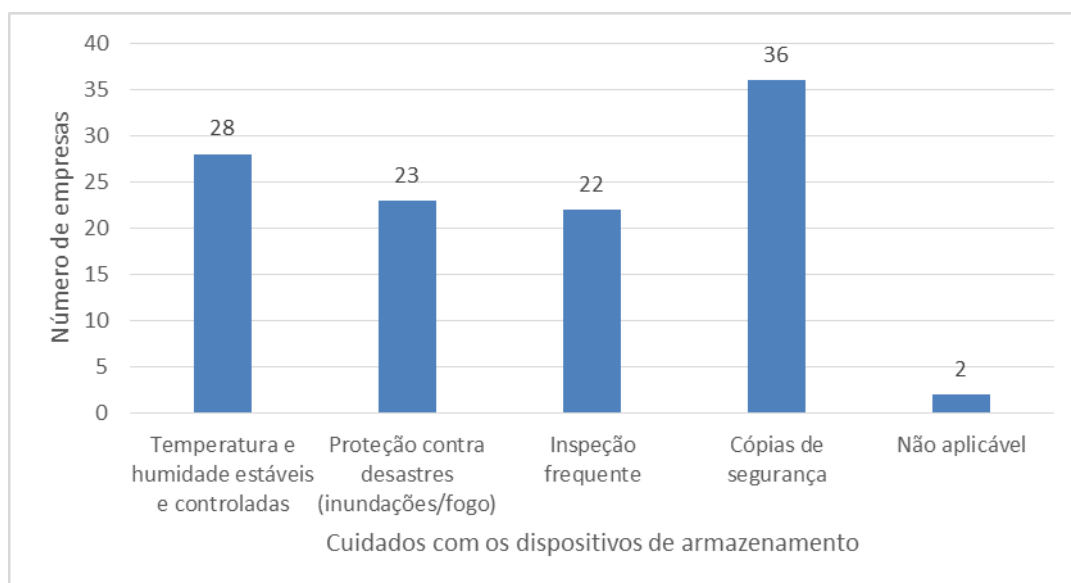


Gráfico 22 – Cuidados com o *hardware* utilizado nas empresas

Como é possível constatar, a realização de cópias de seguranças é a prática mais utilizada por 97% das empresas participantes. Também a verificação do ambiente, nomeadamente da temperatura e humidade, são considerados pelas empresas como um ponto-chave para a sustentabilidade da informação digital a longo prazo.

Já a inspeção frequente dos dispositivos e a implementação de meios prevenção contra desastres são práticas utilizadas por, aproximadamente, 60% das empresas.

No entanto, existe o registo de 2 empresas que referem não aplicar qualquer tipo de diligência para assegurar o bom funcionamento do seu *hardware*.

3.9 – Acesso continuado à informação

As empresas foram questionadas sobre um aspeto da sua cultura organizacional e que influencia o comportamento dos indivíduos que nela trabalham. A questão elaborada teve como objetivo perceber se existe preocupação em aceder à informação por um longo período de tempo.

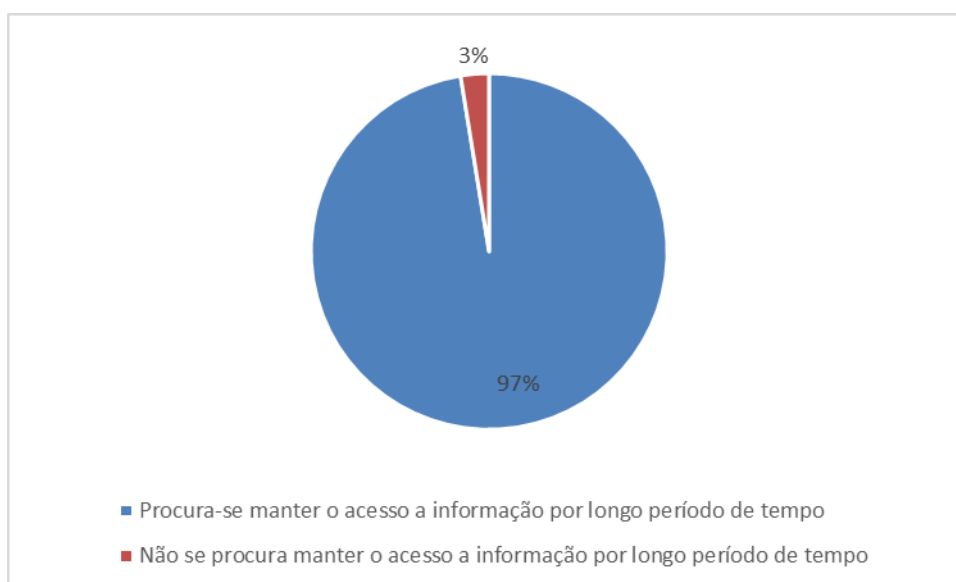


Gráfico 23 – Preocupação em manter o acesso contínuo à informação

Os resultados apresentados no Gráfico 23 são elucidativos: 97% das empresas pretende que a informação criada e utilizada no seu ambiente laboral permaneça acessível por um período de tempo prolongado.

Mas no caso da informação existente em meios tecnológicos há a necessidade de se aplicarem algumas ações para que esse desejo seja cumprido. Nesse sentido, os participantes foram questionados se tinham conhecimento do conceito de preservação digital.

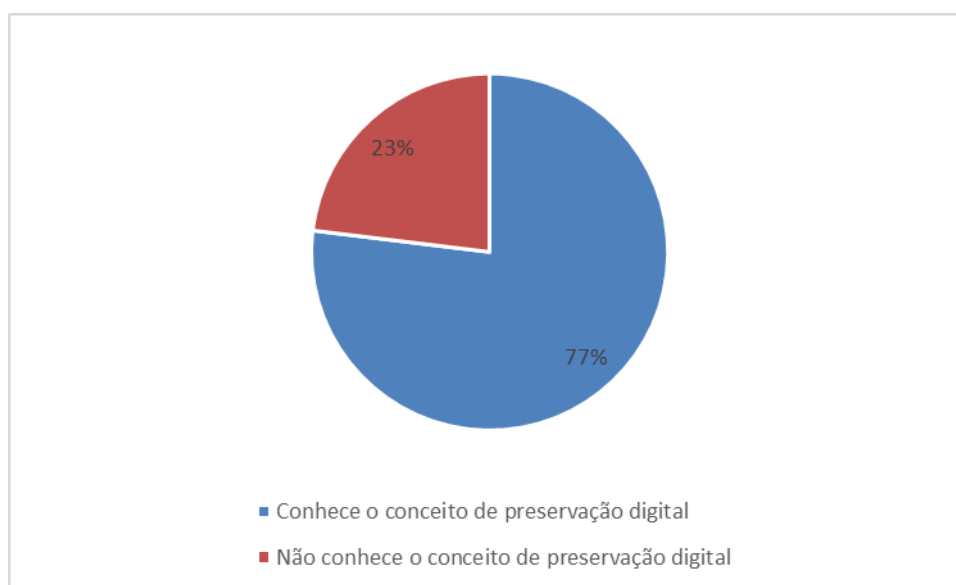


Gráfico 24 – Conhecimento do conceito de preservação digital

Através da observação do Gráfico 24 podemos constatar que 77% dos participantes sabem o que é a preservação digital.

Por sua vez, o próximo gráfico permite verificar a relação entre o conhecimento deste conceito e o tipo de empresa.

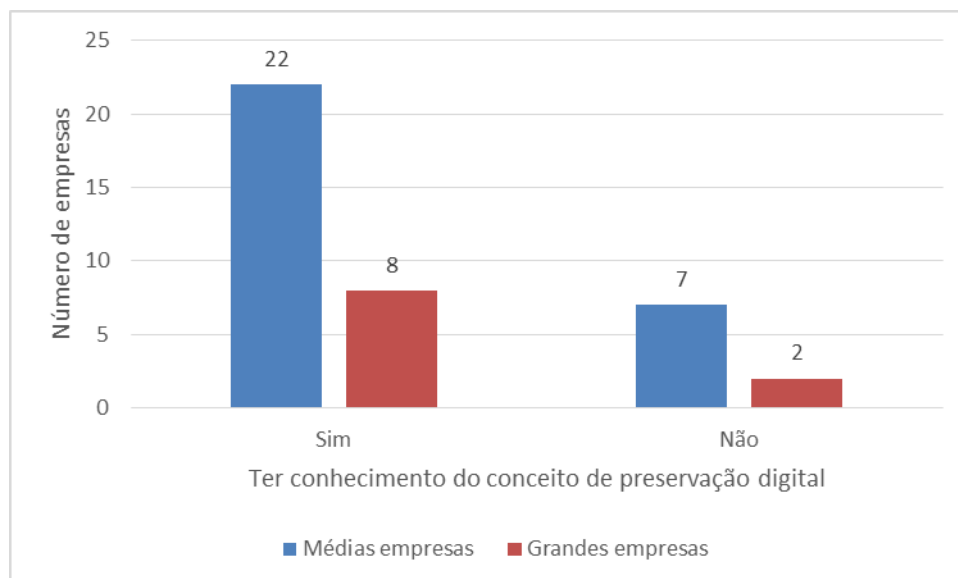


Gráfico 25 – Número de empresas que têm conhecimento do conceito de preservação digital

3.10 – Aplicação de estratégias de preservação digital

Aos participantes foi perguntado se na sua organização já se aplicam estratégias de preservação digital.

Apenas 44% das empresas refere a existência de ações que permitem o acesso continuado à informação em formato digital. Já 33% diz não realizar qualquer tipo de estratégia e 23% afirma não ter conhecimento da aplicação de tais ações.

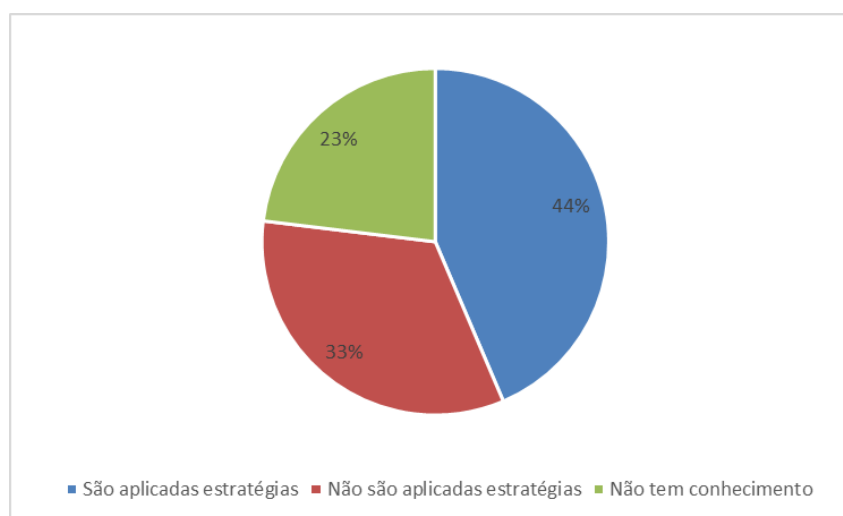


Gráfico 26 – Aplicação de estratégias de preservação digital

No próximo gráfico é possível observar os registos das respostas de acordo com o tipo de empresa.

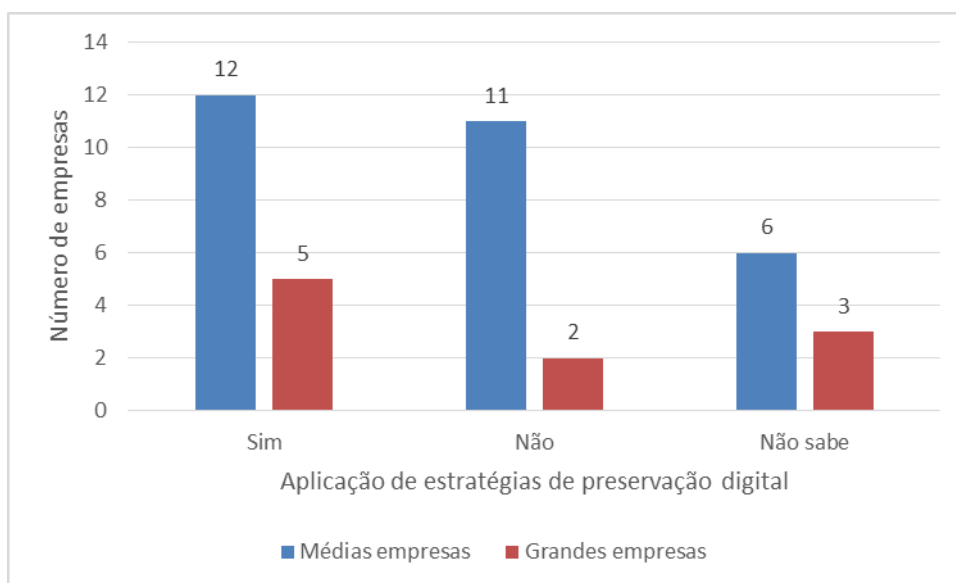


Gráfico 27 – Número de empresas que aplicam estratégias de preservação digital

Para aquelas empresas que afirmaram aplicar estratégias de preservação digital, foram feitas mais algumas questões. Tais questões têm o objetivo de entender as motivações que justificaram a elaboração de um plano de preservação digital e obter mais pormenores sobre a sua implementação.

3.10.1 – Equipa multidisciplinar

As empresas foram questionadas acerca da existência de uma equipa multidisciplinar responsável por estruturar um plano de preservação digital e realizar as intervenções necessárias.

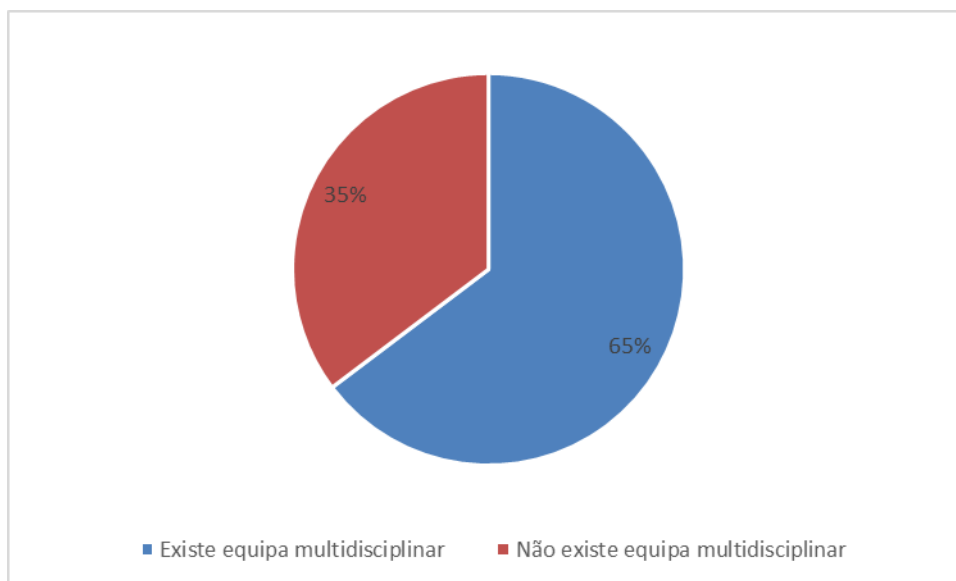


Gráfico 28 – Existência de equipa multidisciplinar para preparar e implementar o plano de preservação digital

É possível verificar a existência de uma equipa multidisciplinar em 65% das empresas participantes, o que significa que profissionais das diversas áreas de cada organização trabalham em conjunto para atingir um objetivo comum.

No entanto, 35% das empresas referem não existir uma equipa multidisciplinar, o que leva a crer que a responsabilidade pela preservação digital esteja apenas focada num departamento. Assim sendo, não se levam em conta as valiosas perspetivas de profissionais de outras áreas e que poderiam trazer benefícios para o plano de preservação digital proposto.

No gráfico seguinte podemos observar a relação das respostas acerca da existência de uma equipa multidisciplinar com o tipo de empresa.

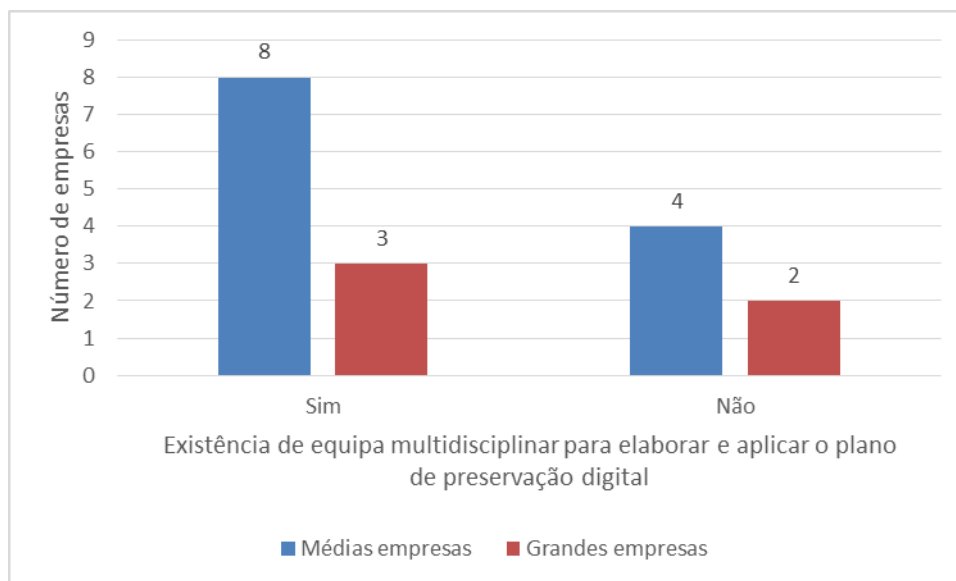


Gráfico 29 – Número de empresas com uma equipa multidisciplinar responsável pela preservação digital

3.10.2 – Responsável pela preservação digital

Para além das informações acerca da existência de uma equipa multidisciplinar, pretendeu-se saber qual o cargo ocupado pelo responsável pela preservação digital na organização. Obtiveram-se as seguintes respostas:

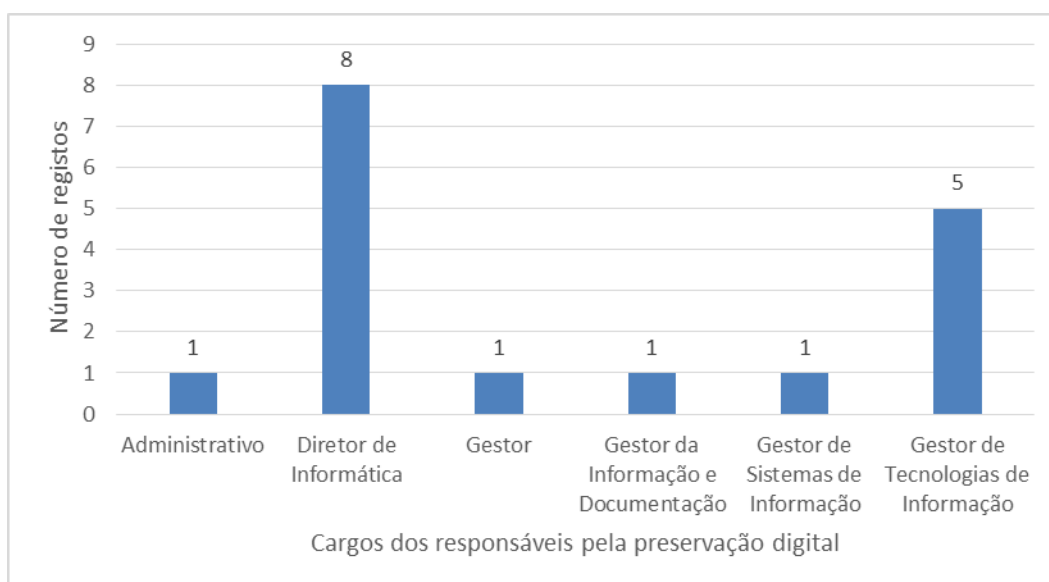


Gráfico 30 – Cargos dos responsáveis pela preservação digital

Consoante as empresas, as denominações atribuídas aos departamentos e aos cargos existentes vão variando. Ou seja, mesmo que as responsabilidades de cada departamento sejam semelhantes em todas as empresas, os seus nomes podem ser distintos. Assim, podemos observar que 8 empresas afirmam que o responsável assume o cargo de Diretor de Informática, mas existem outras 5 empresas que já referem o cargo de Gestor de Tecnologias de Informação. Ou seja, a maior parte dos responsáveis pela preservação digital nas empresas participantes neste estudo trabalha na área dos sistemas tecnológicos de informação que consiste na “plataforma tecnológica – ambiente ou meio físico e lógico – que sustenta a produção, processamento, circulação, armazenamento, transmissão e acesso à informação” (Pinto, 2009^b, p. 25).

Existe ainda um registo para Gestor de Sistemas de Informação e outro para Gestor da Informação e Documentação.

Foram ainda referidos os cargos Administrativo e Gestor, não fazendo qualquer referência ao sector a que pertencem.

3.10.3 – Estratégias de preservação digital

De seguida questionou-se as empresas sobre as estratégias de preservação aplicadas de modo a assegurar a longevidade dos objetos digitais existentes.

A migração foi a estratégia mais referida, logo seguida do refrescamento e da preservação de tecnologia. Já a emulação é a estratégia menos utilizada pelas empresas.

