



Departamento de Ciências e Tecnologias de Informação

## Gestão de Documentos de Arquivo com uso de Sistemas em Software de Código Aberto

Sérgio Alexandre Ferreira

Dissertação submetida como requisito parcial para obtenção do grau de  
Mestre em Software de Código Aberto

Orientador(a):

Professor Doutor Carlos Costa, Professor Associado no ISEG (Lisbon School of Economics)

Co-orientadora

Professora Doutora Manuela Aparicio, Professora Auxiliar no Instituto Universitário de  
Lisboa (ISCTE-IUL)

Julho, 2019

# Agradecimentos

Agradeço aos meus professores do Mestrado, todo o empenho que colocaram para me permitir adquirir mais conhecimentos e competências.

À minha família, por estar sempre presente: Ao Pai uma palavra de especial saudade; À Mãe por estar sempre presente; Ao Filho e Filha por existirem e acreditarem sempre em mim. Aos irmãos pelo companheirismo na procura de direções.

À Eva pela paciência por me aturar durante esta aventura.

# Resumo

A Gestão de Documentos de Arquivo é um processo essencial para as Organizações Públicas ou Privadas e para indivíduos focando-se no registo, descrição, acesso e preservação dos documentos associados às suas atividades. Cada vez mais organizações e indivíduos têm de gerir enormes massas documentais não se referindo estes documentos apenas a textos escritos, mas também arquivos fotográficos, sonoros, entre outros. Entretanto o crescimento da massa documental digital, com as evoluções tecnológicas dos últimos anos, tem vindo a acelerar, contrariamente ao suporte físico, que cada vez mais é digitalizado e destruído ou simplesmente substituído pelo formato digital. A evolução dos Sistemas de Informação permitiu avanços na Gestão de Documentos de Arquivo, pois facilitou a sua operacionalização, contribuindo para maturar práticas e conceitos. Em especial, o movimento do Software Livre trouxe novos sistemas nas mais diversas áreas, incluindo as da Gestão de Documentos de Arquivo. A implementação mais alargada destes sistemas trouxe a descoberto necessidades que podem ser colmatadas com implementações de arquiteturas mais modernas, garantindo maior elasticidade e disponibilidade. Ao verificar a existência limitada de investigações sobre este tema, esta dissertação procura, em termos gerais, identificar se é possível implementar um Sistema de Gestão de Documentos de Arquivo unicamente com software livre. No âmbito da investigação, foi necessário identificar os principais requisitos alinhados com normas e standards relevantes, criar um modelo conceptual para selecionar implementações e, criar um protótipo que demonstre ser possível resolver o problema definido. A investigação foi conduzida socorrendo-se da metodologia DSRM, ao realizar cada uma das atividades definidas, aplicando-as ao problema e objetivos da dissertação. Finalmente, a investigação demonstra, através da concretização do modelo conceptual, desenhado a partir das técnicas investigadas e descritas na revisão de literatura, que é possível implementar um Sistema de Gestão de Documentos de Arquivo, com Software Livre ou de Código Aberto.

## Palavras-chave

Gestão Documental, Gestão de Registos, Arquivo, Preservação Digital, Software Livre, Software de Código Aberto



# Abstract

Document Management is an essential process for Public or Private Organizations and also for individuals, that focuses in recording, describing, accessing and preserving documents linked to their activities. In an increasing matter, individuals and organizations need to manage huge amounts of documents, these documents not being solely text-based but also photographic archives, sound archives, among others. Meanwhile the growth of digital assets, with all the technological evolution from the past years, has been accelerating, and the physical assets have been decreasing in proportion, as they are mostly scanned and destroyed or simply replaced by a digital format. The evolution of Information Systems allowed breakthroughs in Document Management as it has facilitated its operationalization, contributing to practices and concept maturity. In particular, the Free Software Movement brought a wave of new systems in the most diverse fields, including Document Management. The increase in implementations in the field uncovered new requirements that can be covered with the use of modern architectures, that warrant more elasticity and availability. Observing the limited amount of investigations on this theme, this dissertation aims, in general, to identify if it is possible to implement a Document Management System only with Open Source Software. Within the scope of the research, it was necessary to identify the main requirements aligned with relevant norms and standards, create a conceptual model to select implementations and, create a prototype that demonstrates that it is possible to solve the defined problem. The research was conducted using the DSRM methodology, by performing each activity defined, applying them to the problem and goals of the dissertation. Finally, the research demonstrates, through the implementation of the conceptual model, designed based on the investigated techniques and described in the literature review, that it is possible to implement a Document Management System, with Free or Open Source Software.

## Keywords

Document Management, Records Management, Archive, Digital Preservation, Free and Open Source Software

# Table of Contents

Introdução.....	1
1.1 Definição do problema.....	2
1.2 Questão de Investigação.....	4
1.3 Objetivos.....	4
1.4 Abordagem Metodológica.....	5
1.5 Resultados.....	6
1.6 Estrutura da Dissertação.....	7
1.7 Contribuições e Publicações.....	7
2 Revisão de Literatura.....	8
2.1 Princípios e Conceitos Arquivísticos.....	9
2.2 Normativos.....	17
2.2.1 Principais Esforços de Normalização.....	18
2.2.2 Entidades Normalizadoras.....	20
2.2.3 Normativos que definem formatos.....	21
2.2.4 Gestão Documental.....	21
2.2.4.1 ISO15489.....	22
2.2.4.2 MoReQ 2010.....	22
2.2.4.3 DOD 5015.2.....	25
2.2.4.4 MEF.....	26
2.2.4.5 MIP.....	27
2.2.5 Descrição Arquivística.....	27
2.2.5.1 ISAD(G) / ISAAR(CPF).....	28
2.2.5.2 EAC – Encoded Archival Context.....	30
2.2.5.3 CIDOC Conceptual Reference Model.....	30
2.2.6 Preservação Digital.....	31
2.2.6.1 Modelo de Referência OAIS.....	32
2.2.6.2 METS.....	34
2.2.6.3 Projeto E-Ark.....	35
2.2.6.4 PREMIS.....	37
2.3 Tecnologia.....	38
2.3.1 Arquiteturas de Sistemas de Informação.....	39
2.3.1.1 Escalabilidade.....	39
2.3.1.2 Sistemas baseados em Micro Serviços.....	41
2.3.1.3 Integração.....	42
2.3.2 Tecnologia de Base e <i>Commodities</i> .....	43
2.3.3 Software aplicado à Gestão de Documentos e de Arquivos.....	44
2.3.3.1 Sistema de Gestão Documental.....	46
2.3.3.2 Descrição Arquivística.....	47
2.3.3.3 Preservação Digital.....	48
2.3.4 Pesquisa.....	49
2.3.4.1 Pesquisa Federada vs Pesquisa Centralizada.....	50
2.3.4.2 OAI: PMH.....	50
2.3.5 Optical Character Recognition - OCR.....	51
2.3.6 Transformação.....	51
2.4 Software Livre e de Código Aberto.....	52
2.4.1 Metodologia de Avaliação de Software Open Source.....	55
2.4.2 Economia de Serviços e Dependência de Fabricantes.....	58
2.5 Metodologia de Descrição de Arquitetura.....	58

2.6 Síntese.....	63
3 Proposta de Solução Conceptual.....	64
3.1 Arquitetura de referência.....	64
3.2 Decisões Arquiteturais.....	65
3.3 Componentes Arquiteturais.....	66
3.4 Componentes para a Gestão de Documentos de Arquivo.....	67
3.4.1 Gestão Documental.....	67
3.4.2 Descrição Arquivística.....	68
3.4.3 Preservação Digital.....	68
3.4.4 Assinatura Digital.....	69
3.4.5 OCR.....	69
3.4.6 Deteção de Layouts.....	69
3.4.7 Indexação e Pesquisa.....	70
3.4.8 Transformação de Formatos.....	70
3.5 Componentes de infraestrutura.....	70
3.5.1 Sistema(s) Operativo(s).....	71
3.5.2 Virtualização.....	71
3.5.3 Orquestração de Contentores.....	71
3.5.4 Autenticação e Diretório de Utilizadores.....	72
3.5.5 Sistemas de Gestão de Bases de Dados Relacionais.....	72
3.5.6 Filas de Mensagens Assíncronas.....	72
3.5.7 Micro-serviços.....	73
3.5.8 Indexação.....	73
3.6 Visão Global do Modelo Conceptual de Referência.....	74
4 Trabalho Empírico.....	76
4.1 Protótipo.....	77
4.2 Seleção das Componentes.....	78
4.2.1 Gestão Documental.....	79
4.2.1.1 Modelo Utilizado para Análise.....	79
4.2.1.2 Maturidade.....	79
4.2.1.3 Arquitetura.....	80
4.2.1.4 Serviços de Sistema.....	82
4.2.1.5 Funcionalidades.....	83
4.2.1.6 Soluções Identificadas.....	83
4.2.1.7 Alfresco Community.....	84
4.2.1.8 LogicalDoc.....	87
4.2.1.9 Nuxeo.....	90
4.2.1.10 OpenKM.....	93
4.2.1.11 Síntese.....	96
4.2.2 Descrição Arquivística.....	99
4.2.2.1 Modelo Utilizado para Análise.....	99
4.2.2.2 Maturidade.....	100
4.2.2.3 Arquitetura.....	101
4.2.2.4 Funcionalidades.....	101
4.2.2.5 Standards.....	102
4.2.2.6 Soluções Identificadas.....	102
4.2.2.7 Universo de Soluções.....	102
4.2.2.8 AtoM - Access To Memory 2.4.0.....	103
4.2.2.9 Collective Access.....	105
4.2.2.10 Síntese.....	107

4.2.2.11 Conclusão.....	109
4.2.3 Preservação Digital.....	109
4.2.3.1 Modelo Utilizado para Análise.....	109
4.2.3.2 Maturidade.....	110
4.2.3.3 Arquitetura.....	110
4.2.3.4 Funcionalidades.....	111
4.2.3.5 Standards.....	111
4.2.3.6 Soluções Identificadas.....	111
4.2.3.7 Archivemática 1.7.1.....	112
4.2.3.8 RODA 2.2.9.....	113
4.2.3.9 Xena 6.1.0.....	115
4.2.3.10 Síntese.....	116
4.2.4 Componentes para a Gestão de Documentos de Arquivo.....	118
4.2.4.1 Gestão Documental.....	119
4.2.4.2 Descrição Arquivística.....	121
4.2.4.3 Preservação Digital.....	122
4.2.4.4 Assinatura Digital.....	122
4.2.4.5 OCR.....	123
4.2.4.6 Detecção de Layouts.....	123
4.2.4.7 Indexação e Pesquisa.....	123
4.2.4.8 Transformação de Formatos.....	124
4.2.5 Componentes de Infraestrutura.....	124
4.2.5.1 Sistema(s) Operativo(s).....	125
4.2.5.2 Virtualização.....	126
4.2.5.3 Orquestração de Contentores.....	126
4.2.5.4 Autenticação e Diretório de Utilizadores.....	127
4.2.5.5 Sistemas de Gestão de Bases de Dados Relacionais.....	127
4.2.5.6 Filas de Mensagens Assíncronas.....	127
4.2.5.7 Micro-Serviços.....	127
4.2.5.8 Indexação.....	128
4.3 Avaliação Preliminar.....	128
4.4 Aproximação aos normativos.....	129
4.4.1 Alfresco vs MoReq2010.....	129
4.4.2 Alfresco vs ISO 15489.....	130
4.4.3 Alfresco vs MEF.....	130
4.4.4 Alfresco vs MIP.....	130
4.4.5 AtoM versus ISAD(G).....	131
4.4.6 ArchiveMática vs Modelo de Referência OAIS.....	131
4.5 Arquitetura Concreta.....	131
4.5.1 Resumo de Produtos.....	133
4.5.2 Gestão de Documentos de Arquivo.....	134
4.5.2.1 Alfresco: Gestão Documental.....	134
4.5.2.2 AtoM: Descrição Arquivística.....	135
4.5.2.3 Archivemática: Preservação Digital.....	137
4.5.2.4 Esign-Cert: Assinatura Digital.....	137
4.5.2.5 OpenOCR: Reconhecimento Ótico de Caracteres.....	138
4.5.2.6 OpenCV: Detecção de Layouts.....	139
4.5.2.7 ElasticSearch: Indexação e Pesquisa.....	139
4.5.3 Componentes de Infraestrutura.....	140
4.5.3.1 Linux: Sistema Operativo.....	140
4.5.3.2 KVM & Docker: Virtualização.....	140

4.5.3.3 Kubernetes: Orquestração de Contentores.....	141
4.5.3.4 OpenLDAP: Autenticação e Diretório de Utilizadores.....	141
4.5.3.5 PostgreSQL & MariaDB: Sistema de Gestão de Bases de Dados Relacionais.....	141
4.5.3.6 ActiveMQ: Filas de Mensagens Assíncronas.....	142
4.5.3.7 Spring Boot & Micro-serviços.....	142
4.6 Instalação e Configuração do Sistema.....	142
5 Conclusões e trabalhos futuros.....	145
5.1 Principais Conclusões.....	145
5.2 Trabalho Futuro.....	146
6 Referências Bibliográficas.....	148
7 Appendix A – Requisitos Alfresco vs MoReq2010.....	167
7.1.1.1 Serviços do Sistema.....	167

# Índice de Figuras

Figura 1: Arquivismo no Tempo.....	10
Figura 2: Histórico de Publicação de Normas relacionadas com Gestão de Documentos.....	12
Figura 3: Serviços Especificados no MoReq2010 (DLM Forum, 2010).....	24
Figura 4: Modelo de Referência "Open Archival Information System (OAIS)(«Archivematica», 2019)" .....	33
Figura 5: Metodologia QSOS (Moreau, 2013).....	57
Figura 6: Contexto de Descrições de Arquitetura.....	59
Figura 7: Representação das Dimensões do Archimate (The Open Group, 2017).....	62
Figura 8: Arquitetura Conceptual.....	65
Figura 9: Visão Global do Modelo Conceptual.....	75
Figura 10: Critérios de Avaliação da Maturidade de Software de Gestão Documental para aplicação da metodologia QSOS	80
Figura 11: Critérios de Avaliação da Arquitetura de Software de Gestão Documental para aplicação da metodologia QSOS..	81
Figura 12: Critérios de Avaliação da Serviços Disponíveis em Software de Gestão Documental para aplicação da metodologia QSOS.....	82
Figura 13: Critérios de Avaliação das Funcionalidades de Software de Gestão Documental para aplicação da metodologia QSOS.....	83
Figura 14: Análise QSOS - Maturidade do Alfresco Community.....	85
Figura 15: Arquitetura do Alfresco Community.....	86
Figura 16: Análise QSOS - Serviços de Sistema do Alfresco Community.....	87
Figura 17: Análise QSOS - Funcionalidades do Alfresco Community.....	87
Figura 18: Análise QSOS - Maturidade do Logical Doc.....	88
Figura 19: Análise QSOS - Maturidade do LogicalDoc.....	89
Figura 20: Análise QSOS - Serviços de Sistema do LogicalDoc.....	90
Figura 21: Análise QSOS - Funcionalidades do LogicalDoc.....	90
Figura 22: Análise QSOS - Maturidade do Nuxeo.....	91
Figura 23: Análise QSOS - Maturidade do Nuxeo.....	92
Figura 24: Análise QSOS - Serviços de Sistema do Nuxeo.....	93
Figura 25: Análise QSOS - Funcionalidades do Nuxeo.....	93
Figura 26: Análise QSOS - Maturidade do OpenKM.....	94
Figura 27: Análise QSOS - Maturidade do OpenKM.....	95
Figura 28: Análise QSOS - Serviços de Sistema do OpenKM.....	96
Figura 29: Análise QSOS - Funcionalidades do OpenKM.....	96
Figura 30: Análise QSOS - Comparação de Maturidades de Sistemas de Gestão Documental.....	97
Figura 31: Análise QSOS - Comparação de Arquiteturas de Sistemas de Gestão Documental.....	97
Figura 32: Análise QSOS - Comparação de Serviços dos Sistemas de Gestão Documental.....	98
Figura 33: Análise QSOS - Comparação de Funcionalidades de Sistemas de Gestão Documental.....	98
Figura 34: Análise QSOS - Conclusão das comparações de Sistemas de Gestão Documental.....	99
Figura 35: Critérios de Avaliação da Maturidade de Software de Descrição Arquivística para aplicação da metodologia QSOS.....	100
Figura 36: Critérios de Avaliação da Arquitetura de Software de Descrição Arquivística para aplicação da metodologia QSOS .....	101

Figura 37: Critérios de Avaliação das Funcionalidades de Software de Descrição Arquivística para aplicação da metodologia QSOS.....	101
Figura 38: Standards Implementados para análise QSOS.....	102
Figura 39: Análise QSOS - Maturidade do AtoM.....	103
Figura 40: Análise QSOS - Maturidade do AtoM.....	104
Figura 41: Análise QSOS - Funcionalidades do AtoM.....	104
Figura 42: Análise QSOS - Standards implementados pelo AtoM.....	104
Figura 43: Análise QSOS - Maturidade do Collective Access.....	105
Figura 44: Análise QSOS - Maturidade do Collective Access.....	106
Figura 45: Análise QSOS - Funcionalidades do Collective Access.....	106
Figura 46: Análise QSOS - Standards implementados pelo Collective Access.....	106
Figura 47: Análise QSOS - Comparação de Maturidades de Sistemas de Descrição Arquivística.....	107
Figura 48: Análise QSOS - Comparação de Arquiteturas de Sistemas de Descrição Arquivística.....	107
Figura 49: Análise QSOS - Comparação de Funcionalidades de Sistemas de Descrição Arquivística.....	108
Figura 50: Análise QSOS - Comparação de aderência aos standards de Sistemas de Descrição Arquivística.....	108
Figura 51: Análise QSOS - Conclusão das comparações de Sistemas de Descrição Arquivística.....	109
Figura 52: Critérios de Avaliação da Maturidade de Software de Gestão Documental para aplicação da metodologia QSOS.....	110
Figura 53: Critérios de Avaliação da Arquitetura de Software de Preservação Digital para aplicação da metodologia QSOS.....	110
Figura 54: Critérios de Avaliação das Funcionalidades de Software de Preservação Digital para aplicação da metodologia QSOS.....	111
Figura 55: Standards Implementados para análise QSOS.....	111
Figura 56: Análise QSOS - Maturidade do Archivematica.....	112
Figura 57: Análise QSOS - Maturidade do Archivematica.....	113
Figura 58: Análise QSOS - Funcionalidades do Archivematica.....	113
Figura 59: Análise QSOS - Standards Implementados pelo Archivematica.....	113
Figura 60: Análise QSOS - Maturidade do RODA.....	114
Figura 61: Análise QSOS - Maturidade do RODA.....	114
Figura 62: Análise QSOS - Funcionalidades do RODA.....	114
Figura 63: Análise QSOS - Standards Implementados pelo RODA.....	115
Figura 64: Análise QSOS - Maturidade do Xena.....	115
Figura 65: Análise QSOS - Arquitetura do Xena.....	116
Figura 66: Análise QSOS - Funcionalidades do Xena.....	116
Figura 67: Análise QSOS - Standards Implementados pelo Xena.....	116
Figura 68: Análise QSOS - Comparação de Maturidades de Sistemas de Preservação Digital.....	117
Figura 69: Análise QSOS - Comparação de Arquiteturas de Sistemas de Preservação Digital.....	117
Figura 70: Análise QSOS - Comparação de Funcionalidades de Sistemas de Preservação Digital.....	117
Figura 71: Análise QSOS - Comparação de aderência aos standards de Sistemas de Preservação Digital.....	118
Figura 72: Análise QSOS - Conclusão das comparações de Sistemas de Preservação Digital.....	118
Figura 73: Arquitetura Alfresco.....	120
Figura 74: Arquitetura Concreta da Solução.....	133
Figura 75: Arquitetura do AtoM.....	136
Figura 76: Arquitetura do Sistema Archivematica.....	137
Figura 77: Arquitetura do OpenOCR.....	138

# Índice de Tabelas

Principais Normativos.....	20
Entidades com influência na criação de normas.....	22
Formatos eletrônicos para Gestão de Documentos de Arquivo.....	23
Resumo Descritivo de Enquadramento da ISO15489.....	24
Resumo Descritivo do MoReq2010.....	26
Resumo Descritivo da DOD 5015.2.....	28
Resumo Descritivo da MEF.....	29
Resumo Descritivo da MIP.....	30
Resumo Descritivo da definição EAC.....	31
Resumo Descritivo da definição EAC.....	33
Resumo Descritivo do CIDOC CRM.....	34
Resumo Descritivo do Modelo de Referência OAIS.....	37
Resumo Descritivo do esquema de meta-dados METS.....	38
Resumo Descritivo de Enquadramento do Projeto E-Ark.....	40
Resumo Descritivo do Relatório PREMIS.....	42
Metodologias de Avaliação de Software Open Source.....	60
Linguagens de Descrição de Arquitetura (Hussain, 2013).....	65
Taxonomia de Componentes.....	72
Componentes Aplicacionais para a Gestão de Documentos de Arquivo.....	126
Componentes de Infraestrutura.....	132
Resumo de Principais Avaliações de Software de Gestão Documental.....	136
Resumo de Principais Avaliações de Software de Descrição Arquivística.....	136
Resumo de Principais Avaliações de Software de Preservação Digital.....	136
Produtos de Software Selecionados para Gestão de Documentos de Arquivo.....	142
Produtos selecionados como Software de Infraestrutura.....	142
Interfaces Disponibilizados pelo Alfresco.....	144
Módulos instalados em Contentores.....	153
Módulos não instalados em Contentores.....	153
Avaliação de Requisitos de Serviços de Sistema em relação ao MoReq.....	179
Avaliação de Requisitos de Serviços de Pesquisa em relação ao MoReq.....	184
Avaliação de Requisitos de Serviços de Classificação em relação ao MoReq.....	185
Avaliação de Requisitos de Serviços de Registo de Documentos em relação ao MoReq.....	186
Avaliação de Requisitos de Serviços de Meta-dados em relação ao MoReq.....	190
Avaliação de Requisitos de Serviços de Retenção em relação ao MoReq.....	199
Avaliação de Requisitos de Serviços de Pesquisa em relação ao MoReq.....	200
Avaliação de Requisitos de Serviços de Exportação em relação ao MoReq.....	203