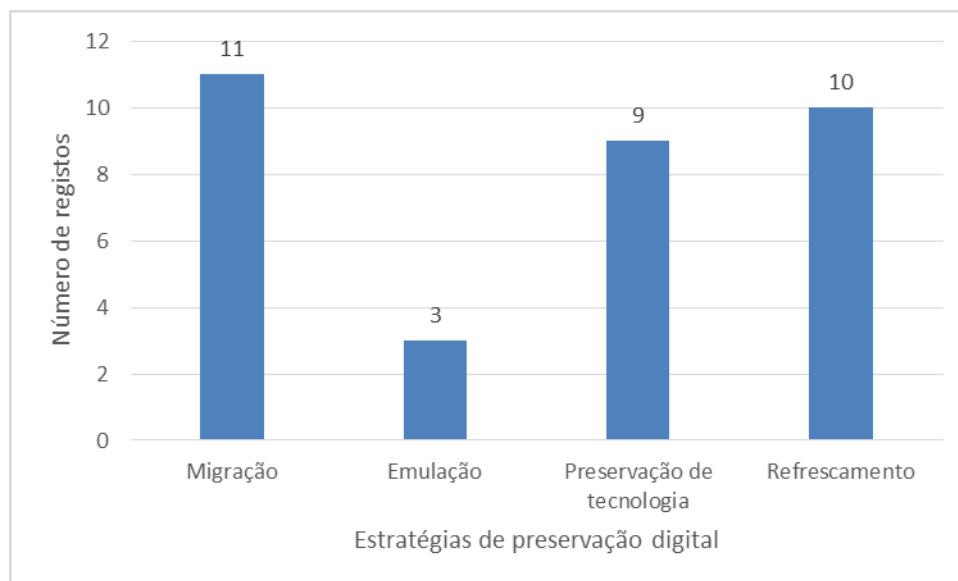


Gráfico 31 – Estratégias de preservação digital aplicadas nas empresas

3.10.4 – Tipos de documentos digitais alvo de preservação

Para além de se querer saber as estratégias aplicadas pelas organizações, também se pretendia conhecer quais os documentos sujeitos às mesmas.

Conclui-se que os documentos que as empresas mais procuram manter acessíveis a longo prazo são os provenientes do sector financeiro e dos recursos humanos. Também outros documentos como contratos, regras e reclamações recebidas são referidos pelas empresas.

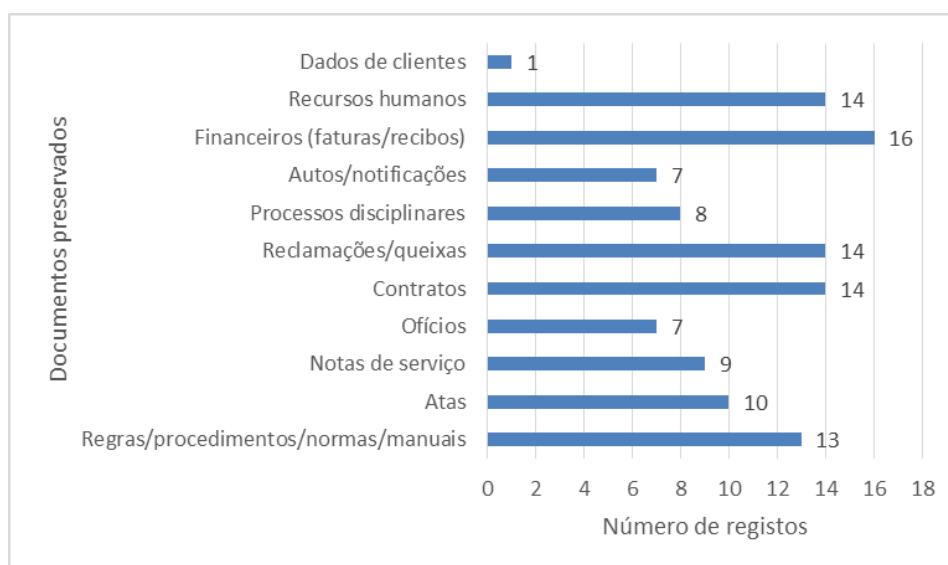


Gráfico 32 – Documentos sujeitos a estratégias de preservação

3.10.5 – Motivação para a preservação digital

Desde logo é importante referir que as empresas são obrigadas, por lei, a conservar determinados documentos durante um certo período de tempo. No final do prazo de vigência administrativa, se o documento tiver valor informativo ou probatório poderá ser permanentemente conservado. Caso disso mesmo é a Portaria nº 412/2001 que estabelece o regulamento arquivístico para as autarquias locais e que apresenta uma tabela de seleção onde podem ser encontradas as diversas séries documentais e os respetivos prazos de conservação administrativa e destino final.

Todavia, apesar da existência dessa obrigação legal, perguntou-se às empresas se essa tinha sido a justificação para manter o acesso dos documentos a longo prazo, ou se o faziam para conservar o seu património histórico-cultural.

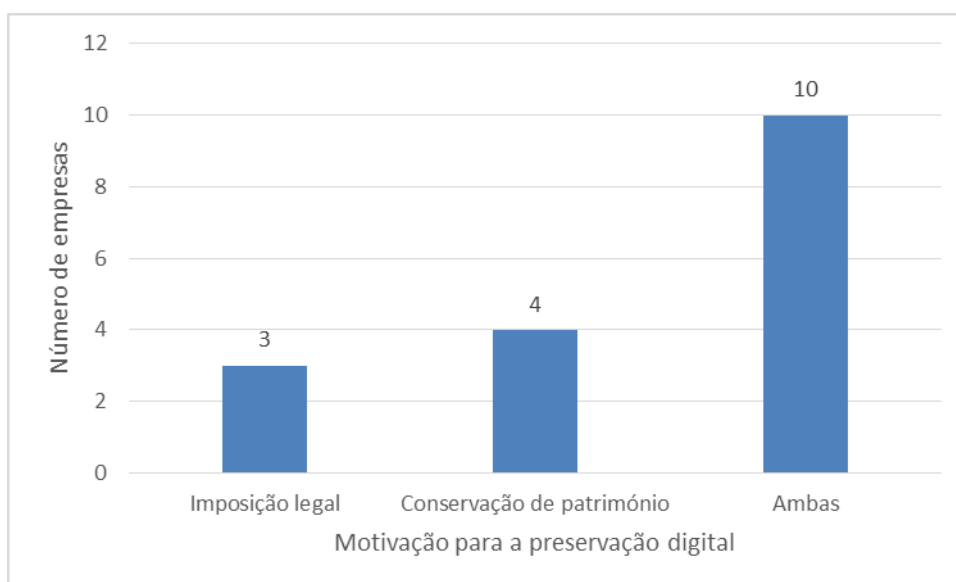


Gráfico 33 – Razões que levaram à implementação do plano de preservação digital

No Gráfico 33 podemos observar que a maioria das empresas participantes consideram ambas as justificações (imposição legal e conservação de património) para o acesso continuado aos documentos.

Continuando a análise, 3 delas referem tê-lo feito somente devido às obrigações legais e 4 para assegurarem a conservação do seu património.

3.10.6 – Investimento realizado

Para terminar as questões feitas às empresas que já possuem um plano de preservação digital, pretendeu-se obter a opinião do participante acerca do investimento realizado (quer em recursos humanos especializados, quer para a infraestrutura tecnológica) para que essa implementação fosse possível.

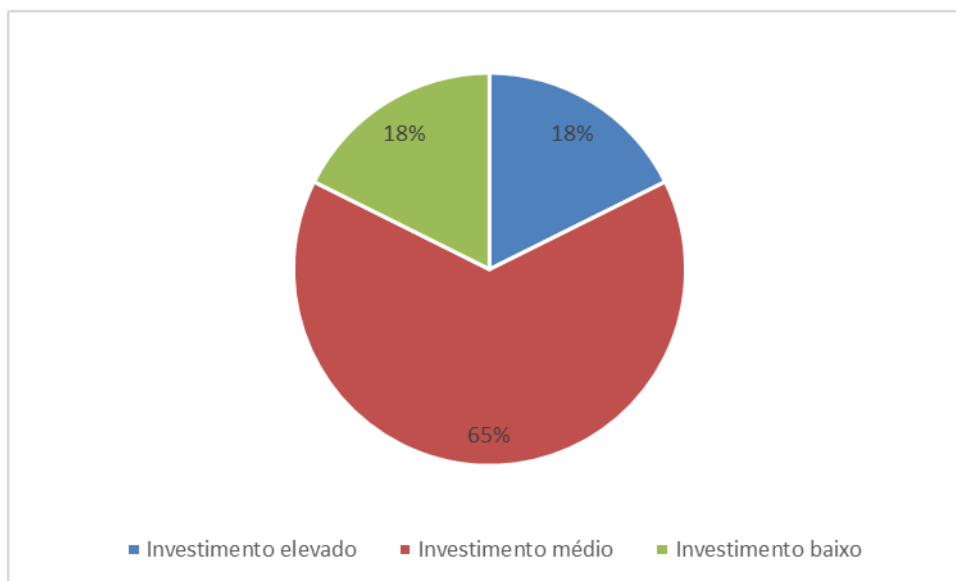


Gráfico 34 – Investimento feito pelas empresas para implementar um plano de preservação digital

A maior parte dos participantes (65%) considerou que o investimento efetuado pela sua empresa foi médio.

É importante referir que aquele que é um investimento elevado para uma média empresa, pode não o ser para uma empresa de grande dimensão, uma vez que se pressupõe que esta última tenha maior capacidade financeira.

3.11 – Desejo de implementação de um plano de preservação digital

Para aqueles participantes que responderam anteriormente que a sua organização não aplica qualquer estratégia de preservação, perguntou-se se gostavam de ver um plano de preservação digital implementado. Os resultados podem ser vistos no próximo gráfico:

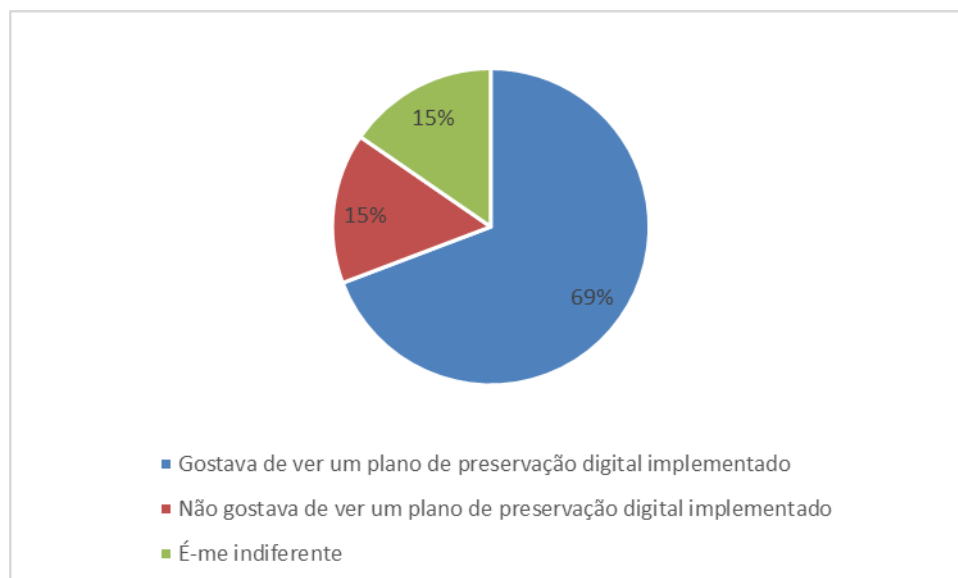


Gráfico 35 – Participantes que gostavam de ter um plano de preservação digital na sua organização

Como podemos constatar através da análise ao Gráfico 35, cerca de 70% dos participantes que não possuem um plano de preservação digital na sua organização gostavam de ver um implementado.

A restante percentagem divide-se por aqueles participantes que não pretendem ou que lhes era indiferente a execução de estratégias de preservação digital na sua organização.

3.12 – Formatos abertos

Para finalizar, questionou-se os participantes sobre um dos fatores mais importantes para garantir o acesso continuado aos objetos digitais: a utilização de formatos abertos.

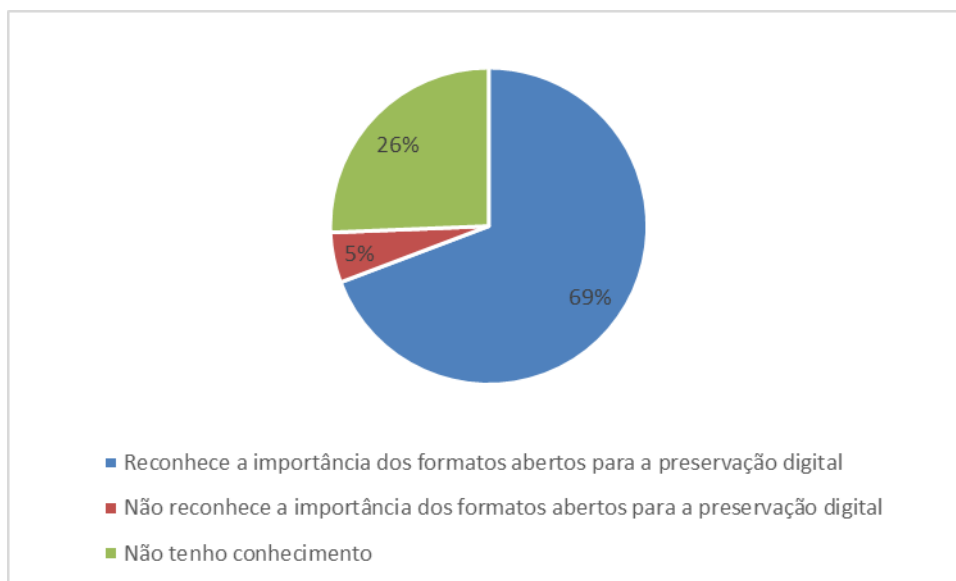


Gráfico 36 – Participantes que reconhecem a importância dos formatos abertos para a preservação digital

Cerca de 70% das empresas reconhecem o papel dos formatos abertos para a preservação digital. Contudo, 5% dizem não reconhecer essa importância e 26% refere desconhecer o assunto.

De seguida, podemos encontrar a relação das respostas dadas com o tipo de empresa.

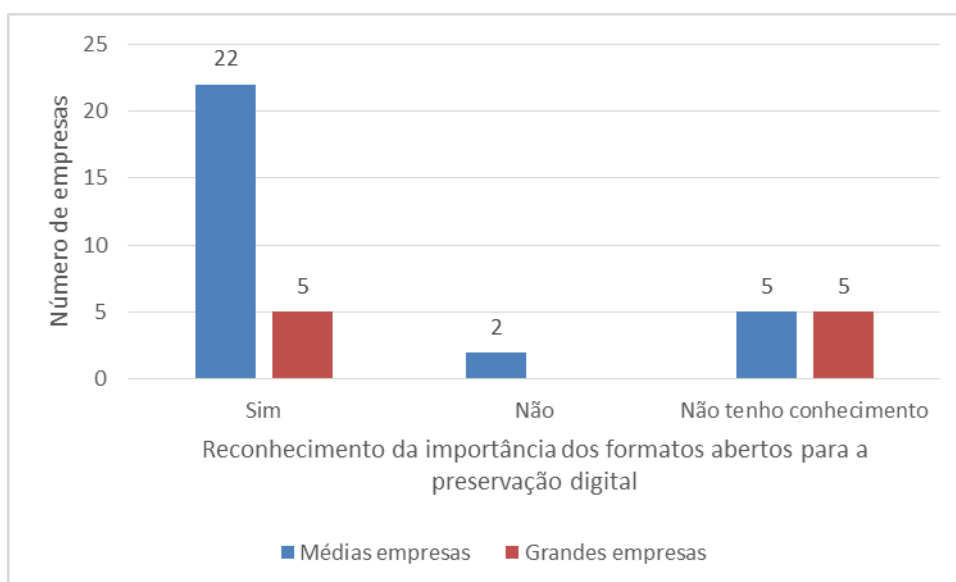


Gráfico 37 – Número de participantes que reconhecem a importância dos formatos abertos para a preservação digital

4 – Conclusões do estudo

Finalizado este estudo, através do envio de questionários *online* para as médias e grandes empresas portuguesas, é possível delinear algumas considerações acerca do estado da informação, da documentação e, conseqüentemente, da preservação digital nas organizações.

Desde logo se evidencia a mínima participação das empresas selecionadas para fazerem parte do estudo. As razões para tal podem ser várias: desde a incapacidade de responder aos vários inquéritos que regularmente surgem; a eliminação imediata do pedido de participação, rejeitando, assim, contribuir para o sucesso de um trabalho académico; o desconhecimento sobre a temática abordada, o que leva a que se descarte ou interrompa o preenchimento do questionário; a inexistência de responsável pela leitura do *email* de contacto⁷¹; ou ainda o simples descuido ou desleixo.

Verificou-se, pois, que a maioria das empresas participantes são de tamanho médio, essencialmente do sector privado e, na sua totalidade, têm a sua sede social em Portugal.

Em virtude quer da colossal evolução tecnológica a que temos assistido, quer da tentativa de redução de custos, grande parte das empresas (90%) refere já procurar utilizar documentos digitais em detrimento dos documentos em papel. Todavia, para que tal utilização seja eficaz, urge a necessidade de recorrer a procedimentos capazes de simplificar a gestão e manutenção dos documentos digitais, nomeadamente através do uso de *software* de gestão documental.

Apesar de o uso e a disseminação de documentos digitais se ter intensificado ao longo dos anos, cerca de 40% das empresas referem não utilizar *software* para auxiliar na gestão dos documentos produzidos. Por sua vez, aquelas que afirmam usar este tipo de programas fazem referência, essencialmente, a sistemas de gestão empresarial que, apesar de não serem indicados para esse efeito, permitem a criação de objetos digitais, nomeadamente de documentos contabilísticos (faturas, recibos, guias, etc.). No entanto, existem alguns programas de gestão documental que podem

⁷¹ Muitos *emails* institucionais são redirecionados automaticamente para um conjunto de pessoas, no sendo de que, cada um, possa responder às solicitações que se enquadram nas suas responsabilidades diárias. Surgindo um pedido fora do comum, é natural que cada um o ignore, na esperança que um outro receptor trate desse pedido (ou porque se ache que outro receptor tenha mais competência para a resposta ou, simplesmente, na esperança de reduzir o seu próprio trabalho).

ser integrados com os sistemas de gestão empresarial, aumentando assim a eficácia nos processos de gestão e manutenção de documentos da organização. Não obstante, também se registaram alguns casos de empresas cujas respostas foram contabilizadas na categoria de *outros programas*, dado não ter sido possível confirmar as respostas dadas.

Relativamente à existência de um conjunto de formatos propostos para os objetos digitais, verificou-se que, aproximadamente, 70% das empresas possui esse tipo de recomendação. Para os documentos de texto destacam-se o formato normalizado e aberto PDF (*Portable Document Format*), logo seguido dos formatos proprietários DOC e DOCX, formatos usados pela aplicação Word, da Microsoft. Também, para as folhas de cálculo, se evidenciam os formatos proprietários XLS e XLSX provenientes do Microsoft Excel. Contudo, neste conjunto, o PDF surge apenas em terceiro lugar. Já para o armazenamento de imagens, os formatos mais usados são o JPEG, PNG e TIFF, tratando-se de formatos cujas especificações são divulgadas publicamente. Por sua vez, os *emails* são maioritariamente armazenados em PST (Outlook), TXT e XML, sendo estes dois últimos formatos abertos. No caso das imagens e *emails* registam-se algumas empresas que não aplicam restrições nos formatos para o armazenamento destes objetos digitais. É possível que tal se deva à baixa produção de documentos nesses formatos. Para concluir, denota-se ainda que para os documentos de texto, folhas de cálculo e *emails*, as empresas ainda se encontram demasiado centradas nos formatos proprietários da Microsoft.

No que diz respeito aos dispositivos responsáveis pelo armazenamento de informação digital, as empresas referem as tecnologias habituais: discos rígidos e sistemas RAID. Apesar do armazenamento *online* estar cada vez mais a impor-se na vida das pessoas, apenas uma empresa referiu usar esse tipo de tecnologia. Acerca dos cuidados tidos com o *hardware*, evidencia-se a realização de cópias de segurança para outros suportes – nomeadamente óticos – de modo a garantir que os dados originais possam ser recuperados em caso de perdas e/ou danos.

No que concerne ao acesso à informação a longo prazo, quase todas as empresas referiram ter essa preocupação. Mas, apesar de a temática da preservação ser conhecida por 75% das empresas, apenas 44% delas dizem aplicar estratégias, com o objetivo de salvaguardar o acesso e manutenção dos objetos digitais.

Relativamente ao planeamento e implementação do plano de preservação digital, 65% das empresas que já aplicam estratégias dizem ter constituído uma

equipa multidisciplinar, enquanto 35% das empresas dizem não o ter feito. Presume-se assim que, no caso das últimas, esta responsabilidade ficou a cargo de um só departamento, descartando as perspetivas que outros profissionais da organização poderiam trazer para este projeto.

O plano de preservação digital naquelas empresas que já o aplicam são, na maioria, liderados por um profissional da área das TIC, registando-se somente um responsável cuja área de especialização é a informação e a documentação.

Quando questionadas acerca de quais as estratégias de preservação digital aplicadas, as empresas distribuíram – de forma quase uniforme – os seus votos pela migração, refrescamento e preservação da tecnologia. Contudo, a migração foi mesmo a estratégia mais escolhida e, por isso, verifica-se uma das hipóteses desta dissertação:

A estratégia de preservação digital mais utilizada pelas médias e grandes empresas portuguesas é a migração.