# Lista de Abreviaturas e Acrónimos

ACL	Access Control Lists
ADL	Architecture Description Language
AIP	Archival Information Package
API	Application Programming Interface
BAD	Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas
CCSDS	Consultative Committee for Space Data Systems
CIDOC	International Committee for Documentation on ICOM
COLD	Computer Output to Laser Disk (o mesmo que ERM)
COLD	Computer to Laser Disk
CPF	Corporate Bodies, Persons and Families (on ISAAR standard)
CRM	Customer Relationship Management
DIP	Dissemination Information Package
DLM	Document Lifecycle Management
DOD	United States Department of Defense
DSRM	Design Science Research for Information Systems
E-Ark	European Archival Records and Knowledge
EAC	Encoded Archival Context
EAD	Encoded Archival Description
ECM	Enterprise Content Management
EDM	Electronic Document Management
EDRMS	Electronic Document and Records Management System
ERM	Enterprise Report Management
ERP	Enterprise Resource Planning
FLOSS	Free/Libre Open Source Software
FSF	Free Software Foundation
GNU	GNU is Not Unix
GPL	GNU General Public License
GWT	Google Web Toolkit
HTTP	HyperText Transfer Protocol
ICA	International Council Organization
ICOM	International Council of Museums
IOT	Internet Of Things
IP	Internet Protocol
ISAAR	International Standard Archival Authority Record
ISAD(G)	General International Standard Archival description
ISDIAH	International Standard for Describing Institutions with Archival Holdings
ISO	International Standards Organization
LOC	United States Library Of Congress
MCRS	MoReq Compliant Record System
MEF	Macro Estrutura Funcional
METS	Metadata Encoding and Transmission Standard
MIP	Meta-informação para a Interoperabilidade

MOM	Message Oriented Middleware
MoReq	Modelo de Requisitos para a gestão de arquivos electrónicos
OAI	Open Archives Initiative
OAI:PMH	Open Archives Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OCR	Optical Character Recognition
ODLIS	Online Dictionary for Library and Information Science
PDF	Portable Document Format
PMH	Protocol for Metadata Harvesting
PREMIS	Preservation Metadata Maintenance Activity
PWG	Premis Working Group
QSOS	Qualification and Selection of Open Source software
RGPD	Regulamento Geral de Proteção de Dados
RLG	Research Libraries Group
RM	Records Managements
SAA	Society of American Archivists
SIEM	Security Information and Event Management
SIP	Submission Information Package
SOA	Service Oriented Architectures
UML	Unified Modeling Language
WWW	World Wide Web
XML	Extended Markup Language

# Introdução

A temática da Gestão de Documentos de Arquivo é cada vez mais relevante, com um volume negócios que este ano atingirá o valor de 37 biliões de dólares («Global enterprise content management market size 2017-2022», 2018), prevendo-se um crescimento para 67 biliões de dólares em 2022.

A difusão de projetos informáticos que são disponibilizados com licenças Livres e(ou) de Código Aberto, é uma realidade incontornável. Estima-se que só nos repositórios do GitHub estejam disponíveis mais de 96 milhões projetos («GitHub Octoverse 2018 | Highlights from the last twelve months», 2018). Estima-se que mais de 55 % da infraestrutura das organizações a nível mundial se suporte em Software Livre ou de Código Aberto (2016 *Future of Open Source Survey Results*, 2016) e que mais de 65% das organizações suportem o seu desenvolvimento de software em produtos de software desenvolvidos com licenças de Software Livre ou de Código Aberto.

Após efetuar uma primeira análise, é fácil verificar que o Software Livre ou o Software de Código Aberto é dominante na área mais genérica da Gestão de Conteúdos. Aparentemente, após uma primeira investigação, tal não ocorre com sistemas de Gestão de Documentos de Arquivo.

A investigação efetuada permitiu identificar os principais trabalhos que se focaram total ou parcialmente neste campo, da utilização de software livre como suporte para um sistema que implemente os principais normativos da área de gestão de Documentos de Arquivo, mais concretamente: *Perspectives on Digital Preservation in Open-Source Digital Library Software* (Madalli, Barve, & Amin, 2012), *Enterprise content management (ECM) implementation in South Africa* (Katuu, 2012), *Software livre: um diferencial competitivo para a Arquivística* (Saraiva & Nogueira, 2015), *XML e preservação digital* (Ramalho et al., 2007). O estudo destas publicações permitiram compreender a necessidade de investigação adicional sobre estes temas, bem como a atualidade da definição de uma solução de sistema para Gestão de Documentos de Arquivo suportada em Software Livre ou de Código Aberto.

Neste contexto, afigura-se como relevante desenvolver investigação que: Permita compreender o estado da arte dos sistemas de Gestão de Documentos de Arquivo desenvolvidos com licenças de Software Livre ou de Código Aberto, dos normativos mais relevantes e até que ponto, estes sistemas implementam os requisitos por eles definidos;

Proponha uma metodologia que ao ser aplicada permita implementar um sistema de informação cumpridor dos principais requisitos definidos pelos normativos usando conjuntos de Software Livre ou de Código Aberto; Implemente um protótipo com produtos concretos que implementem a solução conceptual proposta.

## 1.1 Definição do problema

A quantidade de informação armazenada crescerá nos próximos anos a uma taxa superior a 25% ao ano – prevendo-se que atinja 40 Exabytes em 2020 – sendo a vasta maioria – cerca de 80% - composta por informação não estruturada, maioritariamente não classificada, identificando-se que mais de 23% desta ganharia valor se devidamente tratada (Gantz & Reinsel, 2012). Seguindo a regra de que 80% da informação relevante para as organizações tem origem em informação não estruturada (Grimes, 2008), pode-se facilmente compreender que existe um grande valor inexplorado nesta enorme massa de informação não estruturada.

Paralelamente tem-se vindo a assistir a mudanças na economia, nomeadamente a passagem de uma economia de produto para uma economia de serviços (Lusch, Vargo, & O'Brien, 2007), alterações estas ainda mais acentuadas nos sistemas de informação, em grande parte liderado por fenómenos como o movimento do Software Livre e de Código Aberto (Fitzgerald, 2006). Sendo assunto principal e denominação deste mestrado (Mestrado em Software de Código Aberto), estas questões levaram-me a ponderar se é possível responder a todos os problemas colocados pelas necessidades de Gestão dos Documentos de Arquivo, nomeadamente respeitando os normativos que a regulamentam, com Software *disponível* com Licença de Livre ou de Código Aberto.

As primeiras pesquisas indicam que não existe um Programa, nem mesmo um Conjunto de Programas que resolvam todos os problemas nas organizações, assistindo-se hoje, nas organizações, a uma fragmentação dos seus programas informáticos (Umar & Zordan, 2009). Se antes era comum a existência de um único pacote de *software* que se propunha intervir em todas as áreas da instituição, recorre-se hoje à utilização de programas de elevada qualidade, dedicados a cada uma das diferentes atividades das instituições, a maior parte das vezes implementados por diferentes equipas, altamente especializadas na resolução de um conjunto limitado de problemas. As mais recentes tendências das arquiteturas de computação, de onde emergem conceitos como a escalabilidade, a elasticidade, a sincronização (da Rosa Righi et al., 2018), suportados pela generalização da integração, sobretudo com micro-serviços, seja

através de Interfaces de Programação de Aplicações (API's), seja com Filas de Mensagens Avançadas (Singh, Narendra, Singh, & Shailendra, 2017), serão provavelmente a base que permitirá às instituições retirar o melhor de cada programa especializado, utilizando-os no entanto como se de um todo se tratasse. O problema que esta dissertação estuda não é diferente da realidade atrás descrita: Existe uma proliferação de normativos e de programas, cada um resolvendo um conjunto limitado de problemas. Estudos anteriores (Vieira & Borbinha, 2011) focam-se no levantamento dos normativos existentes escolhendo uma implementação existente como exemplo para estudo (António, 2008).

No sentido de avançar para uma formulação mais completa do problema da implementação dos normativos relativos à Gestão de Documentos de Arquivo com Software Livre ou Open Source, decidiu-se aprofundar as questões relacionadas com as implementações e as possibilidades de estas fornecerem respostas completas aos principais normativos.

Nunca como atualmente se coloca tanto a questão da necessidade de gerir informação não estruturada. O espólio documental anterior relevante - definido como para arquivo - continua a existir, felizmente cada vez mais organizado descrito com meios eletrónicos (Taylor & Joudrey, 2017). Os trabalhos das últimas décadas, que teorizam tanto os ciclos de vida dos ativos de informação não estruturada como os formatos de transferência e preservação criaram uma base sólida para uma era de abertura e transparência destes ativos, seja facilitando o acesso a esta informação, como garantindo que a sua existência tangível ou de representações digitais serão preservadas para as futuras gerações (Ross, 2012).

Além da acumulação histórica de informação não estruturada, o facto de as organizações e pessoas terem acesso a dispositivos e programas que facilitam a criação de conteúdos, acelerou muito a sua produção. Estudos apontam para que a criação de informação nas últimas décadas tenha sido superior ao que a humanidade criou durante os 1000 anos anteriores (Gantz & Reinsel, 2012). É de notar que também os formatos se multiplicaram, existindo a informação não estruturada sob a forma de mensagens de correio eletrónico, imagens, documentos escritos com formatação, filmes, apresentações, registos sonoros e outros (Thibodeau, 2002). Estima-se que a percentagem de informação não estruturada seja mais de 80% da informação existente (Blumberg & Atre, 2003).

As quantidades crescentes de informação não estruturada criam novos problemas aos arquivistas, que, se não ocorrerem modificações tecnológicas, não conseguirão de forma eficaz, classificar, arquivar, descrever, disponibilizar a informação existente (Rowley & Hartley, 2017).

A presente investigação pretende responder à questão se é possível responder às necessidades colocadas pelos Ciclos de Vida dos Documentos e dos Arquivos, em grande parte regulamentadas pelos normativos, utilizando Software Livre ou de Código Aberto, mesmo que recorrendo a várias implementações, integrações e(ou) alteração das componentes existentes.

## 1.2 Questão de Investigação

Com base nos problemas identificados focou-se na seguinte frase a questão de investigação:

*Poder-se-á desenvolver um sistema que suporte o ciclo de vida dos Documentos de Arquivo baseado em tecnologias licenciadas como Software Livre ou Open Source?*

## 1.3 Objetivos

Face à questão de investigação, definiu-se como Objetivo Geral desta tese: Aferir a adequação das implementações de Programas da área da Gestão de Documentos de Arquivo face aos principais normativos por forma a implementar um Sistema em Tecnologias licenciadas como Software Livre ou de Código Aberto.

Para conseguir responder ao Objetivo geral atrás descrito, foram objetivos específicos da investigação:

1. Identificar principais necessidades de Gestão de Documentos de Arquivo;
2. Identificar Standards e Normas Relevantes;
3. Enumerar Requisitos face a necessidades, normas e standards;
4. Procurar implementações de *Software Open Source* para a área da gestão de Documentos de Arquivo;
5. Analisar os Requisitos estudados face às Implementações;
6. Criar modelo conceptual de solução, desenhando uma arquitetura de referência;
7. Desenhar e testar através de um protótipo uma proposta de solução, definido uma arquitetura concreta de aposição dos produtos estudados, ao modelo conceptual, caso tal seja possível.

## 1.4 Abordagem Metodológica

No desenvolvimento da investigação foi aplicada a metodologia DSRM – *Design Science Research for Information Systems* (Hevner, March, Park, & Ram, 2008). Assim, foi usada uma epistemologia positivista, baseada no método científico, fundamentando as aprendizagens na observação, descrição, explicação e previsão da realidade através da construção de modelos que permitem descrever a realidade e conceitos.

Em consonância com a metodologia, durante a investigação, ocorreram as seguintes atividades (Peffer, Tuunanen, Rothenberger, & Chatterjee, 2007):

1. Motivação e Identificação do Problema
2. Definição de Objetivos para a Solução
3. Desenho e Desenvolvimento do Modelo Conceptual
4. Demonstração através de um protótipo
5. Avaliação
6. Comunicação

Na primeira fase, a identificação do problema obrigou à investigação de trabalhos científicos existentes, tendo-se concluído que se trata de uma área que carece de investigações adicionais, motivando a execução da investigação descrita nesta dissertação.

Na segunda fase, com base na identificação do problema definiram-se objetivos gerais e específicos para a implementação da solução. A identificação do problema e os objetivos da solução desenhada estão descritos nos três primeiros subcapítulos desta dissertação. Em complemento foi investigado qual o modelo de seleção de software livre que melhor se adequa à avaliação das aplicações que serão escolhidas como componentes. Esta fase foi especialmente relevante, pois permitiu aferir o Estado da Arte da Gestão de Documentos de Arquivo, bem como suportar a utilização de técnicas e instrumentos durante a investigação, analisando documentos e estudos científicos que se debruçaram sobre esta área. Complementarmente foram analisados standards, normas e boas práticas associadas a cada uma das três subáreas identificadas como nucleares na área da gestão de Documentos de Arquivo: Gestão Documental, Descrição Arquivística e Preservação Digital.

Na terceira fase, com o problema identificado e objetivos claros, deu-se início à fase iterativa que permitiu desenvolver o modelo conceptual, que consiste numa arquitetura de referência, identificando que componentes devem constituir um Sistema de Gestão de Documentos de

Arquivo apenas com software livre. A criação do modelo conceptual, em particular, efetuada com o suporte de instrumentos específicos, revestiu-se de especial importância, na medida em que permitiu claramente descrever e compreender as funcionalidades que tal sistema deverá implementar, incluindo a definição de quais as componentes do sistema e suas interligações (Jarvinen, 2000).

Na quarta fase, aplicando o modelo conceptual desenvolvido e usando o modelo de seleção, foram criadas várias instâncias que visavam resolver o problema utilizando as várias peças de software pré-selecionadas aplicando sobre elas o modelo de seleção. Com base nos Blocos Arquitetados, foi possível descobrir que programas implementam as necessidades, quais as suas lacunas, permitindo desenhar uma proposta de Solução, incluindo a Arquitetura Concreta concretizando a Arquitetura Conceptual com Software Livre e de Código Aberto.

A quinta fase consistiu na avaliação de cada uma das instâncias, observando a sua adequação e eficiência na resolução do problema. Os resultados são apresentados nas sínteses do capítulo 4.2. Com base nos resultados, da avaliação, procedeu-se a modificações do modelo conceptual regressando à terceira fase ajustando as componentes que o compõem (Hevner, 2007). Na última iteração, conseguiu-se chegar a um protótipo que corresponde a uma solução do problema identificado. A utilização de técnicas como Linguagens de Descrição de Arquiteturas e Modelos de Avaliação de Software Livre ou de Código Aberto não só demonstrou, mas também permitiu avaliar e medir as escolhas efetuadas. As principais técnicas foram escolhidas com base em trabalhos científicos previamente investigados.

Na última fase, o texto da dissertação foi cuidadosamente escrito por forma a comunicar o problema, a sua importância, utilidade, o protótipo construído e o rigor usado no seu desenho. Com base neste texto serão escritos artigos e realizadas apresentações a diferentes audiências.

## **1.5 Resultados**

No final da investigação obteve-se um desenho de uma Arquitetura Conceptual de Referência, de uma Arquitetura Concreta, devidamente concretizada com a implementação de protótipos, de um sistema que permite gerir os Documentos de Arquivo usando exclusivamente Software Livre ou de Código Aberto.

Os resultados são apresentados com Diagramas representativos das arquiteturas, da descrição do software usado em cada bloco arquitetural, bem como os normativos em que cada qual se enquadra e(ou) cumpre.

Esperamos ter demonstrado, aplicando estas técnicas já devidamente estudadas e teorizadas, se é ou não possível implementar um sistema global altamente adaptável às mudanças normativas e à evolução dos pacotes envolvidos.

## **1.6 Estrutura da Dissertação**

A dissertação está estruturada de acordo com as fases adotadas na metodologia. Na primeira parte foi definido o problema, a questão de investigação e os objetivos (gerais e específicos). Na segunda parte, Revisão de Literatura, descreve-se o Estado da Arte da área da Gestão de Documentos de Arquivo. Na terceira, com base no investigado na revisão de Literatura, descreve-se a solução, nomeadamente o Modelo Conceptual concretamente através da apresentação de uma Arquitetura de Referência, descrevendo depois no capítulo seguinte o Software Livre ou de Código Aberto existente que preenche os blocos arquitetónicos, numa Arquitetura Concreta. Finalmente como conclusão, avalia-se a resposta à questão investigada.

## **1.7 Contribuições e Publicações**

A presente dissertação pretende contribuir para a comunidade numa primeira fase através de uma visão global dos normativos e da possibilidade de criação de um sistema que os cumpra, e, numa segunda fase, com a publicação de artigos acerca das várias fases e(ou) artefactos do ciclo documental: aquisição, indexação, gestão de registos, preservação digital, arquivo e respetivas implementações usando *Software* Livre ou de Código Aberto.

## 2 Revisão de Literatura

O que atualmente designamos por Arquivos surge quase em simultâneo com a criação da vida em sociedade derivado da necessidade de criar provas dos compromissos sociais celebrados entre as pessoas. A sua importância foi crescendo já que se sentiu grande necessidade de preservar a memória de contratos entre pessoas e organizações, de tal forma que levaram Napoleão a afirmar que “um bom arquivista é mais útil ao governo que um bom general do exército” (Rousseau, Couture, & Arès, 1998). A evolução do Arquivismo tem levado também à evolução de conceitos, por exemplo, o conceito unificado de Arquivo é por autores como Schellenberg (1956) separado em Arquivos e Registos. Ao longo dos séculos muitas formulações foram pensadas e definidas para a organização dos arquivos. Mais recentemente surgiram teorias importantes que não só originaram a criação de normativos como foram implementadas num grande número de arquivos e também sob a forma de requisitos em Sistemas de Informação específicos.

A Gestão de Documentos é uma área mais abrangente (Lage de Medeiros & Gomes do Amaral, 2010): tanto pode ser vista como Gestão de Informação Não Estruturada ou Gestão de Documentos de Arquivo, incluindo ou não fases ou mecanismos como a aquisição, os processos, a meta informação, os automatismos e(ou) a preservação de objetos digitais, Gestão de Registos ou Descrição Arquivística (Atherton, 1985). Esta abrangência e diferenças de visão notam-se na fragmentação existente tanto nos normativos, sejam eles “de facto” ou “de jure” (Silva, 2016), seja nas organizações que se dedicam ao seu estudo e normalização (Vieira & Borbinha, 2011).

Assim, durante a presente revisão de literatura, estudou-se o estado da arte por forma a cumprir os objetivos a que nos propusemos:

- Identificar os conceitos fundamentais relacionados com os Documentos de Arquivo;
- Compreender a importância dos normativos “de facto ou de jure”, relacionados com a Gestão de Documento de Arquivo;
- Identificar tecnologias importantes para a implementação de sistemas modernos de Gestão de Documentos de Arquivo;
- Compreender a importância do Software Livre e de Código Aberto na implementação de sistemas de informação para a gestão de Documentos de Arquivo;

- Identificar e desenhar a implementação de um sistema de Gestão de Documentos de Arquivo com os Produtos de Software Livre ou Código Aberto que melhor se adequam.

## 2.1 Princípios e Conceitos Arquivísticos

O conceito fundamental à volta do qual gira esta temática é o Documento. O *standard ODA – Office Document Architecture*, formalizado sob a forma da norma ISO 8613:1, define um Documento como: “*Uma quantidade de informação destinado à percepção humana, que pode ser trocado entre pessoas e(ou) sistemas*”, muito embora não exista uma definição consensual, esta é suficientemente abrangente dado que pode facilmente ser interpretada em termos físicos como sinais em memória do computador (documentos digitais) (Francke, 2005). Nesta definição está sub-entendido que um documento é registado num meio material que o permite, o qual se define como Suporte do documento. Exemplos de suporte são: papel, fitas magnéticas, micro-filmes, mas também discos virtuais que terão sempre, em última instância, uma existência física (num dispositivo de armazenamento de um computador) (Arellano, 2004).

A definição de Arquivo, bem como termos e conceitos associados estão atualmente bastante estabilizados, embora variem muito entre diferentes países, podemos definir de uma forma abrangente como o Conjunto de Documentos (também referido como Acervo Documental), criados ou recebidos por uma entidade física ou jurídica, pública ou privada, preservados por conterem provas acerca das atividades ou eventos da entidade (Millar, 2017). Autores importantes como Schellenberg (1956) separam o conceito abrangente usado em vários países em dois conceitos: Registo e Arquivo.