Quanto à motivação para implementar um plano de preservação digital, grande parte das empresas dizem tê-lo feito por duas razões. Uma das justificações é a conservação do seu património, permitindo que, mais tarde, os documentos criados possam ser consultados e possam servir de testemunho das atividades realizadas e do conhecimento gerado. A segunda razão deve-se às obrigações que as empresas estão sujeitas de acordo com a legislação do país, o que pode justificar a escolha dos documentos financeiros como aqueles que as empresas mais procuram preservar.

No que concerne ao investimento efetuado para implementar o plano de preservação digital, 66% das empresas diz ter sido um investimento médio. Contudo, uma vez que se consideram empresas de média e grande dimensão, as suas capacidades financeiras são distintas e o investimento realizado por uma empresa pode ser significativo para o seu orçamento mas não o ser para outra.

No que diz respeito ao reconhecimento dos benefícios dos formatos abertos para a preservação digital, 70% das empresas refere estar ciente dessa importância. Assim sendo, podemos dizer que a segunda hipótese desta dissertação também pode ser verificada:

As médias e grandes empresas portuguesas reconhecem a importância dos formatos abertos para a preservação digital.

Contudo, devemos atentar neste resultado e compará-lo com a definição dos formatos usados no armazenamento de documentos de texto, folhas de cálculo, imagens e *emails*: os resultados são contraditórios. Isto é, apesar de as empresas reconhecerem as vantagens dos formatos abertos, na prática estes não são amplamente utilizados. Com exceção feita ao PDF – que se pode concluir já estar devidamente incorporado na mentalidade das organizações –, a maior parte das empresas faz referência a formatos proprietários nos documentos de texto, folhas de cálculo e *emails*. Seria expectável que, em pleno século XXI e numa era marcada pela crise económico-financeira em todo o Mundo, o uso de formatos abertos e de *software* de código aberto fosse superior, até porque a utilização de *software* proprietário implica a aquisição de licenças, situação esta que poderia ser evitada.

Outra conclusão é que 70% dos participantes que dizem não ter um plano de preservação digital implementado, gostavam que essa situação fosse alterada, o que vai de encontro ao que já foi referido anteriormente, nomeadamente com as altas percentagens relativas à substituição dos documentos em papel por digitais e à preocupação em manter o acesso prolongado à informação.

Ao contrário do que era expectável no início da elaboração do estudo, as organizações já se encontram despertas para os benefícios da informação em formatos digitais mas também da sua vulnerabilidade, e entendem que para garantir a sua sustentabilidade ao longo do tempo não podem permanecer paradas e devem aplicar determinadas ações que evitem essa situação. Contudo, a utilização de *software* de código aberto e o uso de aplicações de gestão documental deviam também já ser prática comum, algo que não se verifica.

Conclusão

São inegáveis as vantagens que as Tecnologias da Informação e Comunicação (TIC) introduziram nas nossas vidas. Hoje é possível iniciar, facilmente, uma comunicação com qualquer parte do Mundo ou registar um momento através de uma fotografia e/ou vídeo nas redes sociais, bastando para isso ter um dispositivo com acesso à Internet. Mas as aplicações das TIC não se limitam apenas à conversação, ou partilha de fotografias ou vídeos.

Ao longo da nossa vida – seja a nível pessoal, académico ou profissional – somos responsáveis pela criação de um vasto número de objetos digitais de diferentes tipos. Para além da criação destes objetos, as TIC também permitem um fácil e rápido processamento de uma quantidade cada vez maior de dados, bem como a sua multiplicação, armazenamento e disseminação que, por sua vez, possibilita a partilha de conhecimento.

Infelizmente, alguns dos problemas já existentes com objetos físicos, como a preservação do papel de uma fotografia, também surgiram nas TIC. Basta que queiramos garantir o acesso futuro aos documentos digitais criados, e a gestão de objetos digitais torna-se difícil. Os problemas incluem a obsolescência tecnológica ou a cópia desregrada de um objeto digital para vários suportes (o que pode levar a que não seja possível garantir a sua autenticidade, integridade e fidedignidade).

Surge, então, a necessidade da definição e da aplicação de ações de preservação digital, isto é, um conjunto de ações mais ou menos complexas que venham a garantir a acessibilidade e autenticidade de um objeto em formato digital, por tempo indeterminado. Cabe às gerações atuais assumirem a responsabilidade de realizar as ações necessárias para que, no futuro, os objetos armazenados em formato digital possam ser interpretados com recurso ao *hardware* existente na época.

Nesta dissertação realizou-se uma análise ao estado atual da preservação digital nas médias e grandes empresas portuguesas. Abordou-se esta perspetiva empresarial da preservação digital de modo a colmatar uma falha existente nos estudos atuais, tendo em conta que grande parte da literatura se dedica a estudar outro tipo de instituições, como as bibliotecas ou arquivos.

A metodologia aplicada passou pelo estudo da temática de preservação digital, tendo em conta os diferentes tipos de ações sugeridas pela comunidade;

posteriormente, analisou-se o tecido empresarial português; e com os dados recolhidos delineou-se um conjunto de conclusões que refletem a posição das empresas portuguesas em relação à problemática da preservação digital.

Ao longo deste trabalho salientou-se a importância da aplicação de estratégias ou boas práticas de preservação digital no decorrer do ciclo de vida de um objeto. Tais ações devem ser reunidas e descritas num documento ao qual se atribui a denominação de *plano de preservação digital*. Este plano deve ser produzido por uma equipa multidisciplinar que deve, previamente, efetuar uma análise ao ambiente interno da organização fazendo o levantamento dos recursos estruturais, tecnológicos e humanos existentes, e equacionar a necessidade de investir em algum destes quadrantes de modo a que a implementação do plano de preservação digital seja possível. Nesse sentido, é fulcral que a preservação digital esteja prevista na gestão estratégica de uma organização, uma vez que existe a necessidade da alocação de recursos quer humanos quer financeiros.

Esta equipa multidisciplinar deve ser responsável por identificar as necessidades informacionais de cada departamento da organização e deve definir um conjunto de procedimentos que assegurem a manutenção da informação armazenada em formato digital por um período de tempo prolongado. Nesses procedimentos devem constar, entre outros, a escolha do *software*, formatos e que metainformação deverá acompanhar os objetos digitais.

Relativamente à escolha do *software*, deve-se optar, sempre que possível, pela utilização de *software* de código aberto cujas especificações estejam divulgadas de forma pública e de forma gratuita.

No que concerne aos formatos, também eles devem ser abertos, baseados em normas internacionais, bem documentados e comumente utilizados. A equipa multidisciplinar deve restringir a diversidade de formatos nos quais os objetos digitais podem ser armazenados no sentido de diminuir a quantidade de processos necessários para a sua preservação. No caso de objetos que ainda precisem de edição após a sua criação, aconselha-se o uso de formatos como o ODT e ODS, para documentos de texto e folhas de cálculo, respetivamente. Já o formato PDF/A (adaptação do PDF que propicia o armazenamento a longo prazo de objetos digitais) pode ser aplicado aos documentos que já tenham atingido o propósito da sua criação e assumam carácter probatório e definitivo. Finalmente, a utilização de formatos

baseados em XML garante a possibilidade de edição futura, bem como a interoperabilidade entre diferentes plataformas.

No que diz respeito à metainformação, a equipa multidisciplinar deve definir o esquema de metainformação que mais se adequa às suas necessidades e deve, também, definir a nomenclatura que deve ser utilizada para que todos os colaboradores possam descrever os diversos atributos de acordo com os termos pré-definidos (definição de uma linguagem controlada).

Dada a evolução tecnológica, o criador de um objeto digital assume um papel de grande relevância em todo o processo de preservação. Algumas das ações referidas anteriormente (como o armazenamento do objeto num determinado formato ou a adição de metainformação associada) devem ser aplicadas o mais cedo possível de modo a que o objeto digital não esteja sujeito, mais tarde, a transformações desnecessárias. Para isso, os colaboradores de uma organização devem ser alertados para a importância da aplicação de um conjunto de boas práticas que, por sua vez, permitem uma gestão mais eficaz da informação digital. Tal consciencialização pode ser feita através de ações de formação com o objetivo de evitar a aversão à mudança devido à implementação dos novos procedimentos de trabalho.

Relativamente às estratégias de preservação digital, cabe à equipa multidisciplinar responsável pelo plano de preservação digital definir que estratégias devem ser aplicadas ao longo da vida de cada tipo de objeto digital existente na organização. Cada estratégia acarreta um conjunto de vantagens e desvantagens, pelo que se aconselha a utilização combinada de estratégias de modo a que seja possível prolongar a acessibilidade dos objetos digitais. Deve-se, ainda, elaborar um plano que defina o(s) responsável(is) por cada estratégia.

Neste trabalho foram também descritas algumas das iniciativas mais importantes a nível internacional para a preservação digital: o modelo de referência OAIS e o MoReq2010. O OAIS é considerado por muitos autores como a iniciativa mais importante para a preservação digital, pois foi responsável pela definição de uma nomenclatura que é agora amplamente adotada e que visa estabelecer um conjunto de orientações para a implementação de repositórios digitais. Já o MoReq2010 procura definir as especificações necessárias para a implementação de sistemas de gestão de documentos de arquivo que, por sua vez, devem ser tidas em consideração pelas empresas responsáveis no desenvolvimento deste tipo de *software*. A existência

de *software* criado com base em normas europeias, que visam auxiliar todo o processo de gestão documental, é uma mais-valia para todas as organizações que dele usufruam, pois assim é possível maior homogeneidade entre os processos usados e facilita a reutilização eficiente de abordagens e *software* entre diferentes instituições.

A informação produzida em ambiente laboral é de extrema relevância para qualquer organização, independentemente do seu tamanho ou do tipo de atividade desenvolvida. Este trabalho foca-se na informação registada em documentos digitais e que comprova as atividades realizadas por uma organização, sejam elas de âmbito administrativo, contabilístico ou pessoal. A informação existente nesses documentos reflete o capital intelectual dos colaboradores da organização. O facto de constituírem um meio de prova faz com que muitos dos documentos criados dentro da organização tenham de ser conservados por vários anos. Isso leva a que seja necessário aplicar ações adequadas que permitam garantir a sua manutenção e acessibilidade por longos períodos temporais. No entanto, uma organização só será capaz de implementar ações de preservação de documentos, sejam eles em suporte tradicional ou digital, caso já tenha compreendido a importância que uma correta gestão da informação e da documentação pode trazer ao seu quotidiano.

Realizou-se um estudo – dividido em três partes – com o intuito de entender o ponto de situação da preservação digital nas médias e grandes empresas portuguesas. Inicialmente obteve-se as informações necessárias para caracterizar o perfil da empresa. De seguida, colocaram-se questões de modo a entender como a organização procede à gestão dos seus documentos digitais bem como do *hardware* responsável pelo seu armazenamento. Por último, foram obtidas informações sobre a posição da empresa em relação à preservação digital, nomeadamente em relação à existência de uma equipa multidisciplinar responsável pela sua implementação e às estratégias aplicadas.

O estudo de caso foi realizado através de um questionário *online*, enviado a 1.358 médias e grandes empresas portuguesas. A taxa de respostas foi mínima, mas, mesmo assim, optou-se por prosseguir com a sua análise. No entanto, ressalva-se que os resultados obtidos podem não refletir a realidade das empresas portuguesas já que a amostra não tem significância estatística. De seguida apresentam-se as principais conclusões do estudo realizado.

A maioria das empresas participantes são de dimensão média e atuam no sector privado. A localização da sede social de todas as empresas é em Portugal. No que concerne à substituição de documentos em papel por documentos em formato digital, 90% das empresas já referem fazê-lo. No entanto, apenas 60% referem utilizar *software* de gestão documental, o que mostra que apesar de as empresas já se terem rendido aos objetos digitais, a sua gestão pode ainda não ser muito eficaz.

Cerca de 70% das empresas dizem possuir um conjunto de formatos recomendados para os diferentes tipos de objetos digitais. O formato PDF já é de utilização generalizada para os documentos de texto e folhas de cálculo. Já nas imagens, o formato mais utilizado pelas organizações é o JPEG. Contudo, ainda impera o uso de formatos proprietários para estes objetos digitais e para o *email*.

Quase todas as empresas (97%) afirmam ter preocupação em manter a informação acessível a longo prazo, mas apenas 77% têm conhecimento da temática da preservação digital.

Menos de metade dos participantes (44%) referem ter conhecimento da aplicação de estratégias de preservação digital na sua organização. Dentro destes, 65% indicam que a responsabilidade pela sua implementação está a cargo de uma equipa multidisciplinar. A migração demonstra ser a estratégia de preservação digital mais utilizada, verificando-se assim uma das hipóteses colocadas no início desta dissertação. Questionadas sobre os tipos de documentos que são alvo destas ações, as empresas referem que os mais preservados são os financeiros, contratos e outros relativos aos recursos humanos. A maioria das empresas procura manter os seus documentos digitais acessíveis a longo prazo devido às imposições legais existentes no país e como forma de conservar o seu património histórico-cultural com vista à passagem de testemunho às gerações futuras.

Considerando os participantes que referiram não ter, ainda, um plano de preservação digital implementado na sua empresa, 70% destes afirmam que esta seria uma boa aposta por parte dos seus administradores.

Finalmente, também é possível concluir que 70% dos participantes reconhecem a importância que o uso de formatos abertos tem para a preservação digital, comprovando-se, assim, a segunda hipótese da dissertação. No entanto, a realidade é contraditória já que, através da análise feita aos formatos realmente utilizados se observou que, à exceção dos formatos PDF e JPEG, o uso de formatos proprietários ainda é o mais usual.

Do ponto de vista crítico, pode-se considerar que urge uma mudança na atitude das organizações e dos seus colaboradores em relação à informação e à documentação. Apesar de se comprovar que as organizações procuram utilizar documentos em suporte digital, em detrimento do suporte papel, e que se preocupam em manter o seu acesso por tempo prolongado, ainda existe um longo caminho a percorrer na gestão de objetos digitais. A adoção de *software* de código aberto e de formatos abertos, sempre que possível, potencia o sucesso de qualquer plano de preservação digital e minimiza o investimento feito pelas organizações, tendo em conta que não é necessária a aquisição de licenças. Embora esta seja a situação ideal, não nos podemos esquecer que também *software* proprietário – como Microsoft Word ou Excel – permite o armazenamento de objetos digitais em formatos abertos.

Com este estudo demonstrou-se, até certo ponto, que a preservação digital não é uma preocupação apenas de bibliotecas, arquivos ou instituições de ensino superior. Também as empresas assumem um papel ativo na manutenção e acessibilidade de objetos digitais, de modo a preservar a informação criada no âmbito da sua laboração e que constitui um património histórico das atividades desenvolvidas.

A principal limitação no decorrer da elaboração deste trabalho de investigação foi a mínima participação das empresas que foram selecionadas. Foram enviados vários pedidos, ao longo do tempo em que o questionário esteve disponível, para (re)lembrar as empresas, mas de pouco serviram. Importa salientar, como aliás já foi dito, que a baixa taxa de respostas faz com que a análise efetuada possa ser enviesada.

Como propostas para trabalhos futuros, podemos considerar o alargamento da análise do estado da preservação digital a outras instituições e/ou a imposição de restrições a nível geográfico. Por exemplo, elaborar um estudo às PME, uma vez que são estas empresas as grandes responsáveis pela criação de emprego e por alavancar a economia do nosso país. Outro exemplo pode ser a restrição a uma determinada área de negócio, como os jornais. Estes órgãos de comunicação social produzem grandes quantidades de informação diariamente, seja em forma de texto, fotografia ou vídeo. Seria interessante perceber as ações realizadas e que permitem manter estes objetos digitais acessíveis, já que, muitas vezes, são recuperados no futuro para diferentes finalidades.

Apesar do grande objetivo desta dissertação se centrar na análise do estado da preservação digital nas médias e grandes empresas portuguesas, reuniu-se um conjunto de documentos que apresentam recomendações, redigidas por entidades nacionais e internacionais, que procuram auxiliar na preservação de material digital e que complementam aquilo que veio sendo dito neste trabalho. Estes documentos foram compilados e anexados neste documento. No Anexo A podemos encontrar boas práticas para o armazenamento de objetos digitais. No Anexo C está presente uma lista de vantagens e desvantagens de algumas das estratégias de preservação referidas neste trabalho. No Anexo D referem-se alguns dos critérios que permitem uma correta seleção dos formatos para o armazenamento dos objetos digitais e, por fim, no Anexo E podemos encontrar uma lista dos formatos mais adequados para garantir a preservação dos objetos digitais.

Para além disto, elaborou-se uma síntese com um conjunto de ações que foram sendo referidas ao longo deste trabalho. Com esta síntese – que pode ser encontrada no Anexo I – pretende-se contribuir para esta área da documentação com algumas orientações a todos aqueles que procurem preservar, por tempo indeterminado, os seus objetos digitais.

Referências bibliográficas

AGÊNCIA PARA A COMPETITIVIDADE E INOVAÇÃO – **Definição de PME** [Em linha]. Lisboa: IAPMEI, 2007. [Consult. 08 jan. 2015]. Disponível em www: <<http://www.iapmei.pt/iapmei-art-03.php?id=1790>>.

ALMEIDA, Francisco Alberto Severo de; SILVA, Armando Malheiro da; GUIMARÃES, António Teodoro Ribeiro – O modelo quadripolar aplicado à educação mediada por tecnologia da informação e comunicação: um estudo empírico. **Prisma.com** [Em linha]. 16 (2013), 1-21. [Consult. 05 mai. 2014]. Disponível em www: <<http://repositorio-aberto.up.pt/bitstream/10216/64351/2/arandomalheiromodelo000171341.pdf>>. ISSN 1646-3153.

ANBU, K. John Paul; CHIBAMBO, Marion L. N. – Digital Preservation: Issues and challenges. **Trends in Information Management** [Em linha]. 5:1 (2009), 42-58. [Consult. 17 mar. 2014]. Disponível em www: <<http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=52247976&lang=pt-br&site=ehost-live>>. ISSN 0973-4163.

ANTÓNIO, Rafael – A preservação digital através do MoReq2010. In **Encontro de CTDI**, 8, Vila do Conde. Organizações digitais: entre informação e conhecimento [CD-ROM]. Vila do Conde: Escola Superior de Estudos Industriais e de Gestão, 2014, 1-10.

ARAÚJO, Carlos Alberto Ávila – Correntes teóricas da ciência da informação. **Ciência da Informação** [Em linha]. 38:3 (2009), 192-204. [Consult. 05 mai. 2014]. Disponível em www: <<http://www.scielo.br/pdf/ci/v38n3/v38n3a13.pdf>>. ISSN 1518-8353.

ARAÚJO, Francisco de Assis Noberto Galdino; PINTO, Manuela – D4SiMem: uma proposta de modelo de digitalização para sistemas memoriais. **Ciência da Informação** [Em linha]. 41:1 (2012), 127-139. [Consult. 19 mai. 2014]. Disponível em www: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2124/1798>>. ISSN 1518-8353.

ARQUIVO DA WEB PORTUGUESA – **Formatos adequados para preservação** [Em linha]. Lisboa: FCCN, 2009. [Consult. 12 jan. 2014]. Disponível em [www: <http://sobre.arquivo.pt/colaboracoes/recomendacoes-para-autores-de-sitios-web/formatos-adequados-para-preservacao>](http://sobre.arquivo.pt/colaboracoes/recomendacoes-para-autores-de-sitios-web/formatos-adequados-para-preservacao).

ARQUIVO NACIONAL – **Dicionário brasileiro de terminologia arquivística** [Em linha]. Rio de Janeiro: Arquivo Nacional, 2005. [Consult. 26 mar. 2015]. Disponível em [www: <http://www.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf>](http://www.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf). ISBN 85-7009-075-7.

ASSOCIAÇÃO PARA A PROMOÇÃO E DESENVOLVIMENTO DA SOCIEDADE DE INFORMAÇÃO – **A gestão documental na governança da informação** [Em linha]. Lisboa: Associação para a Promoção e Desenvolvimento da Sociedade de Informação, 2014. [Consult. 28 jan. 2015]. Disponível em [www: <http://www.apdsi.pt/uploads/news/id844/Gest%C3%A3o%20Documental%202014_20141111.pdf>](http://www.apdsi.pt/uploads/news/id844/Gest%C3%A3o%20Documental%202014_20141111.pdf).

BAPTISTA, Ana Alice; MACHADO, Altamiro Barbosa – A utilização do Dublin Core qualificado na descrição semântica de uma revista científica em linha. In **Conferência da Associação Portuguesa de Sistemas de Informação**, 1, Guimarães [Em linha]. APSI, 2000, 1-25. [Consult. 20 out. 2014]. Disponível em [www: <http://repositorium.sdum.uminho.pt/bitstream/1822/381/1/APSI_analice.PDF>](http://repositorium.sdum.uminho.pt/bitstream/1822/381/1/APSI_analice.PDF).

BARBEDO, Francisco – Arquivos Digitais: da origem à maturidade. **Cadernos BAD** [Em linha]. 2 (2005), 6-18. [Consult. 11 mar. 2014]. Disponível em [www: <http://eprints.rclis.org/10202/1/CBAD205Barbedo.pdf>](http://eprints.rclis.org/10202/1/CBAD205Barbedo.pdf). ISSN 1645-2895.

BARBEDO, Francisco [et al.] – RODA: Repositório de Objectos Digitais Autênticos. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 9, Ponta Delgada. Bibliotecas e Arquivos – informação para a cidadania, o desenvolvimento e a inovação [Em linha]. Lisboa: BAD, 2007, 1-9. [Consult. 12 mar. 2014]. Disponível em [www: <http://bad.pt/publicacoes/index.php/congressosbad/article/download/535/320>](http://bad.pt/publicacoes/index.php/congressosbad/article/download/535/320).

BEAGRIE, Neil; JONES, Maggie – **Preservation Management of Digital Materials: The Handbook** [Em linha]. Digital Preservation Coalition, 2008. [Consult. 18 nov. 2014]. Disponível em [www: <http://www.dpconline.org/component/docman/doc_download/299-digital-preservation-handbook-digital-perservation-handbook>](http://www.dpconline.org/component/docman/doc_download/299-digital-preservation-handbook-digital-perservation-handbook).

BOERES, Sonia Araújo de Assis; CUNHA, Murilo Bastos da – Competências básicas para os gestores de preservação digital. **Ciência da Informação** [Em linha]. 41:1 (2012), 103-113. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2115/1796>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2115/1796). ISSN 1518-8353.

BORGES, Maria Alice Guimarães – A compreensão da sociedade da informação. **Ciência da Informação** [Em linha]. 29:3 (2000), 25-32. [Consult. 05 mai. 2014]. Disponível em [www: <http://www.scielo.br/pdf/ci/v29n3/a03v29n3.pdf>](http://www.scielo.br/pdf/ci/v29n3/a03v29n3.pdf). ISSN 1518-8353.

BORGHOFF, Uwe M. [*et al.*] – **Long-Term Preservation of Digital Documents: Principles and Practices**. Heidelberg: Springer, 2010. ISBN 978-3-642-07017-4.

BROWN, Adrian – **PRONOM 3: User Requirements** [Em linha]. The National Archives, 2003. [Consult. 03 nov. 2014]. Disponível em [www: <http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_user_requirements.pdf>](http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_user_requirements.pdf).

BROWN, Adrian – **Selecting File Formats for Long-Term Preservation** [Em linha]. The National Archives, 2003. [Consult. 27 nov. 2014]. Disponível em [www: <https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf>](https://www.nationalarchives.gov.uk/documents/selecting-file-formats.pdf).

BUREAU VAN DIJK – **SABI – Sistema de Análises de Balanços Ibéricos**. Update 173, Versão 61.00. Bureau van Dijk Electronic Publishing, atual. 2014, 06 jan. 2014. [Consult. 09 mai. 2014].

CALAZANS, Angélica Toffano Seidel – Conceitos e uso da informação organizacional e informação estratégica. **Transinformação** [Em linha]. 18:1 (2006), 63-70. [Consult.

06 jan. 2015]. Disponível em www: <<http://www.scielo.br/pdf/tinf/v18n1/06.pdf>>. ISSN 0103-3786.

CAMPOS, Fernanda Maria – Informação digital: um novo património a preservar. **Cadernos BAD** [Em linha]. 2 (2002), 8-14. [Consult. 11 mar. 2014]. Disponível em www: <<http://www.bad.pt/a/CadernosBAD/Caderno22002/Campos.pdf>>. ISSN 1645-2895.

CAMPOS, Fernanda; SARAMAGO, Maria de Lurdes – Preservação digital de longo prazo em instituições patrimoniais – Reutilização e adaptação de metadados. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 9, Ponta Delgada. Bibliotecas e Arquivos – informação para a cidadania, o desenvolvimento e a inovação [Em linha]. Lisboa: BAD, 2007, 1-7. [Consult. 12 mar. 2014]. Disponível em www: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/540>>.

CAPURRO, Rafael; HJORLAND, Birger – O conceito de informação. **Perspectivas em Ciência da Informação** [Em linha]. 12:31 (2007), 148-207. [Consult. 24 mar. 2014]. Disponível em www: <<http://www.scielo.br/pdf/pci/v12n1/11.pdf>>. ISSN 1413-9936.

CHOO, Chun Wei – **A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões**. São Paulo: Editora Senac, 2003. ISBN 85-7359-341-5.

COMISSÃO EUROPEIA – Recomendação nº 2003/361/CE. **Jornal Oficial da União Europeia** [Em linha]. L124 (2003). 36-41. [Consult. 07 jan. 2015]. Disponível em www: <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:124:0036:0041:PT:PDF>>.

COMISSÃO EUROPEIA – **A nova definição de PME: Guia do utilizador e modelo de declaração** [Em linha]. Comissão Europeia, 2006. [Consult. 07 jan. 2015]. Disponível em www: <<http://bookshop.europa.eu/pt/a-nova-defini-o-de-pme-pbNB6004773/downloads/NB-60-04-773-PT->

[C/NB6004773PTC_002.pdf?FileName=NB6004773PTC_002.pdf&SKU=NB6004773PTC_PDF&CatalogueNumber=NB-60-04-773-PT-C](http://www.ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index_pt.htm)>. ISBN 92-894-7913-2.

COMISSÃO EUROPEIA – **O que é uma PME? – Pequenas e Médias Empresas (PME) – Empresas e Indústria** [Em linha]. Comissão Europeia, 2014. [Consult. 07 jan. 2015]. Disponível em www: <http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index_pt.htm>.

CONCEIÇÃO, Alexandre Da Silva – Informação arquivística: o [in]sumo da sociedade contemporânea - a riqueza das organizações. **Archeion Online** [Em linha]. 1:1 (2013), 63-76. [Consult. 27 dez. 2014]. Disponível em www: <<http://periodicos.ufpb.br/ojs/index.php/archeion/article/view/17140/9755>>. ISSN 2318-6186.

CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS – **Reference Model for an Open Archival Information System (OAIS): Recommended Practice, Issue 2** [Em linha]. Washington, D.C.: CCSDS, 2012. [Consult. 23 out. 2014]. Disponível em www: <<http://public.ccsds.org/publications/archive/650x0m2.pdf>>.

CÔRBO, Priscila de Assunção Barreto; CARDOSO, Tatyana Marques de Macedo – Preservação da memória digital no Colégio Pedro II. **Ciência da Informação** [Em linha]. 41:1 (2012), 160-174. [Consult. 19 mai. 2014]. Disponível em www: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2112/1803>>. ISSN 1518-8353.

COSTA, Rejane Moreira da; SANCHES, Rosely – Ferramentas de Engenharia Reversa no Apoio à Qualidade de Software. **Instituto de Ciências Matemáticas e de Computação** [Em linha]. 45 (1996), 1-35. [Consult. 15 dez. 2014]. Disponível em www: <http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_4_5.pdf>. ISSN 0103-2569.

DAPPERT, Angela; ENDERS, Markus – Digital Preservation Metadata Standards. **Information Standards Quarterly** [Em linha]. 22:2 (2010), 4-13. [Consult. 20 out. 2014]. Disponível em www:

<http://www.niso.org/apps/group_public/download.php/4236/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf>. ISSN 1041-0031.

DECRETO-LEI nº 16/93. D.R. I Série-A. 19 (23-01-1993) 264-270.

DECRETO-LEI nº 372/2007. D.R. I Série. 213 (06-11-2007) 8080-8084.

DECRETO-LEI nº 381/2007. D.R. I Série. 219 (14-11-2007) 8440-8464.

DECRETO-LEI nº 143/2009. D.R. I Série. 114 (16-06-2009) 3688-3689.

DEEGAN, Marilyn; TANNER, Simon – Key issues in digital preservation. In DEEGAN, Marilyn; TANNER, Simon. **Digital preservation**. London: Facet Publishing, 2006. ISBN 978-1-85604-485-1. Cap. 1, p. 1-31.

DENIS, Elizabeth Gasiorowski – **ISO and IEC approve OpenDocument OASIS standard for data interoperability of office applications** [Em linha]. ISO, 2006. [Consult. 23 set. 2014]. Disponível em [www: <http://www.iso.org/iso/news.htm?refid=Ref1004>](http://www.iso.org/iso/news.htm?refid=Ref1004).

DIGITAL PRESERVATION POLICY WORKING GROUP – **Establishing a Central Depository for Preserving Digital Image Collections** [Em linha]. Cornell University Library, 2001. [Consult. 18 nov. 2014]. Disponível em [www: <https://www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf>](https://www.library.cornell.edu/preservation/IMLS/image_deposit_guidelines.pdf).

DIREÇÃO-GERAL DE ARQUIVOS – **Recomendações para a produção de Planos de Preservação digital** [Em linha]. Lisboa: DGARQ, 2011. [Consult. 10 mar. 2014]. Disponível em [www: <http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/Recomend_producao_PPD_V2.1.pdf>](http://arquivos.dglab.gov.pt/wp-content/uploads/sites/16/2014/02/Recomend_producao_PPD_V2.1.pdf).

DLM FORUM – **Guidelines on best practices for using electronic information: How to deal with machine-readable data and electronic documents** [Em linha]. Luxembourg: European Communities, 1997. [Consult. 10 set. 2014]. Disponível em [www: <http://dlmforum.typepad.com/gdlines.pdf>](http://dlmforum.typepad.com/gdlines.pdf). ISBN 92-828-2285-0.

DLM FORUM – **MoReq 2010: Modular Requirements for Records Systems – Volume 1: Core Services & Plug-in Modules** [Em linha]. DLM Forum, 2011. [Consult. 29 out. 2014]. Disponível em [www: <http://moreq2010.eu/pdf/moreq2010_vol1_v1_1_en.pdf>](http://moreq2010.eu/pdf/moreq2010_vol1_v1_1_en.pdf).

DSPACE – **DSpace Diagram** [Em linha]. DuraSpace. [Consult. 18 dez. 2014]. Disponível em [www: <http://www.dspace.org/sites/dspace.org/files/media/DSpace%20Diagram_0.pdf>](http://www.dspace.org/sites/dspace.org/files/media/DSpace%20Diagram_0.pdf).

DURASPACE – **DuraSpace | Open technologies for durable digital content** [Em linha]. DuraSpace. [Consult. 21 dez. 2014]. Disponível em [www: <http://www.duraspace.org/faq>](http://www.duraspace.org/faq).

DURASPACE – **Tutorial 1 – Introduction to Fedora – Fedora Create – DuraSpace Wiki** [Em linha]. DuraSpace, 2011. [Consult. 21 dez. 2014]. Disponível em [www: <https://wiki.duraspace.org/display/FEDORACREATE/Tutorial+1+-+Introduction+to+Fedora>](https://wiki.duraspace.org/display/FEDORACREATE/Tutorial+1+-+Introduction+to+Fedora).

DURASPACE – **EndUserFaq** [Em linha]. DuraSpace, 2014^a. [Consult. 18 dez. 2014]. Disponível em [www: <https://wiki.duraspace.org/display/DSPACE/User+FAQ>](https://wiki.duraspace.org/display/DSPACE/User+FAQ).

DURASPACE – **Fedora Repository Home – Fedora Repository – DuraSpace Wiki** [Em linha]. DuraSpace, 2014^b. [Consult. 21 dez. 2014]. Disponível em [www: <https://wiki.duraspace.org/display/FF/Fedora+Repository+Home>](https://wiki.duraspace.org/display/FF/Fedora+Repository+Home).

FEDORA – **Fedora Repository Project | Fedora Repository** [Em linha]. DuraSpace. [Consult. 21 dez. 2014]. Disponível em [www: <http://fedorarepository.org/about>](http://fedorarepository.org/about).

FERREIRA, Carla Alexandra Silva – **Preservação da Informação Digital: uma perspectiva orientada para as bibliotecas** [Em linha]. Coimbra: Faculdade de Letras da Universidade de Coimbra, 2011. [Consult. 10 mar. 2014]. Disponível em [www: <https://estudogeral.sib.uc.pt/bitstream/10316/15001/1/Preserva%C3%A7%C3%A3o%20da%20Informa%C3%A7%C3%A3o%20Digital.pdf>](https://estudogeral.sib.uc.pt/bitstream/10316/15001/1/Preserva%C3%A7%C3%A3o%20da%20Informa%C3%A7%C3%A3o%20Digital.pdf). Dissertação de mestrado.

FERREIRA, Miguel – **Introdução à preservação digital – Conceitos, estratégias e actuais consensos** [Em linha]. Guimarães: Escola de Engenharia da Universidade do Minho, 2006. [Consult. 10 mar. 2014]. Disponível em www: <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. ISBN 972-8692-30-7.

FERREIRA, Miguel; BAPTISTA, Ana Alice; RAMALHO, José Carlos – CriB: A service oriented architecture for digital preservation outsourcing. In **XATA – XML: Aplicações e Tecnologias Associadas**, 4, Portalegre. XATA2006 – XML: Aplicações e Tecnologias Associadas [Em linha]. Portalegre: ESTG, 2006, 1-12. [Consult. 12 mar. 2014]. Disponível em [www: <http://xata.fe.up.pt/2006/papers/7.pdf>](http://xata.fe.up.pt/2006/papers/7.pdf). ISBN 972-99166-2-4.

FERREIRA, Miguel; BAPTISTA, Ana Alice; RAMALHO, José Carlos – An intelligent decision support system for digital preservation. **International Journal on Digital Libraries** [Em linha]. 6:4 (2007), 1-10. [Consult. 12 mar. 2014]. Disponível em [www: <http://repositorium.sdum.uminho.pt/bitstream/1822/6648/1/IJDL-0.4.pdf>](http://repositorium.sdum.uminho.pt/bitstream/1822/6648/1/IJDL-0.4.pdf). ISSN 1432-5012.

FERREIRA, Miguel; SARAIVA, Ricardo; RODRIGUES, Eloy – **Estado da arte em preservação digital** [Em linha]. Lisboa: RCAAP, 2012 [Consult. 10 mar. 2014]. Disponível em [www: <http://repositorium.sdum.uminho.pt/bitstream/1822/17049/2/Estado_da_arte_em_preservacao_digital_v1.0.pdf>](http://repositorium.sdum.uminho.pt/bitstream/1822/17049/2/Estado_da_arte_em_preservacao_digital_v1.0.pdf).

FIGUEROLA, Carlos G. [et al.] – Preservación digital. **Ibersid** [Em linha]. 3 (2009), 265-274. [Consult. 17 mar. 2014]. Disponível em [www: <http://ibersid.eu/ojs/index.php/ibersid/article/view/3749/3510>](http://ibersid.eu/ojs/index.php/ibersid/article/view/3749/3510). ISSN 1888-0967.

FOOT, Mirjam – Preservation: policy, practice and co-operation. **Páginas a&b** [Em linha]. 4 (1999), 7-22. [Consult. 12 mar. 2014]. Disponível em [www: <http://revistas.ua.pt/index.php/paginasab/article/view/1171/1085>](http://revistas.ua.pt/index.php/paginasab/article/view/1171/1085). ISSN 0873-5670.

FREIRE, Gustavo Henrique de Araújo; SILVA, Jonathas Luiz Carvalho – A configuração do campo da ciência da informação: marcas de uma identidade. **Informação & Sociedade: Estudos** [Em linha]. 22:especial (2012), 161-174. [Consult. 21 mar. 2014]. Disponível em [www: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/13775/8216>](http://www.ies.ufpb.br/ojs/index.php/ies/article/view/13775/8216). ISSN 0104-0146.

FREITAS, Cristiana Vieira de – Garantir a autenticidade e o acesso continuado à informação digital: os desafios da preservação digital em arquivos. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 11, Lisboa. Integração, Acesso e Valor Social [Em linha]. Lisboa: BAD, 2012, 1-9. [Consult. 11 mar. 2014]. Disponível em [www: <http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/272/pdf>](http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/272/pdf).

GALINDO, Marcos – O Dilemma do Pharmacon. **Ciência da Informação** [Em linha]. 41:1 (2012), 36-50. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2111/1790>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2111/1790). ISSN 1518-8353.

GLADNEY, Henry M. – **Preserving digital information**. Berlin: Springer, 2007. ISBN 978-3-540-37886-0.

GRÁCIO, José Carlos Abbud – **Preservação digital na gestão da informação: um modelo processual para as instituições de ensino superior** [Em linha]. São Paulo: Cultura Acadêmica, 2012. [Consult. 12 mar. 2014]. Disponível em [www: <http://www.culturaacademica.com.br/img/arquivos/Preservacao_digital_na_gestao_da_informacao-WEB_v2.pdf>](http://www.culturaacademica.com.br/img/arquivos/Preservacao_digital_na_gestao_da_informacao-WEB_v2.pdf). ISBN 978-85-7983-333-5.

GRÁCIO, José Carlos Abbud; FADEL, Bárbara – Preservação digital em instituições de ensino superior no Brasil: aspectos, estratégias e políticas. **Ibersid** [Em linha]. 3 (2009), 275-280. [Consult. 16 mar. 2014]. Disponível em [www: <http://ibersid.eu/ojs/index.php/ibersid/article/view/3750/3511>](http://ibersid.eu/ojs/index.php/ibersid/article/view/3750/3511). ISSN 1888-0967.

GRÁCIO, José Carlos Abbud; FADEL, Bárbara; VALENTIM, Marta Lúcia Pomim – Preservação digital nas instituições de ensino superior: aspectos organizacionais, legais e técnicos. **Perspectivas em Ciência da Informação** [Em linha]. 18:3 (2013),

111–129. [Consult. 11 mar. 2014]. Disponível em [www: <http://www.scielo.br/pdf/pci/v18n3/08.pdf>](http://www.scielo.br/pdf/pci/v18n3/08.pdf). ISSN 1413-9936.

HARPER, Corey A. – Dublin Core Metadata Initiative: Beyond the Element Set. **Information Standards Quarterly** [Em linha]. 22:1 (2010), 20–28. [Consult. 20 out. 2014]. Disponível em [www: <http://www.niso.org/publications/isq/free/FE_DCMI_Harper_isqv22no1.pdf>](http://www.niso.org/publications/isq/free/FE_DCMI_Harper_isqv22no1.pdf). ISSN 1041-0031.

INFOPÉDIA – Agentes económicos. In **Língua Portuguesa com Acordo Ortográfico** [Em linha]. Porto: Porto Editora, 2003-2015^a. [Consult. 10 jan. 2015]. Disponível em [www: <http://www.infopedia.pt/\\$agentes-economicos>](http://www.infopedia.pt/$agentes-economicos).

INFOPÉDIA – Instituições financeiras. In **Língua Portuguesa com Acordo Ortográfico** [Em linha]. Porto: Porto Editora, 2003-2015^b. [Consult. 10 jan. 2015]. Disponível em [www: <http://www.infopedia.pt/\\$instituicoes-financeiras>](http://www.infopedia.pt/$instituicoes-financeiras).

INFOPÉDIA – Nuvem. In **Língua Portuguesa com Acordo Ortográfico** [Em linha]. Porto: Porto Editora, 2003-2015^c. [Consult. 27 fev. 2015]. Disponível em [www: <http://www.infopedia.pt/dicionarios/lingua-portuguesa/nuvem>](http://www.infopedia.pt/dicionarios/lingua-portuguesa/nuvem).

INSTITUTO NACIONAL DE ESTATÍSTICA – **Proporção de empresas com 10 e mais pessoas ao serviço que utilizam sistemas operativos de fonte aberta (Série antiga - %); Anual** [Em linha]. Lisboa: INE, 2008. [Consult. 01 out. 2014]. Disponível em [www: <http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0002227&contexto=bd&selTab=tab2>](http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0002227&contexto=bd&selTab=tab2).

INSTITUTO NACIONAL DE ESTATÍSTICA – **Proporção de empresas com 10 e mais pessoas ao serviço que utilizam sistemas operativos de fonte aberta (Série 2009-2010 - %)** [Em linha]. Lisboa: INE, 2010. [Consult. 01 out. 2014]. Disponível em [www: <http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0002991&contexto=bd&selTab=tab2>](http://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0002991&contexto=bd&selTab=tab2).

INSTITUTO NACIONAL DE ESTATÍSTICA – **A crise e as grandes empresas** [Em linha]. Lisboa: INE, 2014^a. [Consult. 08 jan. 2015]. Disponível em [www: <http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=219181079&att_displasy=n&att_download=y>](http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=219181079&att_displasy=n&att_download=y).

INSTITUTO NACIONAL DE ESTATÍSTICA – **Empresas em Portugal 2012** [Em linha]. Lisboa: INE, 2014^b. [Consult. 08 jan. 2015]. Disponível em [www: <http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=215340824&att_displasy=n&att_download=y>](http://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=215340824&att_displasy=n&att_download=y). ISSN: 0872-9514.

JUNIOR, Laerte Pereira da Silva; MOTA, Valeira Gameleira da – Políticas de preservação digital no Brasil: características e implementações. **Ciência da Informação** [Em linha]. 41:1 (2012), 51-64. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2123/1806>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2123/1806). ISSN 1518-8353.

KASTELLEK, Mike – Practical Limits to the Scope of Digital Preservation. **Information Technology and Libraries** [Em linha]. 31:2 (2012), 63-71. [Consult. 17 mar. 2014]. Disponível em [www: <https://napoleon.bc.edu/ojs/index.php/ital/article/view/2167/pdf>](https://napoleon.bc.edu/ojs/index.php/ital/article/view/2167/pdf). ISSN 2163-5226.

KEEFER, Alice; GALLART, Núria – **La preservación de recursos digitales: el reto para las bibliotecas del siglo XXI**. Barcelona: Editorial UOC, 2007. ISBN 978-84-9788-567-6.

LEE, Kyong-Ho [*et al.*] – The State of the Art and Practice in Digital Preservation. **Journal of Research of National Institute of Standards and Technology** [Em linha]. 107:1 (2002), 93-106. [Consult. 14 nov. 2014]. Disponível em [www: <https://archive.org/download/jresv107n1p93/jresv107n1p93_A1b.pdf>](https://archive.org/download/jresv107n1p93/jresv107n1p93_A1b.pdf). ISSN 1044-677X.

LESSARD-HÉBERT, Michelle; GOYETTE, Gabriel; BOUTIN, Gérald – **Investigação Qualitativa: Fundamentos e Práticas**. 2.^a ed. Lisboa: Instituto Piaget, 2005. ISBN 978-771-737-3.