Schellenberg (1956) define Registo como:

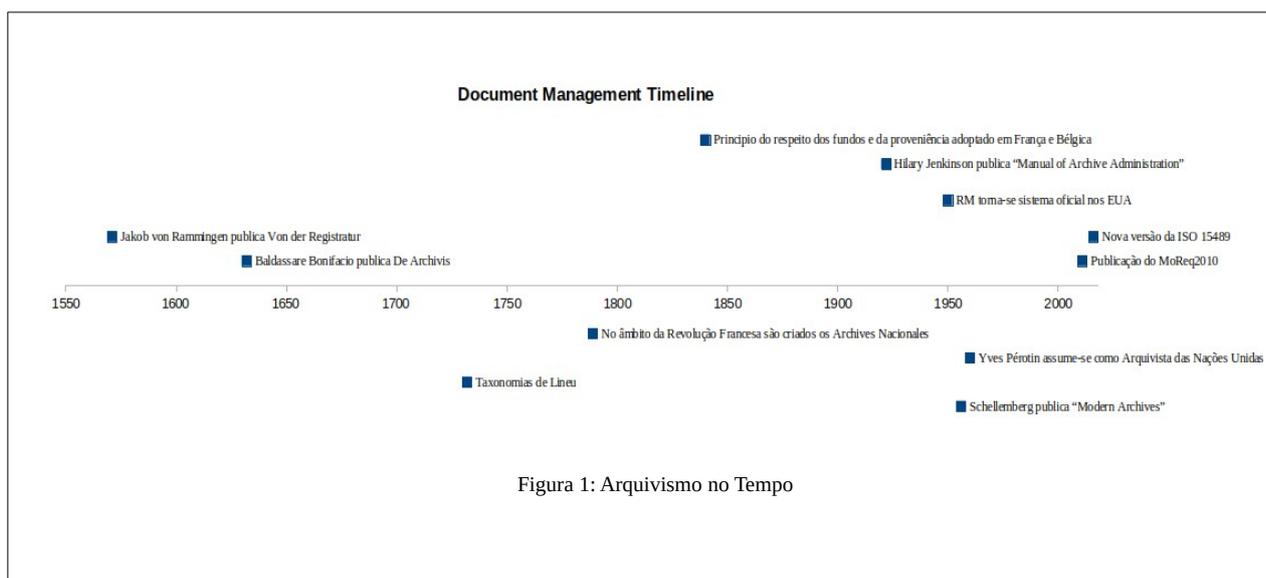
*“Todos os livros, textos, mapas, fotografias ou outros materiais documentais, independentemente da forma ou características físicas, criados ou recebidos por qualquer instituição pública ou privada na persecução das obrigações legais ou em estreita ligação com as transações do negócio, preservado ou preparados para tal pela instituição ou seu legítimo sucessor como prova das suas funções, políticas, decisões, procedimentos, operações, ou outras atividades ou devido ao valor dos dados que contêm.” (Schellenberg, 1956, p.16)*

E de acordo com a sua teorização, define-se um Arquivo da seguinte forma:

“Os registos de qualquer instituição pública ou privada que são julgados dignos de preservação permanente para efeitos de referência e investigação ou que foram selecionados para serem depositados numa instituição de arquivo” (Schellenberg, 1956, p.16)

As definições de Schelemberg (1956), clarificam que todos os materiais documentais recebidos ou criados pelas instituições se tornam registos e só mais tarde, após cumprimento dos prazos de conservação se poderão tornar documentos de Arquivo. São também importantes na medida em que refletem em grande medida a realidade dos tipos de Sistemas Informáticos existentes nesta área (Gestão Documental e de Descrição Arquivística).

É no início do século XX que se começam a esboçar as teorias modernas do arquivismo, concretamente a teoria das três idades, que, por exemplo, no início do século XX é aplicada com a lei geral de arquivos para a Administração Central Italiana, na qual já se define a ideia do Arquivo Corrente, Arquivo de Depósito e Arquivo Histórico (Duchain, 1992).



Vários autores enunciaram princípios orientadores da disciplina arquivística, dentre eles destaque Couture e Rouseau que enumera o princípio da proveniência, a teoria das três idades e o princípio da territorialidade. Outros, como Schellenberg definiram o valor primário e secundário dos documentos, bem como o valor da seleção e definições dos prazos de retenção e destruição (Williams, 2006).

O século XVIII assistiu ao desenvolvimento dos sistemas de classificação, como por exemplo o da Taxonomia de Lineu nas ciências biológicas (Aganette, Alvarenga, & Souza, 2010). O Arquivismo não ficou imune a estas evoluções, sendo neste século que claramente se tentam definir e ordenar os documentos em estruturas lógicas, tendo sido concebidos esquemas de

organização dos arquivos como a distribuição dos documentos em classes ou séries (Duchein, 1992). Na arquivística atual, a organização dos arquivos em taxonomias é implementada pelo instrumento a que se dá o nome de Plano de Classificação. Trata-se de uma organização hierárquica, em árvore. A cada nível podem ser associados meta-dados e(ou) regras. Os níveis inferiores herdam as definições do nó a partir do qual é gerado, podendo, no entanto, redefinir as definições herdadas que assim ganham prioridade.

O Princípio da Proveniência, definido pela primeira vez em 1841 por Natalis de Wailly (Bartlett, 1992) tornando lei que *“todos os documentos que provêm de um corpo, um estabelecimento, uma família ou um indivíduo formam um fundo e devem permanecer unidos”* (Duchein, 1992). Desta forma, segundo o princípio da proveniência *“os arquivos de uma mesma proveniência devem conservar a organização estabelecida pela entidade produtora, a fim de se preservar as relações entre os documentos como testemunho do funcionamento daquela entidade”* (Duchein, 1983). Na base deste princípio está o conceito de Fundo, cuja definição se pode definir como *“Conjunto de documentos de qualquer natureza –isto é, independentemente da sua idade, suporte, modo de produção, utilização e conteúdo– reunidos automática e organicamente – ou seja, acumulados por um processo natural que decorre da própria atividade da instituição –, criados e/ou acumulados e utilizados por uma pessoa física ou moral ou por uma família no exercício das suas atividades ou das suas funções“* (Duchein, 1983). Quando a um fundo já não são adicionados documentos usualmente porque a entidade produtora não se encontra em atividade, chama-se Fundo Fechado. Caso contrário, se a entidade ou seus herdeiros ainda produzirem documentos diz-se que se trata de um Fundo Aberto. Pode ocorrer que a hierarquia das organizações produtoras de documentos implique a subordinação de certos fundos em relação a outros. Neste caso estamos em presença de um Sub-fundo (Duchein, 1983).

Outros princípios não só não são tão consensuais nas várias escolas de arquivística, como não são determinantes na definição dos requisitos dos sistemas de informação e mesmo de normativos, assuntos principais desta dissertação. Fica no entanto a referência à sua existência:

O Princípio do Respeito pela Ordem Original, define que um arquivo deve conservar a organização estabelecida pela entidade coletiva, pessoa ou família que o produziu.

O Princípio da Pertinência, segundo o qual os documentos deveriam ser reclassificados por assunto sem ter em conta a classificação original.

O Princípio da Territorialidade, onde se define que os arquivos públicos, próprios de um território seguem o destino deste último, ou seja os arquivos devem ser entregues ao serviço com jurisdição arquivista sobre o território a que o conteúdo destes se refere.

Importante para efeitos da dissertação é o conceito de Valor do Documento. Define-se que um documento tem valor primário quando o seu interesse é administrativo, legal, jurídico, fiscal, contabilístico ou financeiro. Considera-se que tem valor secundário quando o seu interesse é histórico, para pesquisa, informacional ou cultural. Estes conceitos estão intimamente ligados aos arquivos correntes ou intermédios - valor primário - ou permanente - valor secundário - (Lage de Medeiros & Gomes do Amaral, 2010).

Devido ao aumento exponencial das massas documentais nas sociedades e à ocorrência de reflexão sobre Records Management nos EUA, a abordagem da teoria das três idades teve nesse país especial aceitação. Esta reflexão surge devido à situação de saturação dos depósitos verificada no final da Segunda Guerra Mundial originando mesmo o aparecimento de uma estrutura artificial – o pré-arquivo ou arquivo intermédio – cuja finalidade era preparar os documentos com origem nos serviços produtores para a sua inserção nos arquivos históricos (Schwartz & Cook, 2002).

A publicação de normativos correspondentes à evolução teórica da Gestão de Documentos de Arquivo, ocorreu ao longo de várias décadas e em todos eles existem várias versões. O seu contexto temporal é importante pois ajuda a compreender de que forma pode existir influência entre diferentes trabalhos.

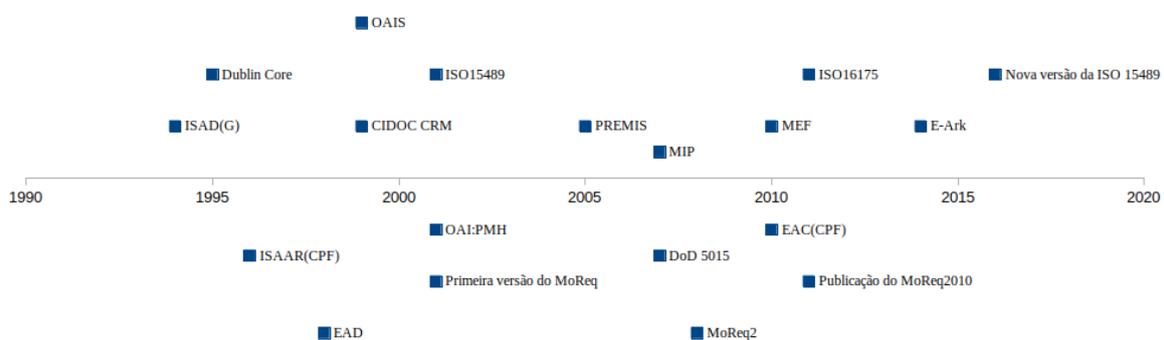


Figura 2: Histórico de Publicação de Normas relacionadas com Gestão de Documentos

Os principais esforços de normalização relacionados com a Gestão de Documentos de Arquivo, foram realizados na última década do século XX e na primeira do século XXI.

O *Record Management* tornou-se um sistema oficial em vigor nos EUA a partir de 1950, tendo como objetivo organizar e sistematizar a gestão de documentos administrativos nas administrações federais até à sua integração nos Arquivos Nacionais. Foram os autores anglo-saxónicos que desenvolveram a teoria e colocaram em prática o novo conceito de arquivo intermédio. Na cultura francófona, o arquivo intermédio surgiu apenas nos anos 60 do século 20, através dos artigos de (Pérotin, 1966).

A teoria das três idades também chamada de ciclo de vida dos documentos tem raízes nas reflexões realizadas pelos arquivistas perante os seus problemas de massa documental acumulada. Embora seja criticada por muitos autores, é a teoria mais usada na gestão de documentos seja em arquivos públicos ou privados. A sistematização obtida com teoria das três idades permitiu o desenvolvimento de técnicas e normativos que se aplicam em formas de Organização, Gestão e implementação de Sistemas Informáticos (A. C. Flores, 2004). A teoria das três idades define a existência de três tipos de arquivos: Arquivos Correntes, onde se encontram os documentos que, mesmo sem movimentação constituem objeto de consultas frequentes, Arquivos Intermédios, que arquivam os documentos de acesso esporádico mas que por razões de interesse administrativo, aguardam eliminação ou recolha para Arquivo Permanente, e os Arquivos Permanentes, que tratam dos documentos que, por razões de valor histórico ou probatório devem ser definitivamente preservados (Pérotin, 1966). É ainda importante voltar a realçar o trabalho de Schellenberg (1956), que distingue as operações nos arquivos correntes dos permanentes, distinção essa que muito contribuiu para a diferenciação dos normativos e sistemas informáticos entre Gestão Documental e Descrição Arquivística (Sousa, 2006). Assim, os arquivos corrente e intermédio estão muito ligados às questões práticas da gestão e os arquivos permanentes ou definitivos mais ligados aos historiadores.