LEXTEC – **Lextec – volume de negócios** [Em linha]. Instituto Camões. [Consult. 08 jan. 2015]. Disponível em [www: <http://www.instituto-camoes.pt/lextec/por/domain_3/definition/17881.html>](http://www.instituto-camoes.pt/lextec/por/domain_3/definition/17881.html).

LIBRARY OF CONGRESS – **PDF/A, PDF for Long-term Preservation** [Em linha]. Library of Congress, 2011^a. [Consult. 23 set. 2014]. Disponível em [www: <http://www.digitalpreservation.gov/formats/fdd/fdd000318.shtml>](http://www.digitalpreservation.gov/formats/fdd/fdd000318.shtml).

LIBRARY OF CONGRESS – **METS Example Documents: Metadata Encoding and Transmission Standard (METS)** [Em linha]. Library of Congress, 2011^b. [Consult. 17 dez. 2014]. Disponível em [www: <http://www.loc.gov/standards/mets/mets-examples.html>](http://www.loc.gov/standards/mets/mets-examples.html).

LIBRARY OF CONGRESS – **Schemas for PREMIS. PREMIS: Preservation Metadata Maintenance Activity** [Em linha]. Library of Congress, 2012. [Consult. 17 dez. 2014]. Disponível em [www: <http://www.loc.gov/standards/premis/schemas.html>](http://www.loc.gov/standards/premis/schemas.html).

LOCKSS – **Formal statement of Conformance to ISO 14721:2003** [Em linha]. Stanford: LOCKSS, 2004 [Consult. 28 out. 2014]. Disponível em [www: <http://www.lockss.org/locksswp/wp-content/uploads/2011/11/OAIS-LOCKSS-Conformance.pdf>](http://www.lockss.org/locksswp/wp-content/uploads/2011/11/OAIS-LOCKSS-Conformance.pdf).

LOUSADA, Mariana – **Informação orgânica como insumo estratégico para a tomada de decisão em ambientes competitivos** [Em linha]. São Paulo: Faculdade de Filosofia e Ciências da Universidade Estadual Paulista, 2011. [Consult. 27 dez. 2014]. Disponível em [www: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/lousada_m_me_mar.pdf>](http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/lousada_m_me_mar.pdf).

Dissertação de mestrado.

LOUSADA, Mariana; VALENTIM, Marta Lígia Pomim – A relação entre a informação orgânica e a gestão documental. In VALENTIM, Marta Lígia Pomim. **Gestão, mediação e uso da informação** [Em linha]. São Paulo: Editora UNESP, 2010. [Consult. 10 mar. 2014]. Disponível em [www: <http://www.unesp.br>](http://www.unesp.br).

<http://www.culturaacademica.com.br/img/arquivos/Gestao_mediacao_e_uso_da_informacao.pdf>. ISBN 978-85-7893-117-1. Cap. 17, p. 361-384.

MÁRDERO ARELLANO, Miguel Ángel – Preservação de documentos digitais. **Ciência da Informação** [Em linha]. 33:2 (2004), 15-27. [Consult. 11 mar. 2014]. Disponível em www: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>>. ISSN 1518-8353.

MÁRDERO ARELLANO, Miguel Ángel – **Critérios para a preservação digital da informação científica** [Em linha]. Brasília: Universidade de Brasília, 2008. [Consult. 11 mar. 2014]. Disponível em www: <http://repositorio.unb.br/bitstream/10482/1518/1/2008_MiguelAngelMarderoArellano.pdf>. Tese de doutoramento.

MÁRDERO ARELLANO, Miguel Ángel – Cariniana: uma rede nacional de preservação digital. **Ciência da Informação** [Em linha]. 41:1 (2012), 74-82. [Consult. 19 mai. 2014]. Disponível em www: <http://cariniana.ibict.br/index.php/publicacoes/item/download/38_5e1dc5644a2a0f3c26f38a672d3f385a>. ISSN 1518-8353.

MÁRDERO ARELLANO, Miguel Ángel; LEITE, Fernando César Lima – Acesso aberto à informação científica e o problema da preservação digital. **Biblios** [Em linha]. 35 (2009), 1-10. [Consult. 11 mar. 2014]. Disponível em www: <http://repositorio.unb.br/bitstream/10482/4937/6/2009_Acesso%20aberto%20%C3%A0%20informa%C3%A7%C3%A3o.pdf>. ISSN 1562-4730.

MORAES, Cássia Regina Bassan de; FADEL, Bárbara – A informação no contexto organizacional: tipos, características e usos. **Ibersid** [Em linha]. 3 (2009), 61-65. [Consult. 27 dez. 2014]. Disponível em www: <<http://www.iversid.eu/ojs/index.php/iversid/article/download/3724/3485>>. ISSN 1888-0967.

NATIONAL INFORMATION STANDARDS ORGANIZATION – **Understanding Metadata** [Em linha]. Bethesda: NISO Press, 2004. [Consult. 20 out. 2014]. Disponível

em www: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. ISBN 1-880124-62-9.

NATIONAL INFORMATION STANDARDS ORGANIZATION – **The Dublin Core Metadata Element Set** [Em linha]. Baltimore: NISO, 2013. [Consult. 20 out. 2014]. Disponível em www: <http://www.niso.org/apps/group_public/download.php/10256/Z39-85-2012_dublin_core.pdf>. ISSN 1041-5635.

NATIONAL LIBRARY OF AUSTRALIA – **Guidelines for the Preservation of Digital Heritage** [Em linha]. UNESCO, 2003. [Consult. 10 set. 2014]. Disponível em www: <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>>.

OCLC/RLG WORKING GROUP ON PRESERVATION METADATA – **Preservation Metadata for Digital Objects: A Review of the State of the Art** [Em linha]. OCLC/RLG, 2001. [Consult. 14 out. 2014]. Disponível em www: <http://www.oclc.org/content/dam/research/activities/pmwg/presmeta_wp.pdf?urlm=161396>.

ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS – **Introduction to UDDI: Important Features and Functional Concepts** [Em linha]. OASIS, 2004. [Consult. 03 nov. 2014]. Disponível em www: <<http://www.uddi.org/pubs/uddi-tech-wp.pdf>>.

PACONTAS – **O que é um Balanço – Contabilidade** [Em linha]. Lousada: PACONTAS, 2012. [Consult. 08 jan. 2015]. Disponível em www: <<http://www.pacontas.pt/blog/201212/porque-e-importante-o-balaco-no-final-do-ano.aspx>>.

PINA, Sofia – A preservação digital na Era da Comunicação. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 10, Guimarães. Políticas de Informação na Sociedade em Rede [Em linha]. Lisboa: BAD, 2010, 1-5. [Consult. 11 mar. 2014]. Disponível em www: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/175/171>>.

PINHO, Micael Ferreira Alves de – **Modelo de Replicação para a Preservação de Dados em Repositórios** [Em linha]. Porto: Faculdade de Engenharia da Universidade do Porto, 2012. [Consult. 11 mar. 2014]. Disponível em www: <<http://repositorio-aberto.up.pt/bitstream/10216/68432/1/000155172.pdf>>. Dissertação de Mestrado.

PINTO, Carlos A. Marques [et al.] – **Fundamentos de Gestão**. 3.^a ed. Lisboa: Editorial Presença, 2010. ISBN 978-972-23-3654-3.

PINTO, Maria Manuela – Do «efémero» ao «sistema de informação»: a preservação na era digital. **Páginas a&b** [Em linha]. 15 (2005), 53-106. [Consult. 12 mar. 2014]. Disponível em www: <<http://ler.letras.up.pt/uploads/ficheiros/3083.pdf>>. ISSN 0873-5670.

PINTO, Maria Manuela – Gestão da Informação e preservação digital: uma perspectiva portuguesa de uma mudança de paradigma. In **Congress Isko-Spain**, 9, València. New Perspectives for the organisation and dissemination of knowledge [Em linha]. València: Universitat Politècnica de València, 2009^a, 323-355. [Consult 11 mar. 2014]. Disponível em www: <http://repositorio-aberto.up.pt/bitstream/10216/25380/2/manuelapintogestao000100395.pdf>.

PINTO, Maria Manuela – **PRESERVMAP: Um roteiro da preservação na era digital**. Porto: Edições Afrontamento, 2009^b. ISBN 978-972-36-1070-3.

PORTARIA nº 412/2001. D.R. I Série-B. 90 (17-04-2001) 2243-2260.

POWELL, Andy; JOHNSTON, Pete – **Guidelines for implementing Dublin Core in XML** [Em linha]. Dublin Core Metadata Initiative, 2002. [Consult. 17 dez. 2014]. Disponível em www: <<http://dublincore.org/documents/2002/12/02/dc-xml-guidelines/>>.

PREMIS Editorial Committee – **PREMIS Data Dictionary for Preservation Metadata** [Em linha]. Library of Congress, 2012. [Consult. 16 abr. 2015]. Disponível em www: <<http://www.loc.gov/standards/premis/v2/premis-2-2.pdf>>.

QUIVY, Raymond; CAMPENHOUDT, Luc Van – **Manual de investigação em ciências sociais**. 6.^a ed. Lisboa: Gradiva, 2013. ISBN 978-972-662-275-8.

RABELLO, Rodrigo; CASTRO, Virgínia Ferreira da Silva – Intermediação da informação e preservação da memória digital. **Ciência da Informação** [Em linha]. 41:1 (2012), 22-35. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2113/1789>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2113/1789). ISSN 1518-8353.

RAMALHO, José Carlos [et al.] – RODA and CRiB: a Service-Oriented Digital Repository. In **International Conference on Preservation of Digital Objects**, 5, London. Joined Up and Working: Tools and Methods for Digital Preservation [Em linha]. London: iPRES, 2008, 1-7. [Consult. 12 mar. 2014]. Disponível em [www: <http://repositorium.sdum.uminho.pt/bitstream/1822/8226/1/RodaAndCrib.pdf>](http://repositorium.sdum.uminho.pt/bitstream/1822/8226/1/RodaAndCrib.pdf).

REICH, Victoria Ann – LOCKSS: ensuring access through time. **Ciência da Informação** [Em linha]. 41:1 (2012), 74-82. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2125>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2125). ISSN 1518-8353.

REPOSITÓRIO CIENTÍFICO DE ACESSO ABERTO DE PORTUGAL – **O Sistema DSPACE** [Em linha]. Lisboa: RCAAP. [Consult. 18 dez. 2014]. Disponível em [www: <http://projeto.rcaap.pt/formar/mod1/contents/o_sistema_dspace.html>](http://projeto.rcaap.pt/formar/mod1/contents/o_sistema_dspace.html).

RIBEIRO, Fernanda – Informação: um campo uno, profissões diversas?. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 8, Estoril. Nas encruzilhadas da Informação e da Cultura: (re)inventar a profissão [Em linha]. Lisboa: BAD, 2004, 1-8. [Consult. 05 mai. 2014]. Disponível em [www: <http://ler.letras.up.pt/uploads/ficheiros/artigo4181.pdf>](http://ler.letras.up.pt/uploads/ficheiros/artigo4181.pdf).

RIBEIRO, Fernanda – A arquivística como disciplina aplicada no campo da ciência da informação. **Perspectivas em Gestão & Conhecimento** [Em linha]. 1:1 (2011), 59-73. [Consult. 24 mar. 2014]. Disponível em [www: <http://periodicos.ufpb.br/ojs/index.php/pgc/article/view/9887/5619>](http://periodicos.ufpb.br/ojs/index.php/pgc/article/view/9887/5619). ISSN 2236-417X.

ROBINSON, Christine – Documentation Dilemmas. **Journal for Quality & Participation** [Em linha]. 31:4 (2009), 35-37. [Consult. 06 jan. 2015]. Disponível em www:

<<http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=36560656&lang=pt-br&site=ehost-live>>. ISSN 1040-9602.

ROMERO, Jorge Candás – El papel de los metadatos en la preservación digital. **El profesional de la información** [Em linha]. 15:2 (2006), 126-136. [Consult. 17 mar. 2014]. Disponível em www: <<http://eprints.rclis.org/8359/1/final.pdf>>. ISSN 1386-6710.

ROUSSEAU, Jean-Yves; COUTURE, Carol – **Os fundamentos da disciplina arquivística**. Lisboa: Dom Quixote, 1998. ISBN 972-20-1428-5.

SANT'ANNA, Marcelo Leone – Os Desafios da Preservação de Documentos Públicos Digitais. **Revista IP** [Em linha]. 3:2 (2001), 123-135. [Consult. 11 mar. 2014]. Disponível em www: <<http://egov.ufsc.br/portal/sites/default/files/anexos/27269-27279-1-PB.pdf>>. ISSN 516-697X.

SARAMAGO, Maria de Lurdes – Preservação digital a longo prazo: boas práticas e estratégias. **Cadernos BAD** [Em linha]. 2 (2002), 54-68. [Consult. 11 mar. 2014]. Disponível em www: <<http://www.bad.pt/a/CadernosBAD/Caderno22002/Saramago.pdf>>. ISSN 1645-2895.

SARAMAGO, Maria de Lurdes – Metadados para preservação digital e aplicação do modelo OAIS. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 8, Estoril. Nas encruzilhadas da Informação e da Cultura: (re)inventar a profissão [Em linha]. Lisboa: BAD, 2004, 1-6. [Consult. 11 mar. 2014]. Disponível em www: <<http://bad.pt/publicacoes/index.php/congressosbad/article/view/640/637>>.

SAYÃO, Luís Fernando – Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação** [Em linha]. 15:30 (2010), 1-31. [Consult. 15 out. 2014].

Disponível em www: <<http://www.brapci.ufpr.br/download.php?dd0=12621>>. ISSN 1518-2924.

SERRANO, António; FIALHO, Cândido – **Gestão do Conhecimento – O novo paradigma das organizações**. 2.^a ed. Lisboa: FCA – Editora de Informática, 2005. ISBN 978-972-722-484-5.

SILVA, Andreia Cunha da; ANTÓNIO, Rafael – A Gestão Documental na perspectiva do MoReq2010. In **Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, 11, Lisboa. Integração, Acesso e Valor Social [Em linha]. Lisboa: BAD, 2012, 1-7. [Consult. 05 mai. 2014]. Disponível em www: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/291/pdf>>.

SILVA, Armando Malheiro da – **A informação: da compreensão do fenómeno e construção do objecto científico**. Porto: Edições Afrontamento, 2006. ISBN 972-36-0859-6.

SILVA, Armando Malheiro da – O Método Quadripolar e a Pesquisa em Ciência da Informação. **Prisma.com** [Em linha]. 26 (2014), 27-44. [Consult. 05 mai. 2015]. Disponível em www: <http://revistas.ua.pt/index.php/prisma.com/article/viewFile/3097/pdf_37>. ISSN 1646-3153.

SILVA, Armando Malheiro da; RIBEIRO, Fernanda – Perspectivar a avaliação como operação metodológica no âmbito da ciência da informação. In **Congress Isko-Spain**, 9, Valência. New Perspectives for the organisation and dissemination of knowledge [Em linha]. Valência: Universitat Politècnica de València, 2009, 288-307. [Consult. 06 mai. 2014]. Disponível em www: <<http://repositorio-aberto.up.pt/bitstream/10216/30085/2/amalheiroperspectivar000109276.pdf>>.

SOUZA, Arthur Heleno Lima Rodrigues de [et al.] – O modelo de referência OAIS e a preservação digital distribuída. **Ciência da Informação** [Em linha]. 41:1 (2012), 65-73. [Consult. 19 mai. 2014]. Disponível em www: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2118/1770>>. ISSN 1518-8353.

TAVARAYAMA, Rodrigo; SILVA, Regina Célia Marques Freitas; MARTINS, José Roberto – A sociedade da informação: possibilidades e desafios. **Nucleus** [Em linha]. 9:1 (2012), 253-262. [Consult. 19 ago. 2014]. Disponível em [www: <http://dialnet.unirioja.es/descarga/articulo/3988649.pdf>](http://dialnet.unirioja.es/descarga/articulo/3988649.pdf). ISSN 1982-2278.

TAVARES, Maria de Fátima Duarte – Preservação digital: entre a memória e a história. **Ciência da Informação** [Em linha]. 41:1 (2012), 9-21. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2119/1807>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2119/1807). ISSN 1518-8353.

TEIXEIRA, Sebastião – **Gestão Estratégica**. Lisboa: Escolar Editora, 2011. ISBN 978-972-592-300-9.

TÉRMENS, Miquel – Las auditorías, una metodología para planificar la preservación digital. Experiencias en España. **Ciência da Informação** [Em linha]. 41:1 (2012), 140-142. [Consult. 19 mai. 2014]. Disponível em [www: <http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2143/1799>](http://revista.ibict.br/ciinf/index.php/ciinf/article/view/2143/1799). ISSN 1518-8353.

TERRA, Ana Lúcia – A metodologia quadripolar de investigação científica aplicada em Ciência da Informação: relato de experiência. **Prisma.com** [Em linha]. 26 (2014), 45-66. [Consult. 05 mai. 2015]. Disponível em [www: <http://revistas.ua.pt/index.php/prisma.com/article/viewFile/3098/pdf_41>](http://revistas.ua.pt/index.php/prisma.com/article/viewFile/3098/pdf_41). ISSN 1646-3153.

THIBODEAU, Kenneth – Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In **The State of Digital Preservation: An International Perspective** [Em linha]. Washington D.C.: Council on Library and Information Resources, 2002, 4-31. [Consult. 01 out. 2014]. Disponível em [www: <http://www.clir.org/pubs/reports/pub107/pub107.pdf>](http://www.clir.org/pubs/reports/pub107/pub107.pdf). ISBN 1-887334-92-0.

THOMAZ, Katia P.; SOARES, Antonio José – A preservação digital e o modelo de referência Open Archival Information System (OAIS). **DataGramZero – Revista de Ciência da Informação** [Em linha]. 5:1 (2004), 1-17. [Consult. 10 mar. 2014].

Disponível em www: <http://www.brapci.ufpr.br/download.php?dd0=7616>>. ISSN 1517-3801.

VIEIRA, Ricardo; BORBINHA, José – MoReq2010 – Uma Apresentação. In **Encontro Nacional de Arquivos Municipais**, 10, Leiria. Gestão da Informação na Administração Municipal: passado, presente e futuro [Em linha]. Lisboa: BAD, 2011, 1-8. [Consult. 29 out. 2014]. Disponível em www: <<http://bad.pt/publicacoes/index.php/arquivosmunicipais/article/download/19/9>>.

W3C WORKING GROUP – **Web Services Architecture** [Em linha]. World Wide Web Consortium, 2004. [Consult. 01 nov. 2014]. Disponível em www: <<http://www.w3.org/TR/ws-arch/#whatis>>.

WENDLER, Robin – The status of preservation metadata in the digital library community. In DEEGAN, Marilyn; TANNER, Simon. **Digital preservation**. London: Facet Publishing, 2006. ISBN 978-1-85604-485-1. Cap. 3, p. 60-77.

YALE UNIVERSITY LIBRARY – **Preservation Metadata Task Force** [Em linha]. Yale University Library, 2006. [Consult. 17 dez. 2014]. Disponível em www: <<http://www.library.yale.edu/cataloging/metadata/pmtf/>>.

Anexos

Anexo A – Boas práticas para o correto armazenamento e manutenção de objetos digitais

O *Digital Preservation Policy Working Group* (2001, p. 16-17) e Beagrie e Jones (2008, p. 104-110) descrevem um conjunto de boas práticas que devem ser consideradas na preparação de um plano de preservação digital, no sentido de garantir uma gestão eficaz dos objetos digitais armazenados e a sua sustentabilidade a longo prazo.

Suportes e Formatos de armazenamento

Em relação aos suportes físicos (*hardware*) usados para o armazenamento dos objetos digitais, e os formatos lógicos (*software*) em que esses mesmos objetos digitais são armazenados, sugere-se que:

- Os ficheiros principais sejam armazenados em suportes digitais contemporâneos de alta qualidade e fabricados de acordo com normas. O uso de suportes estáveis e duradouros reduz a necessidade de estratégias de frescamento e, conseqüentemente, a probabilidade de perda de dados resultantes desse processo.
- O mesmo objeto digital seja guardado com recurso a diferentes formatos de ficheiros e aplicações de *software* (quando possível) de modo a proteger a informação nele existente da obsolescência de algum desses formatos, e de possível corrupção causada por falhas no *software*.
- Após a compra, o suporte seja controlado de modo a evitar que falhas no processo de fabrico venham a prejudicar o armazenamento da informação.
- A legibilidade dos suportes digitais e a integridade dos objetos neles armazenados sejam verificados periodicamente.

Cópias de segurança (*backups*)

O armazenamento dos documentos não deve ser realizado num único dispositivo físico, já que muitos são facilmente danificados com o uso, como é o caso

dos dispositivos óticos, sujeitos a riscos e pós. Por essa razão, sugere-se que se realizem cópias de segurança, tendo em conta que:

- Sejam efetuadas cópias de segurança dos ficheiros principais e armazenadas em locais externos seguros e distantes, de modo a garantir que em caso de desastre (por exemplo, sismo, ou incêndio) as cópias não sejam também afetadas.
- As formas de *backup* mais aconselhadas são as *tapes*, discos rígidos e cópias de segurança remotas, já que são menos suscetíveis a danos (ter em consideração o armazenamento dos dispositivos baseados em gravação magnética longe de fontes magnéticas fortes). Os menos aconselhados são os discos óticos, pelas razões já mencionadas, as disquetes, pela sua fragilidade magnética, e as memórias *flash* (embora a tecnologia destas últimas tenha avançado muito recentemente, o que poderá, no futuro, fazer com que também sejam formas seguras de armazenamento).
- Os dispositivos sejam devidamente monitorizados, assegurando que se encontram de boa qualidade e que estão a ser administrados da forma correta.

Armazenamento

Se por um lado é importante escolher devidamente o formato lógico e o dispositivo físico para o armazenamento dos objetos digitais, também é importante escolher o local onde os dispositivos físicos são guardados. Assim, deve-se:

- Armazenar os suportes em ambiente controlado, no que toca aos níveis de temperatura (entre 18°C e 22°C) e humidade (entre 35% e 45%) (valores indicados para *tapes*). Procurar manter valores consistentes dentro do intervalo aceitável.
- Armazenar os suportes afastados da luz solar.
- Manter distância de zonas com campos magnéticos intensos.
- Manter o ambiente limpo.
- Minimizar o manuseamento e a utilização dos dispositivos de armazenamento magnéticos de modo a reduzir o desgaste.

- Manusear com cuidado a superfície dos suportes óticos para que não fiquem ilegíveis devido a dedadas ou riscos.

Segurança

Os objetos digitais devem ser armazenados de forma segura, não só no sentido de minimizar danos por catástrofes, quer para limitar o acesso indevido, ou mesmo a substituição da informação.

- O acesso às zonas de armazenamento deve ser controlado.
- Os suportes de armazenamento devem ser guardados em zonas seguras.
- Devem ser aplicados mecanismos de segurança necessários para salvaguardar a autenticidade dos objetos digitais e dessa forma garantir que não ocorram modificações acidentais ou propositadas.

Refrescamento

Do ponto de vista de preservação lógica duradoura dos objetos digitais sugere-se que:

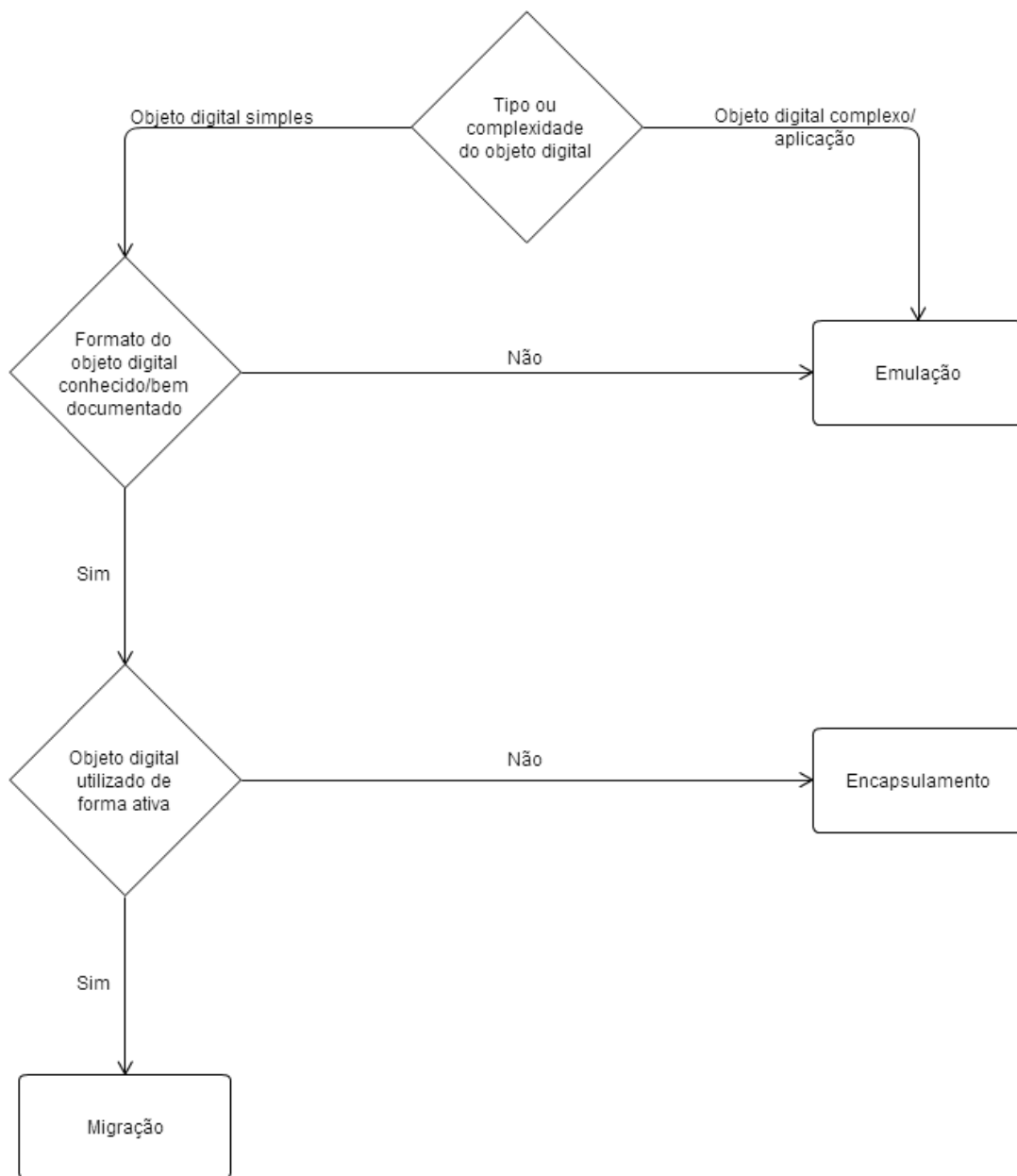
- Os objetos digitais sejam atualizados, periodicamente, sendo usados novos suportes sempre que seja indicado pelas recomendações do fabricante, quando novos dispositivos sejam instalados ou quando uma auditoria aos dados revele a existência de erros. Devem seguir-se as tendências das tecnologias de armazenamento mais seguras e eficientes.
- Sejam implementados procedimentos de verificação para garantir a autenticidade e integridade dos suportes após o processo de refrescamento.

Documentação

De forma a assegurar a acessibilidade dos objetos digitais a longo prazo e de forma autêntica durante o seu ciclo de vida, deve-se:

- Relatar as ações levadas a cabo durante o processo de refrescamento ou qualquer outra operação que possa por em causa a integridade dos objetos digitais.
- Desenvolver um plano de recuperação em caso de desastre, devido a ações naturais ou humanas, e realizar sessões de formação para capacitação dos colaboradores sobre esses procedimentos.

Anexo B – Processo de escolha da estratégia de preservação digital



Fonte: Lee *et al.*, 2002, p. 103

Figura 24 – Esquema para seleção da estratégia de preservação digital

Anexo C – Vantagens e desvantagens de estratégias de preservação digital

Existe um vasto conjunto de estratégias e boas práticas que permitem a manutenção e acesso a longo prazo aos objetos em formato digital.

São também vários os autores que realizam estudos acerca desta temática. De seguida, apresenta-se um resumo das vantagens e desvantagens de algumas das estratégias apresentadas nesta dissertação, com base nos trabalhos de Beagrie e Jones, de Keefer e Gallart e da *National Library of Australia*.

Preservação de tecnologia

A preservação de tecnologia procura manter operacional a configuração tecnológica existente no momento de criação dos objetos digitais, enquanto estes se mostrem necessários.

Vantagens

- Preserva as funcionalidades e aparência do objeto digital original.
- Pode atrasar a necessidade de aplicação de outras estratégias de preservação.
- Pode ser a solução provisória mais prática para objetos digitais complexos.

Desvantagens

- Trata-se de uma estratégia a curto e médio prazo.
- Manter as plataformas digitais antigas vai acabar por se tornar problemático, quer em termos de ocupação de espaço, como na cessação inevitável do suporte técnico.

Emulação

A emulação implica a criação de um *software* que simula a plataforma tecnológica existente no momento de criação dos objetos digitais e que entretanto se tornou obsoleta.

Vantagens

- Reproduz o documento original, o que inclui as suas funcionalidades e aparência.
- Mantém a sequência de *bits* original, o que traz garantias de autenticidade.
- Não necessita de uma monitorização constante ao formato do recurso digital.
- Pode oferecer os melhores resultados para a preservação de objetos digitais complexos.
- Não obriga à manutenção de dispositivos físicos.

Desvantagens

- É uma solução dispendiosa. A cada alteração significativa nas plataformas de *hardware* e *software* é necessário criar novos emuladores, o que implica pessoal especializado para tal efeito.
- É fulcral que as plataformas tecnológicas estejam bem descritas de modo a que os profissionais da computação sejam capazes de construir emuladores de acordo com as características existentes.
- Também os emuladores têm de ser alvos de ações de preservação para combater a evolução tecnológica.

Migração

A migração procura transferir objetos digitais de uma configuração de *software/hardware* mais antiga para uma outra recente e, dessa forma, superar a obsolescência tecnológica.

Vantagens

- Os procedimentos para migração simples estão bem definidos para alguns formatos.
- É a estratégia mais utilizada, o que permite maior troca de conhecimento sobre o processo e consequências.
- Não requer conhecimento técnico muito especializado.

- A tendência de utilização de *software* e formatos estabelecidos por normas e com codificação aberta permitem facilitar a migração.
- Converte o documento para um formato capaz de ser interpretado pelos sistemas atuais.
- Permite comparar os objetos migrados com os originais para verificar que características essenciais foram mantidas.

Desvantagens

- É difícil programar uma ação de migração de objetos digitais complexos, podendo ser demorada e financeiramente dispendiosa.
- Altera o documento correndo o risco de se perderem características importantes, como funcionalidades ou aparência.
- Pode comprometer a autenticidade e integridade do objeto digital e, por isso, devem ser aplicados procedimentos de controlo da qualidade.
- A sua aplicação não é recomendada em conjuntos de objetos digitais que possuam uma grande quantidade de diferentes tipos de formatos.
- Vai repetir-se ao longo do ciclo de vida do objeto digital, podendo acumular várias modificações incrementais que farão com que o objeto migrado seja invariavelmente diferente do original.

Migração para suportes analógicos

A migração para suportes analógicos consiste na conversão de objetos digitais para suportes análogos, como é o caso do papel ou microfilme.

Vantagens

- O objeto deixa de estar sujeito à obsolescência tecnológica, assumindo que o suporte de destino (microfilme, papel) é de qualidade.
- O objeto deve ser tratado como um documento de arquivo e devem-lhe ser aplicadas as estratégias de preservação física tradicionais, já bem conhecidas.
- Os suportes tradicionais permitem o acesso ao objeto por dezenas de anos.

- Pode ser considerada uma prática provisória até à implementação da infraestrutura necessária para albergar estratégias de preservação digital mais adequadas.

Desvantagens

- Perde funcionalidades do objeto digital original, o que pode levar à inutilização e à perda de integridade do documento.
- Só pode ser considerada como opção para aqueles recursos que não utilizem ou não necessitem das funcionalidades disponíveis apenas em ambiente digital.
- O facto de o objeto se encontrar num suporte tradicional, faz com que o seu acesso não seja tão facilitado e seja necessário alocar espaço físico para o receber no arquivo.
- Devem ser considerados os custos da conversão do objeto digital para analógico e também o custo das estratégias de preservação que serão aplicadas no âmbito do arquivo.

Normalização

A normalização sugere que se limite a quantidade de formatos que podem ser utilizados. A escolha deve recair sobre formatos abertos baseados em normas, amplamente usados e documentados.

Vantagens

- A utilização de padrões abertos estáveis e documentados publicamente retarda a necessidade de se aplicarem estratégias de longo prazo e diminui a complexidade de todo o processo.
- A adoção de padrões amplamente usados possibilita a existência de ferramentas capazes de os interpretar. No entanto, caso se tenham tornado obsoletas existe uma grande probabilidade de que sejam desenvolvidas novas ferramentas que garantam o acesso e interpretação dos objetos nesses formatos.
- Permite simplificar e reduzir custos de migração.

- Auxilia na distribuição de esforços ao longo do ciclo de vida dos objetos digitais.

Desvantagens

- Padrões abertos estáveis podem não estar disponíveis para alguns formatos.
- Mesmo sendo padrões estáveis podem estar sujeitos a mudanças acabando por originar novas versões. Por isso, devem ser constantemente avaliados, e se necessário, os objetos atualizados para novas versões.
- Extensões proprietárias são muito comuns, no entanto não são tão bem documentadas como os próprios padrões.
- A conversão para formatos normalizados pode implicar algum investimento.
- Do resultado da conversão podem ocorrer perdas.
- Não resolve necessariamente os problemas do acesso a longo prazo, a não ser que esta ação esteja associada a uma estratégia efetiva.

Atualização de versões de formato

A atualização de versões de formato consiste na capacidade que uma dada aplicação de *software* tem de interpretar e apresentar os objetos digitais criados por versões anteriores dessa mesma aplicação.

Vantagens

- Retarda, por um dado período, a necessidade de estratégias de preservação primárias.
- A retrocompatibilidade tem vindo a ser disponibilizada por um número crescente de fabricantes.
- Em grande parte dos casos, permite funcionalidades semelhantes às originais.

Desvantagens

- Nem todos os fabricantes disponibilizam esta funcionalidade.
- Nos casos em que o fabricante disponibiliza a funcionalidade, não é expectável que perdue indefinidamente.
- Algumas funcionalidades utilizadas podem ser descontinuadas por parte do fabricante.
- Pode introduzir modificações indesejadas nos objetos digitais, sendo que esta situação pode ser agravada se forem feitas conversões constantes.
- Trata-se de uma existência volátil pois está dependente dos mercados e pode deixar de estar disponível sem aviso prévio.
- Trata-se de uma ação que deve ser aplicada a curto e médio prazo.

Migração a-pedido

O processo de conversão baseado na migração a-pedido faz uso do objeto digital original, o que permite evitar a deturpação que pode ocorrer no final de cada ação de conversão.

Vantagens

- Faz uso do objeto digital original, o que reduz o risco de deturpações acumuladas resultantes dos processos de conversão.
- O objeto digital só é transformado no momento em que o seu acesso é requisitado.

Desvantagens

- No futuro podem não existir ferramentas disponíveis para objetos digitais complexos.
- Caso existam, podem não apresentar a totalidade dos elementos de um objeto.
- As ferramentas de acesso e os respetivos metadados devem também ser alvos de preservação.

Encapsulamento

O encapsulamento permite agrupar junto do objeto digital toda a informação necessária para a sua interpretação.

Vantagens

- Garante que todas as informações de apoio necessárias para assegurar o acesso ao objeto digital são mantidas como uma só entidade.
- Fornece uma forma útil de chamar a atenção para os elementos necessários ao acesso.

Desvantagens

- Pode originar ficheiros muito grandes devido à informação armazenada.
- Mesmo existindo toda a informação necessária para a construção de ferramentas, tal processo pode ser complicado.
- *Software* encapsulado continua sujeito à rápida obsolescência tecnológica

Anexo D – Critérios para a seleção de formatos

O relatório *Selecting File Formats for Long-Term Preservation* elaborado por Adrian Brown em 2008 para a instituição *The National Archives* reflete doze critérios com vista a auxiliar no processo de escolha dos formatos, sendo eles: ubiquidade (*ubiquity*), suporte (*support*), divulgação (*disclosure*), estabilidade (*stability*), facilidade de identificação e validação (*ease of identification and validation*), direitos de propriedade intelectual (*intellectual property rights*), suporte de metainformação (*metadata support*), complexidade (*complexity*), interoperabilidade (*interoperability*), viabilidade (*viability*) e, por fim, reutilização (*re-usability*).

Ubiquidade

As leis do mercado (oferta/procura) ditam que os formatos bem estabelecidos e comumente usados têm mais probabilidade de garantirem suporte técnico a longo prazo pelos fabricantes de *software* e amplo apoio por parte de comunidades interessadas e especializadas, em comparação com aqueles formatos apenas utilizados por nichos de mercado.

A adoção de formatos populares e amplamente utilizados deve ser uma hipótese a considerar, uma vez que existe maior probabilidade de serem desenvolvidas outras ferramentas que permitam o seu acesso e, por isso, não se tornem obsoletos tão rapidamente.

Suporte

Um dos principais fatores a considerar é a extensão do suporte do *software* atual. Deve ser considerado o número de ferramentas de *software* – e a sua ubiquidade – que podem e devem ser uma alternativa a um único *software* utilizado atualmente para interpretar um dado formato.

Divulgação

A equipa multidisciplinar responsável pela preservação digital da sua organização deve conhecer os formatos que melhor se adaptam a cada tipo de objeto

digital existente e, por isso, os seus elementos precisam de ter acesso às suas especificações técnicas para que possam tomar a melhor decisão de acordo com a informação que possuem.

Nesse seguimento, recomendam-se os formatos abertos cujas especificações estejam divulgadas publicamente.

No entanto, também as empresas que criam formatos proprietários podem publicar as suas especificações, como foi o caso do formato PDF por parte da Adobe.

Não obstante, os formatos abertos também podem ter desvantagens. No caso do ASCII (*American Standard Code for Information Interchange*) podem existir perdas a nível da estrutura, conteúdo e funcionalidades. Já no caso do PDF, preserva-se a aparência visual mas perde-se a reutilização do conteúdo.

A sofisticação e utilização de formatos abertos tende a aumentar embora, por vezes, existam funcionalidades/características que apenas o respetivo formato proprietário consegue dar resposta.

A *The National Archives* aconselha o uso de formatos abertos sempre que tal seja possível.

Qualidade da Documentação

Devem ser escolhidos formatos que possuam a sua documentação divulgada de forma pública, bem fundamentada, descrita de forma precisa e compreensível. É importante que a descrição seja de boa qualidade para que possa ser interpretada corretamente pelo ser humano e, conseqüentemente, possam ser desenvolvidas (se necessário) as ferramentas de *software* adequadas para aceder e interpretar esses formatos.

Geralmente os formatos com melhor documentação são os não-proprietários.

Estabilidade

A escolha deve recair num formato com uma especificação estável, isto é não deve estar sujeito a modificações constantes e de grande relevo ao longo do tempo.

É fulcral que novas versões de um dado formato sejam retrocompatíveis.

Facilidade de identificação e validação

A identificação de um dado formato e a confirmação de que se trata de um exemplar válido deve ser feita de forma precisa e é fulcral para o acesso contínuo.

Um formato bem projetado deve permitir a sua identificação através da análise das suas codificações.

Direitos de Propriedade Intelectual

Os formatos podem fazer uso de tecnologias protegidas por patentes ou outras restrições a nível da propriedade intelectual, sendo que tais situações podem trazer limitações no acesso presente e futuro dos objetos existentes nesse formato.

A *The National Archives* recomenda o uso de formatos sem limitações no que toca a patentes e direitos de autor.

Suporte de Metainformação

Alguns formatos permitem a inclusão de metadados, que podem ser gerados automaticamente pela aplicação e/ou inseridos pelo utilizador.

A existência da metainformação anexa ao objeto traz grandes benefícios para o processo de preservação digital, pois graças a ela é possível descrever a origem desse objeto a nível organizacional mas também o contexto tecnológico desde o momento da sua criação. Como exemplo, o formato TIFF pode conter metainformação que descreve a marca e modelo do *scanner*, versões do sistema operativo e aplicação de *software* usados, nome do criador e descrição da imagem.

De preferência devem ser escolhidos formatos que suportem metainformação.

Complexidade

Um formato deve ser escolhido baseando-se na premissa de que possui as características adequadas para suportar os requisitos funcionais previstos.

Não deve ser escolhido um formato com especificações/funcionalidades em excesso, pois quanto mais complexo for o formato, mais caros se irão tornar os processos de gestão e preservação.

Interoperabilidade

A *The National Archives* considera de grande relevância a possibilidade de troca de informação entre utilizadores e sistemas. É, então, desejável o uso de formatos que possam ser interpretados por uma ampla gama de aplicações e, por isso, sejam independentes de plataforma, como é o caso dos formatos abertos.

O facto de um dado formato ser independente de plataforma é ainda uma garantia de sustentabilidade a longo prazo pois permite que um objeto seja transferido de um ambiente digital para outro sem que haja perda de informação.

Viabilidade

Alguns formatos oferecem meios para deteção e correção de erros, de modo a revelar uma possível corrupção do ficheiro (por exemplo no processo de transferência).

Os formatos que fazem uso deste tipo de mecanismos (como o PNG que possui sequências de *bytes* para verificar a ocorrência de diferentes tipos de erros) são mais robustos e, por isso, o seu uso é aconselhado.

Reutilização

Existem alguns tipos de dados que necessitam de manter a sua capacidade de serem processados. Por exemplo, se um documento de folha de cálculo for convertido para PDF perde a capacidade de processamento dos dados. Assim sendo, deve ser considerada a manutenção de uma versão desse objeto digital que permita o seu processamento futuro.

Anexo E – Lista de formatos adequados para preservação

O Arquivo da Web Portuguesa distingue cinco tipos de formatos (texto, imagem, áudio, vídeo e outros) e subdivide cada um destes tipos de acordo com o nível de adequação para preservação (alta, média e baixa).