No ciclo de vida dos documentos, um documento mantém-se no arquivo corrente ou intermédio durante um prazo, ao qual se chama de Prazo de Retenção, findo o qual se decide qual a Ação de Destino, isto é, acerca da sua Destruição, Conservação (total ou parcial) com transferência para o Arquivo Permanente. Após a decisão, é comum iniciar-se o Período de Confirmação durante o qual se aguarda uma confirmação para efetivar a ação de destino. O Prazo de Retenção e o destino dos documentos varia consoante a sua natureza. A definição do Prazo de Retenção e da Ação de Destino é efetuada através de um instrumento de gestão chamado Tabela de Seleção. A sua definição está intimamente ligada ao Plano de Classificação: sendo implementada sobre a forma da definição do Prazo de Retenção e Ação de Destino para cada Classificação (fazendo uso os mecanismos de hereditariedade da

hierarquia). Desta forma, quando um documento é classificado, fica-lhe associado o Prazo de Retenção e Ação de Destino (Schellenberg, 1956).

Também nos arquivos, tem vindo a ser dada uma importância crescente ao conceito de Interoperabilidade Semântica. O termo interoperabilidade é definido pelo Online Dictionary for Library and Information Science (ODLIS), como:

*“The capability of a computer hardware or software system to communicate and work effectively with another system in the exchange of data, usually a system of a different type, designed and produced by a different vendor.”* («ODLIS», 2018).

À face mais visível da Interoperabilidade chama-se Interoperabilidade Técnica. Esta define padrões e formatos tanto para as comunicações como para a sequência e forma de organização dos dados transferidos entre sistemas e organizações. No entanto, pelo facto de ser possível transferir informação entre sistemas e ambos saberem como a gerar e carregar, a informação pode não ter exatamente o mesmo significado. A Interoperabilidade Semântica relaciona-se com o significado ou semântica das informações nos vários sistemas. Por exemplo, em sistemas com Interoperabilidade Semântica, será mais provável que o significado de “autor” seja o mesmo em todos os sistemas que interoperam. Para atingir a Interoperabilidade Semântica usam-se técnicas como esquemas de meta-dados, classificações, tesouro e cada vez mais as Ontologias (Sayão & Marcondes, 2008) . A Interoperabilidade Semântica coloca-se também (mas não só) ao nível dos Documentos de Arquivo. A nível nacional, a Direção Geral do Livro, Arquivos e Bibliotecas desenvolveu algumas iniciativas para promover maior nível de Interoperabilidade Semântica (Penteado, Henriques, & Lourenço, 2012). Também a nível Europeu, com o Quadro Europeu de Interoperabilidade se estabelecem princípios orientadores, bem como um modelo conceptual, com o objetivo de aumentar a partilha de informações e conhecimentos entre as organizações («BAD», 2019).

Retomando o tema da divisão, estabelecida na teoria das três idades, e pelos trabalhos de Schelemberg (1956), é importante referir que os Arquivos Correntes e Intermédios estão muito ligados às questões práticas da gestão e os arquivos permanentes ou definitivos são mais ligados aos historiadores, possuindo naturezas e formas de gestão diferentes (Schellenberg, 1956), facto que se reflete na divisão entre Registos e Arquivos também observada nos normativos, e conseqüentemente nas implementações de Sistemas Informáticos, existindo especificações relativas à Gestão Documental (Gestão de Registos) e a Arquivos Definitivos ou Permanentes (Descrição Arquivística). A área da Gestão Documental – *Records Management* na nomenclatura Anglo-saxónica - referem-se geralmente aos

arquivos correntes e intermédios, sendo que a Descrição Arquivística se aplica aos documentos permanentes (Alberts, Schellinck, Eby, & Marleau, 2010). Alguns exemplos de normativos relativos à Gestão Documental é a norma ISO15482 e a especificação MoReq 2010. No arquivo permanente frequentemente aplica-se a ISAD(G) do International Council of Archives mas também o CIDOC CRM.

A Preservação Digital tem também vindo a ser alvo de grande preocupação, uma vez que, devido a fatores tão diversos como a obsolescência de formatos de ficheiros ou de suportes físicos, ou mesmo o desaparecimento de programas de leitura adequados, existe o perigo de desaparecimento da Informação armazenada em formato Digital. De tal forma que esse perigo pode ser maior do que o desaparecimento dos documentos existentes em formatos Físicos. Para mitigar este risco, nas últimas décadas a comunidade ligada aos arquivos tem colocado uma grande ênfase na definição de políticas, tecnologias e normas que aumentem a probabilidade de os documentos que existem apenas em formato digital, poderem ser acessíveis dentro de algumas décadas.

Assim, a investigação efetuada sobre trabalhos científicos na área da gestão de Documentos de Arquivo, levou à limitação do âmbito desta dissertação em três grandes áreas:

- Gestão Documental - também conhecida como *Records Management*, foca-se sobre os documentos de acesso maioritariamente administrativo e frequente e dos quais ainda não se decidiu se devem ser armazenados de forma permanente;
- Descrição Arquivística – insere-se sobretudo nos documentos que, pelo seu interesse histórico ou probatório devem ser preservados de forma definitiva ou permanente;
- Preservação Digital – área que se preocupa com a garantia de acesso futuro aos documentos independentemente do formato do documento ou do suporte onde se encontra armazenado.

Outras teorias documentais, como o do *Continuum* são importantes, mas por enquanto não há um consenso quanto à sua adoção por parte dos grandes arquivos nacionais e estrangeiros (Frank Upward, 2000), Durante a década de 90 do século XX alguns arquivistas, nomeadamente Frank Upward (1996), da Austrália, elaboraram o novo modelo dos “registos contínuos”. Esta teoria aproxima-se mais do que os Sistemas Informáticos de Gestão de Documentos de Arquivos podem disponibilizar, referindo-se ao conjunto global da existência de registos e a um conjunto coerente de processos de gestão desde o tempo da criação dos registos até à sua preservação e(ou) utilização dos registos como arquivos (Upward, 1996).

A Gestão de Arquivos suporta-se em alguns objetivos: disponibilizar documentos de forma completa e oportuna a fim de permitir a tomada de decisões eficientes; processar os documentos armazenados de forma eficiente; disponibilizar o acesso a documentos com o menor custo possível; tornar os documentos mais úteis para os utilizadores; eliminar documentos quando estes deixam de ser necessários (Schellenberg, 1956).

A Gestão Documental, muitas vezes reduzida à fase de armazenamento e(ou) descrição, refere-se no entanto a um grande conjunto de fases, cada uma das quais implementada com recurso a tecnologias que muitas vezes são bastante abrangentes (Rousseau et al., 1998).

A Produção de documentos corresponde à fase durante a qual os documentos são construídos. É uma fase cada vez mais suportada por sistemas digitais, sejam eles processadores de texto, máquinas fotográficas ou de filmar, ou relatórios com origem em sistemas de gestão mais genéricos das organizações (ex: *Enterprise Resource Planning* - ERP, *Customers Relationship Management* - CRM). Atualmente, ainda é comum considerar a geração automática de documentos com base em *templates* uma parte da gestão documental, não sendo, no entanto uma assunção muito consensual (Smith & McKeen, 2003).

Dado que muitos documentos não são gerados nos sistemas de gestão documental, podendo nem sequer ter existência digital, uma das tarefas, a de aquisição que quase pode ser considerada tarefa de integração, consiste em criar uma cópia digital de um documento em formato físico de forma a que possa ser devidamente ingerido e classificado no sistema de gestão documental (Laumer, Beimborn, Maier, & Weinert, 2013). São exemplos de módulos de aquisição: software de digitalização (mais ou menos complexos, com mais ou menos automatismos), integração com sistema de mensagens de correio eletrónico, capturando as mensagens de acordo com regras, ou os sistemas antes chamados *Computer Output to Laser Disk* – COLD, que capturam os documentos “à saída” da sua emissão no ERP e os inserem devidamente indexados no sistema de gestão documental (de Lima Baldam, 2002).

O tratamento manual da documentação, dada a crescente quantidade, é cada vez mais complementado com mecanismos de extração de meta-informação dos conteúdos originais. Alguns exemplos de extração são: descoberta automática de formatos dos documentos (*Layout Detection*), extração de texto existente em imagens (*Optical Character Recognition* - OCR), conversão de som em texto (Ribeiro, 2018).

Os mecanismos de extração são cada vez mais complementados com mecanismos de transformação da meta-informação com base no que foi possível extrair e em regras que permitem derivar informação com base noutra mais vaga. Neste particular, mecanismos de

*machine learning* (Samuel, 2000) podem tornar-se extremamente úteis para reduzir custos e melhorar produtividade.

A tarefa em que os sistemas de gestão documental mais se têm preocupado é no arquivo coerente e seguro dos documentos em formato digital. Este arquivo é conseguido por uma peça de *software* a que se chama Repositório (United States Patent N. US7941431B2, 2011).

Ao repositório estão normalmente associadas funcionalidades que permitem descrever o documento e seu contexto com meta-dados (dados acerca dos dados). A grande maioria dos sistemas, além de implementar conjuntos de meta-dados normalizados, são expansíveis, permitindo a criação dos campos de descrição adequados a cada arquivo.

Os planos de classificação, apesar de também serem meta-dados, têm um contexto muito específico pois são hierárquicos, têm permissões e regras associadas, relacionadas com a conservação e acesso aos documentos. Sendo uma tarefa sobretudo manual, as capacidades computacionais atuais permitirão a breve trecho automatizar a classificação de uma cada vez maior percentagem dos documentos (Sousa, 2006).

É também fundamental garantir a autenticidade dos documentos, nomeadamente quanto ao seu conteúdo e ao momento da sua geração e(ou) inserção no sistema de gestão documental. Estas tarefas são garantidas pela utilização de certificados digitais e selos temporais que, ao ser adicionados aos documentos, e estando ligados a autoridades certificadoras reconhecidas pelos estados, garantem a sua autenticidade. Não é ainda consensual a perenidade dos certificados, dado que uma entidade certificadora pode no futuro desaparecer (Warasart & Kuacharoen, 2012).

O arquivo, também efetuado num repositório, é muitas vezes concretizado em repositórios específicos, criados para documentos que chegaram à fase de arquivo definitivo ou histórico. Pode, no entanto, em muitos casos, ser mantido no repositório de arquivo corrente ou intermédio com modificação dos meta-dados por forma a estes cumprirem os normativos dos arquivos definitivos.

## **2.2 Normativos**

As áreas da Gestão Documental, por razões históricas não incluem apenas os campos da conceção, exploração e administração de documentos de arquivo, incluindo também a definição de políticas, procedimentos, recomendações e responsabilidades, sendo que cada um

destes campos pode ser referido em vários normativos *de facto* ou *de jure*, mais ou menos especializados em sub-áreas dos Documentos de Arquivo (Vieira & Borbinha, 2011).