Texto

Alta adequação para preservação

- HTML, XHTML ou XML, contendo referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Textos planos usando codificação de caracteres UTF-8, USASCII ou UTF-16 com Byte Order Mark (.txt)
- PDF/A-1 segundo a norma ISO 19005-1 (.pdf)
- Open Document Text (.odt)

Média adequação para preservação

- HTML, XHTML ou XML, sem referência para o DTD ou esquema de validação de formato (.html, .xhtml, .xml)
- Cascading Style Sheets (.css)
- Texto plano com codificação de caracteres em ISO-8859-1 (.txt)
- PDF com fontes embebidas (.pdf)
- Rich Text Format 1.x (.rtf)
- HTML 4.x incluindo declaração de DOCTYPE (.html)
- Open Office Text Document (.sxw)
- Office Open XML (.docx)
- DTD (.dtd)
- SGML (.sgml)

Baixa adequação para preservação

- Microsoft Word (.doc)
- Postscript (.ps)
- PDF cifrado (.pdf)

- WordPerfect (.wpd)
- DVI (.dvi)

Imagem

Alta adequação para preservação

- PNG (.png)
- JPEG2000 usando compressão sem perda de informação (.jp2)
- TIFF sem compressão
- SVG (.svg)

Média adequação para preservação

- JPEG2000 usando compressão com perdas de informação (.jp2)
- GIF (.gif)
- JPEG/JFIF (.jpg)
- TIFF com compressão
- BMP (.bmp)
- Digital Negative (.dng)
- Computer Graphic Metafile e WebCGM (.cgm)

Baixa adequação para preservação

- Macromedia Flash (*.swf)
- PhotoShop (.psd)
- JPEG 2000 Part 2 (.jpf, .jpx)
- MrSID (.sid)
- TIFF em formato Planar (.tiff)
- FlashPix (.fpx)
- RAW
- Encapsulated Postscript (.eps)

Áudio

Alta adequação para preservação

- AIFF com Pulse-code modulation (.aif, .aiff)
- WAV com Pulse-code modulation (.wav, bwf)
- Ogg Vorbis (.ogg, .oga)

Média adequação para preservação

- MP3 (MPEG-1/2, Layer 3) (.mp3)
- Free Lossless Audio Codec (.flac)
- SUN Audio sem compressão (.au)
- Standard MIDI (.mid, midi)
- Advance Audio Coding (.mp4, .m4a, .aac)

Baixa adequação para preservação

- RealNetworks 'Real Audio' (.ra, .rm, .ram)
- Windows Media Audio (.wma)
- WAV com compressão (.wav)
- AIFC com compressão (.aifc)
- NeXT SND(.snd)

Vídeo

Alta adequação para preservação

- QuickTime Movie sem compressão (.mov)
- AVI sem compressão (.avi)
- Motion JPEG 2000 (ISO/IEC 15444-4) (mj2)
- Motion JPEG (.avi, .mov)

Média adequação para preservação

- MPEG-1, MPEG-2 (.mpg, .mpeg)
- MPEG-4 (.mp4)
- Ogg Theora (.ogg, .ogm, .ogv)

Baixa adequação para preservação

- Windows Media Video (.wmv)
- AVI com compressão (.avi)
- QuickTime Movie com compressão (.mov)
- RealNetworks 'Real Video' (.rv, .rm)

Outros formatos

Alta adequação para preservação

- Comma Separated Values (.csv)
- SQL DDL

Média adequação para preservação

- OpenOffice (.sxc/.ods, .sxi/.odp)
- OOXML segundo a norma ISO/IEC DIS 29500 (.xlsx, .pptx)

Baixa adequação para preservação

- Microsoft Excel (.xls)
- Microsoft PowerPoint (.ppt)
- Microsoft Access (.mdb)
- Microsoft Visio (.vsd)

Anexo F – Exemplos de esquemas de metainformação

Dicionário de Dados PREMIS

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns="info:lc/xmlns/premis-v2"
targetNamespace="info:lc/xmlns/premis-v2" elementFormDefault="qualified" attributeFormDefault="unqualified">
<xs:import namespace="http://www.w3.org/1999/xlink" schemaLocation="http://www.loc.gov/standards/xlink/xlink.xsd"/>

<xs:element name="premis" type="premisComplexType"/>
<xs:element name="object" type="objectComplexType"/>
<xs:element name="event" type="eventComplexType"/>
<xs:element name="agent" type="agentComplexType"/>
<xs:element name="rights" type="rightsComplexType"/>

<xs:complexType name="premisComplexType">
  <xs:sequence>
    <xs:element ref="object" maxOccurs="unbounded"/>
    <xs:element ref="event" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="agent" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="rights" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="version" type="versionSimpleType" use="required"/>
</xs:complexType>

<xs:complexType name="objectComplexType" abstract="true"/>
<xs:complexType name="file">
  <xs:complexContent>
    <xs:extension base="objectComplexType">
      <xs:sequence>
        <xs:element ref="objectIdentifier" maxOccurs="unbounded"/>
        <xs:element ref="preservationLevel" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="significantProperties" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="objectCharacteristics" maxOccurs="unbounded"/>
        <xs:element ref="originalName" minOccurs="0"/>
        <xs:element ref="storage" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="environment" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="signatureInformation" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="relationship" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingEventIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingIntellectualEntityIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingRightsStatementIdentifier" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="xmlID" type="xs:ID"/>
      <xs:attribute name="version" type="versionSimpleType"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

Fonte: Library of Congress, 2012 [Em linha]

Figura 25 – Excerto do esquema XML que suporta a implementação do Dicionário de dados PREMIS (1/2)

```

<xs:complexType name="representation">
  <xs:complexContent>
    <xs:extension base="objectComplexType">
      <xs:sequence>
        <xs:element ref="objectIdentifier" maxOccurs="unbounded"/>
        <xs:element ref="preservationLevel" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="significantProperties" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="originalName" minOccurs="0"/>
        <xs:element ref="environment" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="relationship" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingEventIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingIntellectualEntityIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingRightsStatementIdentifier" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="xmlID" type="xs:ID"/>
      <xs:attribute name="version" type="versionSimpleType"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

<xs:complexType name="bitstream">
  <xs:complexContent>
    <xs:extension base="objectComplexType">
      <xs:sequence>
        <xs:element ref="objectIdentifier" maxOccurs="unbounded"/>
        <xs:element ref="significantProperties" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="objectCharacteristics" maxOccurs="unbounded"/>
        <xs:element ref="storage" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="environment" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="signatureInformation" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="relationship" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingEventIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingIntellectualEntityIdentifier" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="linkingRightsStatementIdentifier" minOccurs="0" maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="xmlID" type="xs:ID"/>
      <xs:attribute name="version" type="versionSimpleType"/>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>

<xs:complexType name="eventComplexType">
  <xs:sequence>
    <xs:element ref="eventIdentifier"/>
    <xs:element ref="eventType"/>
    <xs:element ref="eventDateTime"/>
    <xs:element ref="eventDetail" minOccurs="0"/>
    <xs:element ref="eventOutcomeInformation" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="linkingAgentIdentifier" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="linkingObjectIdentifier" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="xmlID" type="xs:ID"/>
  <xs:attribute name="version" type="versionSimpleType"/>
</xs:complexType>

<xs:complexType name="agentComplexType">
  <xs:sequence>
    <xs:element ref="agentIdentifier" maxOccurs="unbounded"/>
    <xs:element ref="agentName" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="agentType" minOccurs="0"/>
    <xs:element ref="agentNote" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="agentExtension" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="mdSec" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="linkingEventIdentifier" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element ref="linkingRightsStatementIdentifier" minOccurs="0" maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="xmlID" type="xs:ID"/>
  <xs:attribute name="version" type="versionSimpleType"/>
</xs:complexType>

<xs:complexType name="rightsComplexType">
  <xs:choice maxOccurs="unbounded">
    <xs:element ref="rightsStatement"/>
    <xs:element ref="rightsExtension"/>
    <xs:element ref="mdSec"/>
  </xs:choice>
  <xs:attribute name="xmlID" type="xs:ID"/>
  <xs:attribute name="version" type="versionSimpleType"/>
</xs:complexType>
</xs:schema>

```

Fonte: Library of Congress, 2012 [Em linha]

Figura 26 – Excerto do esquema XML que suporta a implementação do Dicionário de dados PREMIS (2/2)

```

<object xmlns="http://www.loc.gov/standards/premis/v1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/standards/premis/v1 http://www.loc.gov/standards/premis/v1/Object-v1-1.xsd">
  <objectIdentifier>
    <objectIdentifierType>handle</objectIdentifierType>
    <objectIdentifierValue>yale.dp/yc03825-001</objectIdentifierValue>
  </objectIdentifier>
  <preservationLevel>2</preservationLevel>
  <objectCategory>file</objectCategory>
  <objectCharacteristics>
    <compositionLevel>0</compositionLevel>
    <fixity>
      <messageDigestAlgorithm>CRC32</messageDigestAlgorithm>
      <messageDigest>49a566ed</messageDigest>
    </fixity>
    <fixity>
      <messageDigestAlgorithm>MD5</messageDigestAlgorithm>
      <messageDigest>3a659f1f9bbb477841cc72ecbdb48b5</messageDigest>
    </fixity>
    <fixity>
      <messageDigestAlgorithm>SHA-1</messageDigestAlgorithm>
      <messageDigest>f73f70fe941cf13c49c5070a5f74477b72d1535</messageDigest>
    </fixity>
    <size>33012564</size>
    <format>
      <formatDesignation>
        <formatName>TIFF</formatName>
        <formatVersion>6.0</formatVersion>
      </formatDesignation>
      <formatRegistry>
        <formatRegistryName>PRONOM</formatRegistryName>
        <formatRegistryKey>fmt/10</formatRegistryKey>
        <formatRegistryRole>specification</formatRegistryRole>
      </formatRegistry>
      <formatRegistry>
        <formatRegistryName>IANA</formatRegistryName>
        <formatRegistryKey>image/tiff</formatRegistryKey>
        <formatRegistryRole>specification</formatRegistryRole>
      </formatRegistry>
    </format>
  </objectCharacteristics>
  <creatingApplication>
    <creatingApplicationName>Adobe Photoshop CS2</creatingApplicationName>
    <creatingApplicationVersion>9.0</creatingApplicationVersion>
    <dateCreatedByApplication>2005-10-06</dateCreatedByApplication>
  </creatingApplication>
  <originalName>yc03825.tif</originalName>
  <linkingEventIdentifier>
    <linkingEventIdentifierType>Yale Repository Event ID</linkingEventIdentifierType>
    <linkingEventIdentifierValue>E2006-07-27T18:39:37Z</linkingEventIdentifierValue>
  </linkingEventIdentifier>
</object>

```

Fonte: Yale University Library, 2006 [Em linha]

Figura 27 – Exemplo de um registo correspondente ao elemento Objeto

```

<event xmlns="http://www.loc.gov/standards/premis/v1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.loc.gov/standards/premis/v1 http://www.loc.gov/standards/premis/v1/Event-v1-1.xsd">
  <eventIdentifier>
    <eventIdentifierType>Yale Repository Event ID</eventIdentifierType>
    <eventIdentifierValue>E2006-08-02T14:27:10Z</eventIdentifierValue>
  </eventIdentifier>
  <eventType>message digest calculation</eventType>
  <eventDateTime>2006-08-02T14:27:10Z</eventDateTime>
  <eventDetail>E2006-08-02T142710Z.txt</eventDetail>
  <eventOutcomeInformation>
    <eventOutcome>1</eventOutcome>
  </eventOutcomeInformation>
  <linkingAgentIdentifier>
    <linkingAgentIdentifierType>Yale Net ID</linkingAgentIdentifierType>
    <linkingAgentIdentifierValue>ab12</linkingAgentIdentifierValue>
  </linkingAgentIdentifier>
  <linkingObjectIdentifier>
    <linkingObjectIdentifierType>handle</linkingObjectIdentifierType>
    <linkingObjectIdentifierValue>yale.dp/yc03825-002</linkingObjectIdentifierValue>
  </linkingObjectIdentifier>
</event>

```

Fonte: Yale University Library, 2006 [Em linha]

Figura 28 – Exemplo de um registo correspondente ao elemento Evento

Dublin Core

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/ http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/">

  <dc:title>UKOLN</dc:title>
  <dcterms:alternative>UK Office for Library and Information Networking</dcterms:alternative>
  <dc:subject>
    national centre, network information support, library
    community, awareness, research, information services,public
    library networking, bibliographic management, distributed
    library systems, metadata, resource discovery,
    conferences,lectures, workshops
  </dc:subject>
  <dc:subject xsi:type="dcterms:DDC">062</dc:subject>
  <dc:subject xsi:type="dcterms:UDC">061(410)</dc:subject>
  <dc:description>
    UKOLN is a national focus of expertise in digital information
    management. It provides policy, research and awareness services
    to the UK library, information and cultural heritage communities.
    UKOLN is based at the University of Bath.
  </dc:description>
  <dc:description xml:lang="fr">
    UKOLN est un centre national d'expertise dans la gestion de l'information
    digitale.
  </dc:description>
  <dc:publisher>UKOLN, University of Bath</dc:publisher>
  <dcterms:isPartOf xsi:type="dcterms:URI">
    http://www.bath.ac.uk/
  </dcterms:isPartOf>
  <dc:identifier xsi:type="dcterms:URI">
    http://www.ukoln.ac.uk/
  </dc:identifier>
  <dcterms:modified xsi:type="dcterms:W3CDTF">
    2001-07-18
  </dcterms:modified>
  <dc:format xsi:type="dcterms:IMT">
    text/html
  </dc:format>
  <dcterms:extent>
    14 Kbytes
  </dcterms:extent>
</metadata>
```

Fonte: Powell e Johnston, 2012 [Em linha]

Figura 29 – Exemplo de um registo Dublin Core

METS

```
< mets: mets xmlns: mets="http://www.loc.gov/METS/" xmlns: mods="http://www.loc.gov/mods/v3" xmlns: rts="http://cosimo.stanford.edu/sdr/metsrights/"
xmlns: mix="http://www.loc.gov/mix/v10" xmlns: xlink="http://www.w3.org/1999/xlink" xmlns: xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi: schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd http://cosimo.stanford.edu/sdr/metsrights/
http://cosimo.stanford.edu/sdr/metsrights.xsd http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-2.xsd http://www.loc.gov/mix/v10
http://www.loc.gov/standards/mix/mix10/mix10.xsd" OBJID="ark:/13030/hb4199p148" LABEL="Extending the Lexicon by Exploiting Subregularities"
PROFILE="http://www.loc.gov/mets/profiles/00000013.xml">

  < mets: metsHdr CREATEDATE="2006-01-03T11:25:07" LASTMODDATE="2009-09-22T04:28:25">
    < mets: agent ROLE="CREATOR" TYPE="ORGANIZATION">
      < mets: name>Library Systems Office, University of California, Berkeley</mets: name>
    </mets: agent>
  </mets: metsHdr>

  < mets: dmdSec ID="DMR1">
    < mets: mdRef xlink: href="http://oskicat.berkeley.edu/search/v?SEARCH=GLADN184985770" LOCTYPE="URL" MDTYPE="MARC" LABEL="Catalog Record"/>
  </mets: dmdSec>

  < mets: dmdSec ID="DM1">
    < mets: mdWrap MDTYPE="MODS" LABEL="Extending the Lexicon by Exploiting Subregularities">
      < mets: xmlData>
        < mods: mods>
          < mods: titleInfo>
            < mods: title>Extending the Lexicon by Exploiting Subregularities</mods: title>
          </mods: titleInfo>
          < mods: name type="personal">
            < mods: namePart>Wilensky, Robert</mods: namePart>
            < mods: role>
              < mods: roleTerm type="text" authority="marcrelator">Author</mods: roleTerm>
            </mods: role>
          </mods: name>
          < mods: typeOfResource>text</mods: typeOfResource>
          < mods: genre authority="aat">technical reports</mods: genre>
          < mods: originInfo>
            < mods: place>
              < mods: placeTerm type="text">University of California at Berkeley, Berkeley, California</mods: placeTerm>
            </mods: place>
            < mods: publisher>Computer Science Division</mods: publisher>
            < mods: dateCreated>February 1991</mods: dateCreated>
          </mods: originInfo>
          < mods: language>
            < mods: languageTerm type="code" authority="iso639-2b">eng</mods: languageTerm>
          </mods: language>
          < mods: physicalDescription>
            < mods: extent>32 p</mods: extent>
          </mods: physicalDescription>
          < mods: relatedItem displayLabel="Metacollection" type="host">
            < mods: titleInfo>
              < mods: title>technical reports</mods: title>
            </mods: titleInfo>
            < mods: identifier type="local search">hb1k40068r</mods: identifier>
          </mods: relatedItem>
          < mods: relatedItem displayLabel="Metacollection" type="host">
            < mods: titleInfo>
              < mods: title>EECS</mods: title>
            </mods: titleInfo>
            < mods: identifier type="local search">bk0000m744s</mods: identifier>
          </mods: relatedItem>
          < mods: identifier type="local">CSD-91-618</mods: identifier>
          < mods: location>
            < mods: physicalLocation>Engineering Library, 110 Bechtel, University of California, Berkeley, CA 94720-1796</mods: physicalLocation>
          </mods: location>
        </mods: mods>
      </mets: xmlData>
    </mets: mdWrap>
  </mets: dmdSec>
```

Fonte: Library of Congress, 2011^b [Em linha]

Figura 30 – Exemplo de um registo METS (1/2)

```

<mets:amdSec>
  <mets:rightsMD ID="RMD1">
    <mets:mdWrap MDTYPE="OTHER" OTHERMDTYPE="METSRights">
      <mets:xmlData>
        <rts:RightsDeclarationMD RIGHTSCATEGORY="COPYRIGHTED">
          <rts:RightsHolder>
            <rts:RightsHolderName>The author(s)</rts:RightsHolderName>
            <rts:RightsHolderContact>
              <rts:RightsHolderContactAddress>
                Computer Science Division, Department of Electrical Engineering & Computer Sciences, University of California, Berkeley
              </rts:RightsHolderContactAddress>
            </rts:RightsHolderContact>
          </rts:RightsHolder>
          <rts:Context CONTEXTCLASS="GENERAL PUBLIC">
            <rts:Constraints>
              <rts:ConstraintDescription>
                Transmission or reproduction of materials protected by copyright beyond that allowed by fair use requires the written permission of
                the copyright owners. Works not in the public domain cannot be commercially exploited without permission of the copyright owner.
                Responsibility for any use rests exclusively with the user.
              </rts:ConstraintDescription>
            </rts:Constraints>
          </rts:Context>
        </rts:RightsDeclarationMD>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
</mets:amdSec>

<mets:fileSec>
  <mets:fileGrp USE="application">
    <mets:file ID="FID1" MIMETYPE="application/pdf" SEQ="1" GROUPID="GID1">
      <mets:Flocat xlink:href="http://nma.berkeley.edu/ark:/28722/bk0005j7x2r" LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
  <mets:fileGrp USE="application">
    <mets:file ID="FID2" MIMETYPE="application/ps" SEQ="1" GROUPID="GID1">
      <mets:Flocat xlink:href="http://nma.berkeley.edu/ark:/28722/bk0005j7x39" LOCTYPE="URL"/>
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>