## 2.2.1 Principais Esforços de Normalização

A evidente fragmentação, apresentada na tabela que abaixo se apresenta, mostra também a existência de diferentes organizações que se têm vindo a preocupar com uma maior definição e sistematização destes campos.

Durante a investigação efetuada, os principais estudos científicos estudados levaram-nos a uma seleção dos principais normativos.

Desta forma, na tabela que se segue, apresenta-se um resumo dos principais documentos que autores de estudos e artigos definem como relevantes para a sistematização da Área dos Documentos de Arquivo.

**Tabela 1: Principais Normativos**

Norma	Entidade	Sub-Área	Descrição
ISAD(G)	ICA	Descrição Arquivística	Norma Geral Internacional para Descrição Arquivística (ICA, 2011)
ISAAR(CPF)	ICA	Descrição Arquivística	Norma Internacional de Registo de Autoridade Arquivística para Pessoas Coletivas, Pessoas Singulares e Famílias (Comitê de Normas de Descrição, ICA, 2004)
ISO15489/ NP4438	ISO/IPQ	Records Management	Informação e documentação – Records Management («Norma NP 4438-1», 2005)
EAD	LOC	Descrição Arquivística	Formato concreto (XML) para meta-dados de descrição de fundos e coleções («EAD», 2019)
EAC(CPF)	SAA	Descrição Arquivística	Formato concreto (XML) para descrição de criadores de material arquivísticos («EAC-CPF», 2019)
ISO16175	ISO	Records Management	Princípios e Requisitos funcionais para Records Managements em ambientes eletrónicos («ISO», 2018)
MoReq2010	DLM Forum	Records Management	Especificação de Requisitos

Norma	Entidade	Sub-Área	Descrição
			funcionais e arquiteturas para Sistemas de Gestão Documental (DLM Forum, 2010)
DOD5015.2	DOD	Records Management	Standard para definição de Critérios de Desenho de Aplicações de Software para Records Management («DoD Standard 5015.2», 2007)
OAIS ISO14721	/ CCSDS	Preservação Digital	Modelo de Referência para a implementação de Sistemas de Preservação de Informação em Formato Digital (CCSDS, 2012)
PREMIS	PWG	Preservação Digital	Estratégias de implementação de meta-dados para Preservação Digital («Premis Working Group», 2016)
OAI-PMH	OAI	Indexação e Pesquisa	Protocolo para Recolha de Meta-dados («Open Archives Initiative», 2019)
E-ARK	CE	Preservação Digital	Projeto de Investigação de Métodos e Tecnologias para Arquivo («E-Ark Project», 2014)
MEF	DGLAB	Records Management	Macro Estrutura Funcional – Representação das funções desempenhadas na Administração Pública Portuguesa; Modelo de Classificação (Lourenço, Henriques, Penteadó, 2013)
MIP	DGLAB	Records Management	Esquema de Meta-dados para troca de documentos na Administração Pública Portuguesa (Barbedo & Corujo, 2012)
CIDOC CRM /	ICM	Descrição	Ontologia para Conceitos e

Norma	Entidade	Sub-Área	Descrição
ISO 21127		Arquivística	Informação de Instituições de Herança Cultural e Museus («CIDOC CRM», 2019)
METS	Library Of Congress	Descrição Arquivística	Esquema XML de Meta-dados para descrever objetos de bibliotecas digitais (Library of Congress, 2015)

## 2.2.2 Entidades Normalizadoras

Os normativos estudados são resultado do esforço de um conjunto de entidades de vários tipos e com várias existências jurídicas. Parte deles foram criados por grupos mais ou menos informais, nacionais ou internacionais, sendo algumas delas mais tarde transformados em standards internacionais, por exemplo o modelo de referência OAIS transformou-se na norma internacional ISO 14721 (DeNardis, 2010).

No caso concreto da área da Gestão de Documentos de Arquivo, as normas e(ou) standards tem sido criados por várias organizações. A investigação efetuada permitiu identificar as principais instituições produtoras, resumindo-as na tabela que se segue.

**Tabela 2: Entidades com influência na criação de normas**

Nome	Descrição
ISO	International Organization for Standardization («ISO», 2018)
IPQ	Instituto Português da Qualidade («IPQ», 2018)
ICA	International Council of Archives («ICA», 2019)
DLM Forum	Grupo de instituições europeias ligadas a arquivos, gestão de registos e de informação («DLM Forum», 2018)
Biblioteca do Congresso dos EUA	(«Library of Congress», 2019)
CCSDS	Consultative Committee for Space Data Systems. Fundada em 1982, tem como membros as principais Agências Espaciais Nacionais («CCSDS», 2019)
Digital Library Federation	Comunidade organizada pelo Conselho de Bibliotecas dos EUA («DLF», 2019)
SAA	Society of American Archivists («SAA», 2019)
PWG	(«Premis Working Group», 2016)

Nome	Descrição
OCLC	(«Online Computer Library Center», 2019)
RLG	Research Libraries Group («RLG», 2019)
OAI	(«Open Archives Initiative», 2019)
DGLAB	Direção Geral do Livro Arquivos e Bibliotecas. Criada em 2012, da fusão da DGARQ - Direção Geral dos Arquivos e DGLAB – Direção Geral do Livro (DGLAB, 2013)
ICM	International Council of Museums («ICOM», 2019)

### 2.2.3 Normativos que definem formatos

Parte dos normativos definem concretamente o formato de ficheiros eletrónicos de suporte. Decidiu-se enumerar quais os formatos definidos dado que estes são definições objetivas que os programas podem implementar.

**Tabela 3: Formatos eletrónicos para Gestão de Documentos de Arquivo**

Nome	Descrição
EAD	Encoded Archival Description. Norma baseada em XML para a descrição de fundos e coleções.
E-ARK SIP, DIP, AIP	Definição concreta de formato para os Pacotes de Informação usados na Preservação Digital.
PREMIS XML Schema	Esquema XML para definição dos meta-dados para preservação de objetos digitais.
EAC	Encoded Archival Context. Esquema XML para descrição de instituições, pessoas e famílias
MoReq2010 Export Format	Esquema XML para exportação de documentos de um Sistema de Gestão Documental por forma a poder ser importado noutra sistema

### 2.2.4 Gestão Documental

Na área da Gestão Documental, ou de *Records Management* nos países Anglo Saxónicos, conseguimos encontrar normativos que são desenvolvidos ou patrocinados por instituições com bastante poder efetivo, como o Departamento de Defesa dos EUA ou a Comissão Europeia. Tal facto deve-se à essência desta sub-atividade que é importante para o

funcionamento diário das instituições e para seu relacionamento com as entidades estatais reguladoras. Destacam-se as seguintes normas:

- DOD 5015.2 («DoD Standard 5015.2», 2007)
- ISO 15489 (ISO, 2016)
- MoReq2010 (DLM Forum, 2010)

Outras normas, de âmbito nacional, como a MIP (Barbedo & Corujo, 2012) e a MEF (Lourenço, Henriques, Penteado, 2013), têm como objetivo melhorar a organização, classificação e interoperabilidade semântica dos documentos nas instituições públicas nacionais.

### 2.2.4.1 ISO15489

A norma ISO15489 (ISO, 2016), cuja primeira versão remonta a 2001, define de que forma se devem Gerir Registos de Negócio, termo anglo-saxónico para “Gestão de Documentos de Arquivo” (Healy, 2010)

Divide-se em duas partes: ISO15489-1:2001, onde são definidos os Princípios Orientadores e a ISO15489-2:2001 que consiste num Guia Prático para a implementação de Sistemas de Gestão Documental.

Recentemente foi publicada uma nova versão, a ISO15489-1:2016, que define princípios e conceitos para a criação, captura e Gestão de Registos.

**Tabela 4: Resumo Descritivo de Enquadramento da ISO15489**

O quê? Princípios Orientadores e Guia Prático para implementação de Sistemas de Gestão Documental	
Quem? International Organization of Standards	Como? Através de um sub-comité relacionado com Gestão de Registos
Quando? 2001	Porquê? Para normalizar conceitos e práticas de governança da Informação Documental

### 2.2.4.2 MoReQ 2010

O MoReq (DLM Forum, 2010) foi criado pelo DLM Forum («DLM Forum», 2018), em cooperação com a Comunidade Europeia.

O extenso do acrónimo: *Modular Requirements for Records Systems*, deixa antever que se trata de uma especificação formal de requisitos, passíveis de ser usada como base de trabalho para o desenvolvimento de Sistemas de Gestão Documental.

Foi apresentada com o objetivo claro de: Criar uma Referência Mundial de Requisitos para Sistemas de Gestão Documental.

A primeira versão, denominada MoReq foi criada em 2005 pelo Fórum DLM, uma associação de arquivos públicos, de especialistas e outras partes interessadas do governo e da indústria, fortemente patrocinados pela Comunidade Europeia.

Em 2009 foi efetuada uma nova versão, que principalmente acrescentou testes e um modelo de meta-dados com especificação XML.

Em 2010, face às evoluções tecnológicas, evoluiu para a versão MoReq2010, para a qual foram efetuadas modificações na estrutura e nomes de conceitos, com todas as implicações destas alterações (Vieira & Borbinha, 2011).

O MoReq2010 inclui:

- Definição de conceitos;
- Arquitetura de referência;
- Definição de Requisitos não funcionais;
- Modelo de meta-dados;
- Documentação de testes;
- Esquema XML para o modelo de meta-dados.
- Textos explicativos de todo o modelo de referência e dos artefactos disponibilizados

Um sistema que implementa o MoReq2010 chama-se *MoReq Compliant Record System* – MCRS. Para que o seja, deve estar construído com uma arquitetura modular, implementar os conceitos chave, cumprir os requisitos definidos pelo MoReq2010 e passar nos testes definidos pela *framework* de testes.

Os principais conceitos do MoReq2010 são as Entidades, que consistem na informação gerida pelo sistema - sendo as principais os Registos/Documentos e os Utilizadores - e Serviços, que são funcionalidades que manipulam e gerem as entidades.

São definidos serviços principais, nos quais se incluem: serviço de registo, serviço de utilizadores e grupos, serviço de classificação, serviço de exportação, serviço de pesquisa, serviços de destruição e(ou) conservação, serviços de calendarização da destruição (*disposal*). São também definidos serviços relativos ao modelo: gestão do modelo e papeis do modelo.