<mets:structMap>
  <mets:div TYPE="text" LABEL="Extending the Lexicon by Exploiting Subregularities" ADMID="RMD1" DMDID="DMR1 DM1">
    <mets:fptr FILEID="FID1"/>
    <mets:fptr FILEID="FID2"/>
  </mets:div>
</mets:structMap>
</mets:mets>

```

Fonte: Library of Congress, 2011^b [Em linha]

Figura 31 – Exemplo de um registo METS (2/2)

Anexo G – Apresentação do questionário

Este inquérito surge no âmbito da Dissertação de Mestrado em Informação Empresarial da Escola Superior de Estudos Industriais e de Gestão (ESEIG) do Instituto Politécnico do Porto (IPP) e tem como tema de estudo "**Preservação digital nas médias e grandes empresas portuguesas: estudo de caso e recomendação de boas práticas**".

Com este questionário pretende-se perceber se as empresas portuguesas estão já sensibilizadas para a temática da preservação digital e quais as estratégias que privilegiam. Pretende-se também compreender que tipos de documentos são preservados digitalmente e tentar concluir se as empresas portuguesas que já adotaram estratégias de preservação o fizeram por imposição legal ou como forma de registo do seu âmbito de atuação cultural e patrimonial.

A sua participação neste inquérito é um contributo para a realização e sucesso de um trabalho académico. Não lhe serão colocadas questões de foro privado ou de natureza confidencial.

As informações por si prestadas são confidenciais e apenas serão acedidas pela autora do estudo e pelos seus orientadores, não sendo possível identificar as instituições participantes na análise de resultados.

Caso pretenda, os resultados do estudo serão enviados para V. Ex.^a bastando para isso preencher o campo indicado no formulário de questionário.

A autora agradece antecipadamente a sua participação neste estudo.

Anexo H – Inquérito por questionário

Grupo I – Perfil da empresa

1. Como caracteriza a sua organização?

Escolha uma das seguintes respostas

- Média (Entre 50 a 250 efetivos; entre 10 a 50 M€ de Volume de Negócios ou entre 10 a 43 M€ de Balanço Total)
- Grande

Fonte: *Enterprise Europe Network*

(<http://www.enterpriseeuropenetwork.pt/info/pol%20serv/pol%20C3%ADticas/Paginas/p1.aspx>)

2. A sua empresa é uma organização pública ou privada?

Escolha uma das seguintes respostas

- Empresa pública
- Empresa privada

3. Em que distrito ou região autónoma de Portugal se localiza a sua empresa?

Escolha uma das seguintes respostas

- | | |
|---|---|
| <input type="checkbox"/> Aveiro | <input type="checkbox"/> Lisboa |
| <input type="checkbox"/> Beja | <input type="checkbox"/> Portalegre |
| <input type="checkbox"/> Braga | <input type="checkbox"/> Porto |
| <input type="checkbox"/> Bragança | <input type="checkbox"/> Santarém |
| <input type="checkbox"/> Castelo Branco | <input type="checkbox"/> Setúbal |
| <input type="checkbox"/> Coimbra | <input type="checkbox"/> Viana do Castelo |
| <input type="checkbox"/> Évora | <input type="checkbox"/> Vila Real |
| <input type="checkbox"/> Faro | <input type="checkbox"/> Viseu |
| <input type="checkbox"/> Guarda | <input type="checkbox"/> Região Autónoma da Madeira |
| <input type="checkbox"/> Leiria | <input type="checkbox"/> Região Autónoma dos Açores |

4. A sede social da sua empresa situa-se no local escolhido na pergunta anterior? (se empresa privada na questão 2)

- Sim
- Não

5. A sede social da sua empresa situa-se em Portugal? (se Não na questão 4)

- Sim
- Não

6. Em que país se encontra a sede social da sua empresa? (se Não na questão 5)

7. Qual o principal âmbito de atuação da sua empresa ou aquele que corresponde ao maior volume de negócios? (se empresa pública na questão 2)

Escolha uma das seguintes respostas

- | | |
|---|--|
| <input type="checkbox"/> Administração Central | <input type="checkbox"/> Indústria extrativa |
| <input type="checkbox"/> Administração Local | <input type="checkbox"/> Indústria farmacêutica |
| <input type="checkbox"/> Apoio social | <input type="checkbox"/> Indústria petrolífera/combustível/gás natural |
| <input type="checkbox"/> Comércio a retalho | <input type="checkbox"/> Indústria têxtil |
| <input type="checkbox"/> Comércio por grosso | <input type="checkbox"/> Instituição bancária/financeira |
| <input type="checkbox"/> Comunicação social | <input type="checkbox"/> Publicidade |
| <input type="checkbox"/> Consultoria | <input type="checkbox"/> Produção de máquinas/equipamentos elétricos |
| <input type="checkbox"/> Cuidados de saúde | <input type="checkbox"/> Produção de papel |
| <input type="checkbox"/> Engenharia civil | <input type="checkbox"/> Produção de vidro |
| <input type="checkbox"/> Indústria agropecuária | <input type="checkbox"/> Segurança |
| <input type="checkbox"/> Indústria automóvel | <input type="checkbox"/> Siderurgia |
| <input type="checkbox"/> Indústria de bens alimentares | <input type="checkbox"/> Tecnologias |
| <input type="checkbox"/> Indústria de energia elétrica | <input type="checkbox"/> Telecomunicações |
| <input type="checkbox"/> Indústria livreira | <input type="checkbox"/> Turismo |
| <input type="checkbox"/> Indústria de produtos químicos | <input type="checkbox"/> Transportes |
| <input type="checkbox"/> Indústria de mobiliário | <input type="checkbox"/> Outro: _____ |

8. Qual o principal âmbito de atuação da sua empresa ou aquele que corresponde ao maior volume de negócios? (se empresa privada na questão 2)

Escolha uma das seguintes respostas

- | | |
|---|--|
| <input type="checkbox"/> Apoio social | <input type="checkbox"/> Indústria farmacêutica |
| <input type="checkbox"/> Comércio a retalho | <input type="checkbox"/> Indústria petrolífera/combustível/gás natural |
| <input type="checkbox"/> Comércio por grosso | <input type="checkbox"/> Indústria têxtil |
| <input type="checkbox"/> Comunicação social | <input type="checkbox"/> Instituição bancária/financeira |
| <input type="checkbox"/> Consultoria | <input type="checkbox"/> Publicidade |
| <input type="checkbox"/> Cuidados de saúde | <input type="checkbox"/> Produção de máquinas/equipamentos elétricos |
| <input type="checkbox"/> Engenharia civil | <input type="checkbox"/> Produção de papel |
| <input type="checkbox"/> Indústria agropecuária | <input type="checkbox"/> Produção de vidro |
| <input type="checkbox"/> Indústria automóvel | <input type="checkbox"/> Segurança |
| <input type="checkbox"/> Indústria de bens alimentares | <input type="checkbox"/> Siderurgia |
| <input type="checkbox"/> Indústria de energia elétrica | <input type="checkbox"/> Tecnologias |
| <input type="checkbox"/> Indústria livreira | <input type="checkbox"/> Telecomunicações |
| <input type="checkbox"/> Indústria de produtos químicos | <input type="checkbox"/> Turismo |
| <input type="checkbox"/> Indústria de mobiliário | <input type="checkbox"/> Transportes |
| <input type="checkbox"/> Indústria extrativa | <input type="checkbox"/> Outro: _____ |

Grupo II – Gestão Documental

1. Na sua organização procura-se utilizar documentos digitais como forma de substituição dos documentos em papel?

- Sim
- Não

2. Na sua organização existem programas (*software*) que auxiliem na gestão documental?

- Sim
- Não

3. Quais os programas (*software*) de gestão de documental utilizados na sua empresa? (se sim na questão 2)

Selecione todas as que se apliquem

- Alfresco
- filedoc
- iPortalDoc
- M-Files
- MoreDoc
- Mytho
- Primavera
- Outros: _____

4. Na sua empresa existe alguma recomendação de utilização de determinados formatos na criação, exportação e armazenamento de objetos digitais?

- Sim
- Não

5. Quais os formatos recomendados para o armazenamento de documentos de texto?

(se Sim na questão 4)

Selecione todas as que se apliquem

- DOC
- DOCX

- ODT
- PDF
- XML
- Outros: _____

6. Quais os formatos recomendados para o armazenamento de folhas de cálculo? (se Sim na questão 4)

Selecione todas as que se apliquem

- XLS
- XLSX
- CSV/TSV
- ODS
- PDF
- XML
- Não aplicável
- Outros: _____

7. Quais os formatos recomendados para o armazenamento de imagens? (se Sim na questão 4)

Selecione todas as que se apliquem

- JPEG
- PNG
- GIF
- TIFF
- SVG
- Não aplicável
- Outros: _____

8. Qual o formato recomendado para o armazenamento de *emails*? (se Sim na questão 4)

Selecione todas as que se apliquem

- PST
- CSV
- TXT

- XML
- Não aplicável
- Outros: _____

9. Que tecnologias são utilizadas, na sua empresa, para o armazenamento e cópia de dados?

Selecione todas as que se apliquem

- Discos rígidos
- RAID
- USB
- CD
- DVD
- Tape
- Outros: _____

10. Que cuidados existem com o *hardware* responsável pelo armazenamento e cópia de documentos na sua organização?

Selecione todas as que se apliquem

- Temperatura e humidade estáveis e controladas
- Proteção contra desastres (inundações/fogo)
- Inspeção frequente
- Cópias de segurança
- Não aplicável
- Outros: _____

Grupo III – Preservação digital

1. Na sua organização existe preocupação em manter o acesso à informação por um longo período de tempo?

- Sim
- Não

2. Conhece o conceito de preservação digital?

Ajuda: Preservação digital é o “conjunto de actividades ou processos responsáveis por garantir o acesso continuado a longo-prazo à informação e restante património cultural existente em formatos digitais” (Webb citado por Ferreira, 2006, p. 20).

- Sim
- Não

3. Na sua organização são aplicadas estratégias de preservação digital?

Escolha uma das seguintes respostas

- Sim
- Não
- Não tenho conhecimento

4. Na sua organização a elaboração do plano de preservação digital e respetiva intervenção são realizadas por uma equipa multidisciplinar? (se Sim na questão 3)

- Sim
- Não

5. Qual o cargo que o responsável pela preservação digital ocupa na empresa? (se Sim na questão 3)

6. Quais as estratégias de preservação digital utilizadas na sua empresa? (se Sim na questão 3)

Selecione todas as que se apliquem

- Migração (consiste na transferência de objetos digitais para outro tipo de tecnologias ou para a geração de tecnologia seguinte)

- Emulação (é a utilização de *software* capaz de reproduzir o funcionamento de uma plataforma, estando a atuar sob outra que é incompatível)
- Preservação da tecnologia (trata-se da conservação do ambiente físico e lógico existente no momento da criação do objeto digital que se pretende preservar)
- Refrescamento (consiste na transferência dos dados de determinado suporte físico de armazenamento, antes que essa tecnologia se torne obsoleta ou se deteriore, para outros mais recentes)
- Outras: _____

7. Que tipos de documentos a sua organização procura manter acessíveis a longo prazo? (se Sim na questão 3)

Selecione todas as que se apliquem

- Regras/procedimentos/normas/manuais
- Atas
- Notas de serviço
- Ofícios
- Contratos
- Reclamações/queixas
- Processos disciplinares
- Autos/notificações
- Financeiros (faturas/recibos)
- Recursos humanos
- Outros: _____

8. A manutenção de documentos a longo prazo é feita por imposição legal ou para conservar o património histórico-cultural da organização? (se Sim na questão 3)

Selecione todas as que se apliquem

- Imposição legal
- Conservação de património
- Outros: _____

9. Como quantifica o investimento feito pela sua organização na implementação do plano de preservação digital, quer em infraestrutura tecnológica como em recursos humanos? (se Sim na questão 3)

Escolha uma das seguintes respostas

- Elevado
- Médio
- Baixo

10. Gostaria de ver implementado um plano de preservação digital na sua empresa? (se Não na questão 3)

Escolha uma das seguintes respostas

- Sim
- Não
- É-me indiferente

11. Reconhece a importância da utilização de formatos abertos para a preservação digital?

Escolha uma das seguintes respostas

Ajuda: Formato aberto é uma especificação para *software* disponibilizada de forma livre e gratuita para qualquer indivíduo sem que existam limitações na sua utilização.

- Sim
- Não
- Não tenho conhecimento

Grupo IV – Interesse na receção dos resultados e comentários

1. Pretende receber os resultados deste estudo?

Sim

Não

2. Por favor, indique o *email* para o qual devem ser enviadas as conclusões do estudo:

(se Sim na questão 1)

3. Se assim o desejar pode deixar um comentário ou esclarecer alguma informação relativa às questões do questionário realizado.

Anexo I – Recomendação de ações que visam garantir a sustentabilidade da informação digital

Embora a utilização de informação em formato digital traga muitas vantagens como sejam a aptidão que os dispositivos atuais têm para processar grandes quantidades de dados com relativa facilidade, ou a possibilidade de armazenar esses dados em dispositivos cada vez mais pequenos mas com maior capacidade, a verdade é que também existem algumas desvantagens.

Curiosamente, até a própria rapidez com que esta tecnologia evolui pode ser uma grande desvantagem. Os ciclos de atualização de *software* e *hardware*, cada vez mais rápidos – graças a um mercado altamente competitivo – fazem com que a tecnologia existente depressa se torne obsoleta e, conseqüentemente, a informação se torne vulnerável.

Para evitar que o acesso aos objetos digitais se perca é necessário aplicar um conjunto de ações desde o momento da sua criação/digitalização, ao longo do seu ciclo de vida ativa e, também, ao longo da sua vida em arquivo. São algumas destas ações que se pretendem redigir como síntese deste trabalho, com o objetivo principal de auxiliar as empresas no processo de preservação digital. Não obstante, algumas destas ações podem também ser aplicadas por indivíduos comuns que tenham essa preocupação no seu quotidiano.

Estipular regras de nomeação de ficheiros

Em virtude da colossal evolução tecnológica a que temos assistido, cada vez mais se prefere usar documentos digitais em detrimento dos documentos em papel. Todavia, para que tal utilização seja eficaz, urge a necessidade de recorrer a procedimentos capazes de simplificar a gestão e manutenção dos documentos digitais. Entre elas está o levantamento dos tipos de documentos digitais criados e recebidos e a respetiva criação de normas de nomeação. A aplicação destas regras por parte de todos os colaboradores da organização permite uniformizar os nomes dos documentos e, dessa forma, identificar facilmente cada documento, nomeadamente o seu tipo, auxiliando no processo futuro de pesquisa de informação.

Criação de perfis de acesso à informação

Muitas vezes os documentos são colocados em pastas partilhadas com diversos colaboradores. No entanto, pode acontecer que nem todos os colaboradores devam ter acesso a todos os conteúdos aí armazenados.

Por isso, recomenda-se a criação de pastas com diferentes perfis de acesso, nas quais são colocados os objetos digitais que apenas um determinado grupo de utilizadores pode aceder. Deve-se também definir os privilégios de cada perfil de utilizador que acede a uma determinada pasta, isto é, para cada utilizador, se este pode criar objetos digitais, editar, eliminar ou apenas visualizar.

Armazenamento *online*

O armazenamento de objetos digitais através de sistemas *online* (na *cloud*, como é habitualmente designado) é cada vez mais utilizado. Contudo, tal como foi referido anteriormente, deve-se ter atenção a que utilizadores podem aceder a cada pasta e aos seus privilégios de acesso.

Existem dois tipos de plataformas: aquelas que permitem apenas o armazenamento e outras que, para além do armazenamento, permitem também o trabalho colaborativo.

O trabalho colaborativo consiste na possibilidade de diferentes pessoas estarem a editar o documento ao mesmo tempo e esse facto pode acarretar alguns riscos (vejam-se como exemplos o Google Docs ou o novo Office 360). Por isso, é importante que se faça a divisão de tarefas entre os colaboradores, para que não estejam a editar em simultâneo a mesma secção do documento. Deve também ser feito um controlo de versões de modo a que fiquem registadas as várias modificações feitas a um documento e quem as realizou. Isso permite que se possa voltar a uma versão anterior caso algo falhe e não se perca todo o trabalho efetuado, bem como para ser possível obter informação sobre o responsável por determinada parte de um documento.

Estas plataformas digitais apresentam quer planos gratuitos, com capacidades de armazenamento e funcionalidades limitadas, como planos *premium* onde algumas destas limitações são minimizadas. Nos planos *premium* podemos incluir planos orientados para as organizações, em que se disponibilizam quotas de

armazenamento para cada colaborador e se adicionam funcionalidades como a recuperação ilimitada de ficheiros, que pode ser uma mais-valia caso ocorram ações indevidas nos ficheiros.

É importante que ao escolher a plataforma mais adequada para a sua organização se tenha em atenção a política de privacidade praticada. Muitas das vezes, uma organização gera uma grande quantidade de documentos confidenciais e é conveniente que a plataforma selecionada dê garantias de não aceder, recolher ou transmitir a terceiros as informações armazenadas.

As organizações ou indivíduos que façam uso deste tipo de serviço devem estar atentas às comunicações enviadas pela empresa prestadora do serviço e que comuniquem eventuais alterações aos termos de serviço, particularmente no caso dos planos gratuitos. Estas comunicações devem ser feitas atempadamente para que seja dada a possibilidade de os utilizadores se adaptarem à nova realidade, especialmente em caso de diminuição da capacidade de armazenamento.

Utilização de formatos abertos e *software* de código aberto

Para que a longevidade dos documentos digitais seja assegurada é fulcral a utilização de *software* de código aberto, ou seja, *software* cujas especificações estão divulgadas de forma pública, passíveis de serem modificados e livres do pagamento de licenças de utilização. É, assim, aconselhável a utilização de *software* de código aberto em detrimento de *software* proprietário, sempre que seja possível.

Por outro lado, em relação aos formatos, também se deve dar preferência aqueles que são amplamente utilizados, baseados em normas estáveis e cujas especificações estejam divulgadas publicamente. A escolha de formatos deve ser feita tendo em consideração a sua capacidade de suportar as características inerentes aos objetos digitais. Para além disso, e sempre que possível, devem ser selecionados formatos baseados em XML (uma vez que, ao usarem uma estrutura bem definida e para a qual existem já várias bibliotecas computacionais disponíveis, garantem que seja possível a implementação de serviços de interoperabilidade entre plataformas) e que suportem metainformação.

No caso de documentos de texto aconselha-se o uso dos formatos PDF, ODT, XML e TXT. Note-se que o formato PDF não deve ser usado se o documento a

preservar puder, no futuro, vir a ter de ser alterado. Além disso, convém também referir que o formato ODT não deixa de ser uma aplicação do formato XML.

Para as folhas de cálculo recomendam-se os formatos PDF, ODS (também baseado em XML), XML e CSV. Para as imagens, os formatos JPEG e TIFF. Já os *emails* devem ser armazenados em HTML, TXT e CSV.

Não obstante, aconselha-se o controlo dos formatos no momento de criação dos objetos digitais de modo a evitar conversões desnecessárias.

Utilização de *software* de gestão documental

De modo a que os documentos provenientes do interior e exterior de uma organização possam ser geridos de forma eficaz ao longo do seu ciclo de vida é essencial a utilização de *software* de gestão documental. Muitas das aplicações de gestão documental já permitem a integração com os sistemas de gestão empresarial, possibilitando a gestão dos documentos desde o momento da sua criação. É imperativo que a informação não se perca, possa ser recuperada facilmente e siga os fluxos adequados. Um sistema de gestão documental permite a pesquisa quer no conteúdo, quer na metainformação do documento.

Este tipo de aplicação permite digitalizar, catalogar, classificar e administrar o fluxo de circulação dos objetos digitais na organização. Assim é possível reduzir o uso de papel (assumindo uma preocupação ambiental), libertar espaço do arquivo físico e aumentar a produtividade (dado que os documentos são facilmente identificados e geridos).

É, ainda, possível realizar o controlo de versões dos documentos, quando existem diversas colaborações sobre os mesmos. Não obstante, é também importante garantir que um documento digital se mantenha autêntico e íntegro, e isso pode ser feito recorrendo a assinaturas digitais e temporais.

É fulcral que todas as ações levadas a cabo sob os documentos – quer no seu conteúdo, como no objeto digital propriamente dito – estejam devidamente documentadas, para que no futuro se entendam quais as alterações efetuadas.

Geralmente este tipo de aplicação permite que se aceda aos documentos em qualquer lugar, bastando apenas ter uma ligação à Internet.

Cuidados com a tecnologia

O *hardware* responsável por garantir o acesso e armazenamento de objetos digitais deve ser manuseado de acordo com um conjunto de boas práticas, nomeadamente a instalação dos dispositivos em compartimentos com mecanismos de segurança, de modo a que apenas pessoas autorizadas tenham acesso. Não obstante, devem-se controlar os níveis de temperatura e humidade a que os dispositivos estão expostos, mas, também, devem estar situados fora da zona de exposição solar direta.

Para além deste conjunto de boas práticas é conveniente realizar inspeções frequentes ao *hardware* de modo a verificar se ainda tem condições para armazenar a informação digital da organização ou se se está a tornar obsoleto. Antes que o *hardware* se torne ultrapassado devem-se transferir os dados existentes para novos dispositivos que devem ser atuais e de boa qualidade. Devem ser seguidas as recomendações propostas pelos fabricantes de modo a garantir a longevidade e bom funcionamento dos equipamentos.

Cópias de segurança

A informação digital da empresa não deve estar exclusivamente armazenada nos dispositivos em funcionamento ou naqueles em que se pretende realizar o arquivo dos objetos digitais. Deve-se considerar a realização de cópias de segurança periódicas de modo a garantir a recuperação dos dados, em caso de perda acidental (falha tecnológica ou desastre ambiental) ou propositada (ação humana).

As cópias de segurança dos ficheiros devem ser distribuídas por diversos locais, geograficamente afastados das instalações onde se encontra o armazenamento primário. A escolha deve recair em locais estratégicos para que se evite que, em caso de desastre, também estes sejam afetados. Deve ser escolhido *hardware* de boa qualidade e assegurar o seu correto manuseamento.

Por outro lado, pode-se também optar por armazenar as cópias de segurança *online*, devendo-se seguir as recomendações anteriormente sugeridas.

Metainformação

A metainformação assume um papel preponderante na sustentabilidade dos documentos digitais ao longo do tempo. Ela auxilia no processo de identificação e recuperação de documentos, ao permitir definir elementos como título, autor ou assunto. Não obstante, também auxilia no processo de preservação digital, uma vez que permite descrever o ambiente tecnológico existente no momento de criação do objeto digital, as suas permissões de acesso e todas as ações efetuadas.

Desta forma é possível obter um histórico do documento desde o momento da sua criação até à altura em que necessita de ser alvo de ações de preservação de modo à manutenção do seu acesso.

É importante que a organização escolha o esquema de metainformação mais apropriado e formalize a nomenclatura a utilizar (linguagem controlada). Assim, é possível que os colaboradores definam os seus documentos e as ações realizadas de forma idêntica, auxiliando no processo de recuperação.

Plano de preservação digital

Para a implementação de um plano de preservação digital aconselha-se a constituição de uma equipa multidisciplinar composta por colaboradores de diversas áreas da organização. Assim, é possível auscultar as necessidades informacionais e os diferentes pontos de vista de cada departamento. No entanto, para que uma ação deste tipo seja bem-sucedida chama-se particular atenção para os profissionais das áreas da gestão documental e da informática. Enquanto os primeiros possuem conhecimentos acerca da classificação e seleção de documentos, os segundos conhecem as plataformas digitais. A ação conjunta destes profissionais permitirá desenvolver e implementar soluções eletrónicas que garantam a longevidade dos documentos digitais de forma íntegra e autêntica.

É fulcral que o planeamento da preservação digital esteja integrado na gestão estratégica da organização, dado que são necessários recursos financeiros, materiais e humanos que devem ser contemplados no seu orçamento.

Formação dos colaboradores

Um dos pontos mais importantes em todo o processo de preservação digital é a formação dos colaboradores. É conveniente que todos estejam informados das modificações ou novas implementações que irão ocorrer na organização, de modo a que possam implementar no seu quotidiano essas novas práticas. Por exemplo, caso uma empresa opte por estipular regras de nomeação de ficheiros, todos os funcionários devem ser informados das novas regras para que possam agir em conformidade.

Contudo, os gestores devem estar cientes que, muitas vezes, os funcionários resistem às mudanças, sendo necessário investir em estratégias mais eficazes que demonstrem a importância que novas formas de atuar trarão para o futuro da organização.