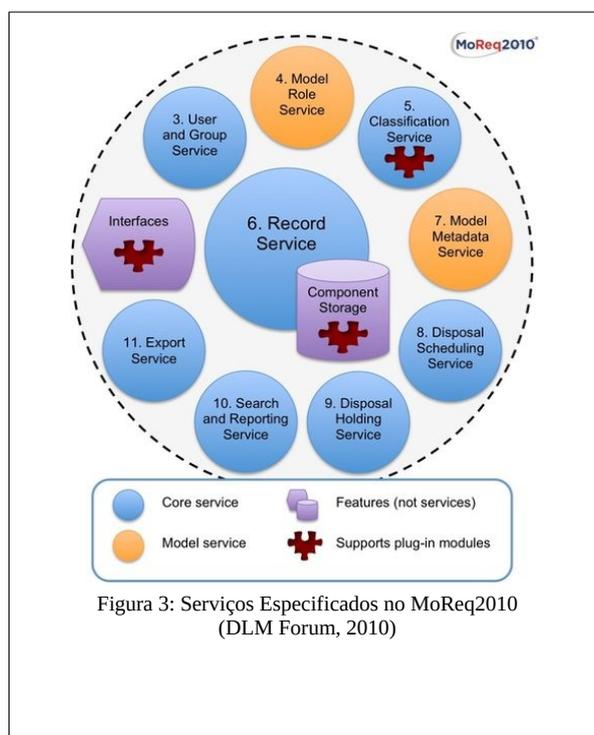
**Tabela 5: Resumo Descritivo do MoReq2010**

O quê? Modelo de Requisitos para implementação de Sistemas de Gestão Documental	
Quem? DLM Fórum, Comissão Europeia	Como? Através da criação de um grupo de trabalho com financiamento da União Europeia
Quando? 2011	Porquê? Para concretizar a definição dos requisitos necessários para implementar um Sistema de Gestão Documental

O MoReq é uma especificação baseada em serviços. As principais componentes de um sistema que implemente o MoReq são:

- Serviços principais (*core*)
- Serviços de Modelo
- Funcionalidades
- Módulos de extensibilidade (*plug-in*)

A imagem que se segue esquematiza a visão serviços do MoReq2010.



Cada serviço está encarregue de gerir um conjunto de entidades como de seguida resumidamente se descreve:

- *Model Role Service*, inclui as Listas de Controlo de Acessos (ACL) e os *Roles* de cada utilizador;

- *Record Service*, gere as entidades de Agregações, Componentes e Registos;
- *Classification Service*, encarregue de gerir as Classes;
- *Disposal Holding Service*, que gere as políticas de retenção;
- *Disposal Scheduling Service*, encarregue de gerir os calendários de retenção ou destruição;
- *System Services*, gere os tipos de Entidades, Eventos, Definições de Funções e Serviços;
- *User and Group Service*, que contem a gestão das entidades Utilizadores e Grupos;
- *Model MetaData Service*, encarregue de gerir as definições dos elementos de Metadados.

### 2.2.4.3 DOD 5015.2

Em 2007 o Departamento de Defesa dos Estados Unidos da América definiu sob a forma de norma, um conjunto de requisitos para a Gestão Documental (países Anglo-saxónicos usam o termo *Records Management*). A norma foi batizada de “*Department of Defense Standard 5015.2* («DoD Standard 5015.2», 2007) *Electronic Records Management Software Applications Design Criteria Standard*” (Seymour, 2017).

Esta norma para além de ser definida como obrigatória no DoD, foi também adotada por departamentos centrais ou locais da Administração Americana e foi uma grande influência na criação de outras normas, como por exemplo a ISO15489 (Duarte, 2017).

A norma define conceitos, estabelece políticas, atribui responsabilidades na gestão de registos em quaisquer meios, incluído eletrónicos e “estabelece requisitos funcionais para identificação, classificação, armazenamento, segurança e outros procedimentos para a gestão da informação” (Duarte, 2017)

**Tabela 6: Resumo Descritivo da DOD 5015.2**

O quê? Standard do Departamento de Defesa dos EUA acerca de Gestão de Documentos de Arquivo	
Quem? Departement Of Defense of United States of America	Como? Criação e aplicação da norma no DoD e na Administração Pública, central e local dos EUA
Quando? 25 de Abril de 2007	Porquê? Para garantir que todo o DOD usa os mesmos conceitos políticas e responsabilidades na

#### 2.2.4.4 MEF

A MEF (Lourenço, Henriques, Penteado, 2013), Macro Estrutura Funcional foi criada em 2010 pela Direção Geral do Livro Arquivos e Bibliotecas, tendo como destinatários de aplicação a Administração Pública Central e Local de Portugal (Lourenço, Henriques, Penteado, 2013).

A sua criação teve como principais objetivos:

- Aumentar a probabilidade de existir interoperabilidade semântica na troca de documentos entre organismos do estado;
- Disponibilizar aos organismos da administração pública um modelo de classificação cujo significado e sintaxe sejam idênticos nos vários organismos da Administração Pública;
- Reduzir custos através da reutilização de instrumentos;
- Harmonização de linguagens.

A MEF é uma estrutura hierárquica com três níveis:

- Nível 1 – Funções;
- Nível 2 – Sub Funções;
- Nível 3 – Processo de negócio (produto/serviço).

Num sistema que cumpra a MEF o plano de classificação é hierárquico cumpre pelo menos os três níveis, incluindo o significado e as regras definidas (Torres, 2018).

**Tabela 7: Resumo Descritivo da MEF**

O quê? Modelo Classificador das funções desempenhadas por organizações do sector público.	
Quem? Direção Geral do Livro Arquivos e Bibliotecas.	Como? Definindo um modelo de Classificação, que inclui a estrutura e as orientações para o desenvolvimento dos planos de classificação concretos de cada organização.
Quando? Versão 1 – 2011 Versão 2 - 2012	Porquê? Para que as instituições do estado procedam à criação de um plano de classificação documental orientado às funções exercidas nas organizações por oposição à sua orgânica.

## 2.2.4.5 MIP

A MIP (Barbedo & Corujo, 2012)– Meta-informação para a Interoperabilidade foi criada pela DGLAB com o objetivo de facilitar a troca de documentos entre Organismos da Administração Pública de Portugal, mais concretamente para facilitar a interoperabilidade semântica. Os documentos, como já estudado, não são apenas constituídos pelos ficheiros que apresentam o seu conteúdo, sendo acrescidos de um conjunto de informação adicional que os descrevem nomeadamente em relação ao seu contexto. A informação descritiva é armazenada sob a forma de meta-informação, isto é, informação acerca da informação (o conteúdo de cada documento). Se cada instituição definir de forma diferente o conjunto de meta-informação associada aos seus documentos, torna-se inexecutável ocorrer uma troca de completa de documentos entre as organizações uma vez que teria de se efetuar a correspondência do significado da informação nas diferentes organizações, podendo esta correspondência não ser possível. A interoperabilidade semântica, quando obtida, é um conceito que define igual estrutura e significado dos elementos de meta-informação nas várias instituições. A MIP é um esforço do Estado Português para garantir que as instituições nacionais conseguem trocar de forma mais completa os documentos entre si (Barbedo & Corujo, 2012).

**Tabela 8: Resumo Descritivo da MIP**

O quê?: Definição de um esquema de meta-informação	
Quem? Direção Geral do Livro Arquivos e Bibliotecas.	Como? Definindo 17 elementos de meta-informação, incluindo os sub-elementos, a sua cardinalidade, tipos de dados aceites, constrangimentos e(ou) inter-relações.
Quando? Primeira Versão - 2007 Versão 1.0c - 2012	Porquê? Para facilitar a interoperabilidade semântica na troca de documentos entre instituições, ou seja garantir que os campos descritivos dos documentos são os mesmos e têm o mesmo significado em diferentes instituições

## 2.2.5 Descrição Arquivística

A Descrição Arquivística tem como principais objetivos: identificar e explicar o contexto e o conteúdo da Documentação de Arquivo.

De acordo com o conceito do ciclo de vida das três idades, pode ser considerada a fase do Arquivo Definitivo ou Histórico, no contexto da Gestão de Documentos de Arquivo.

Embora atualmente normas como a ISAD(G) (ICA, 2011) sejam colocadas em causa devido às novas necessidades de um modelo de descrição arquivística mais abrangente (Runa, Barbedo, & Almeida, 2018), a ISAD(G) é ainda a norma mais importante e com maior utilização no âmbito dos Arquivos Históricos ou Definitivos (Cook, 2011).

### 2.2.5.1 ISAD(G) / ISAAR(CPF)

A *General International Standard Archival Description*, consiste em regras de descrição de objetos com valor arquivístico. Os objetos descritos podem ter ou não existência ou representação digital em qualquer formato. Ao normalizar regras e formatos de descrição é de grande valor para a recuperação e troca de informação de descrições de objetos arquivísticos.

Foi definida pelo ICA – International Council of Archives («ICA», 2019), em 1994.

**Tabela 9: Resumo Descritivo da definição EAC**

O quê? Norma Internacional de Descrição Arquivística	
Quem? International Council of Archives	Como? Com a criação de um subgrupo da comissão AdHoc para standards de descrição, com o apoio da UNESCO
Quando? 1994	Porquê? Para promover a acessibilidade ao contexto e conteúdo da documentação de arquivo através da criação de um modelo de metadados e regras

As regras da ISAD(G) estão organizadas em sete zonas de informação descritiva:

- Identificação;
- Contexto;
- Conteúdo e Estrutura;
- Condições de acesso e de utilização;
- Documentação Associada;
- Notas;
- Controlo da Descrição.

A ISAD(G) descreve os objetos digitais incluindo-os na seguinte hierarquia que os classifica: (T. Lee & Iio, 2015)

Fundos – Conjunto de registos criados ou acumulados por pessoas, famílias ou instituições;

Sub-Fundos – Unidade administrativa subordinada aos fundos;

Séries – Unidade básica que se refere a um conjunto de documentos resultantes da mesma função ou atividade. Uma série pode ser dividida em Sub-Séries;

Ficheiros – Unidades de documentos organizados ou agrupados pela sua utilização ou pelo seu criador. Normalmente a unidade básica dentro de uma série;

Item – A unidade de arquivo mais pequena e indivisível;

Tanto como os conteúdos arquivados e a sua descrição (normalizada entre outras pela Normal Geral para Descrição Arquivística ISAD(G)), o contexto da peça arquivada é de extrema importância, concretamente o contexto da sua proveniência (Cunningham, 2011).

Para endereçar a informação da autoridade dos arquivos, a ISAAR(CPF) (Comitê de Normas de Descrição, ICA, 2004)– International Standard Archival Authority Record (Corporate, Personal, Families) foi elaborada pelo Conselho Internacional dos Arquivos, referindo-se ao registo e descrição de Autoridades Arquivísticas para pessoas Coletivas, Pessoas Singulares e Famílias. A primeira versão foi disponibilizada em 1996, a segunda versão em 2004. A ISAAR(CPF) permite identificar e registar as entidades produtoras de documentos, de outras entidades que fazem parte do ambiente e da documentação das histórias e administrativas e biográficas daquelas entidades, as suas responsabilidades funcionais, as relações que mantêm entre elas e as relações com os sistemas de arquivo (Comitê de Normas de Descrição, ICA, 2004).

A norma ISAAR(CPF) define quais os tipos de informação que podem ser incluídos num registo de autoridade e orienta acerca da forma como os registos podem ser desdobrados num sistema de controlo descritivo. Desta forma, para além das orientações, consiste sobretudo na definição de um modelo de meta-dados que devem ser usados sempre que necessário descrever uma autoridade.

De certa forma associado à norma ISAD(G), o ICA definiu também a («ISDIAH: International Standard for Describing Institutions with Archival Holdings | International Council on Archives», 2008), cujo objetivo é a normalização da meta-informação que descreve as instituições que detêm acervos arquivísticos

### 2.2.5.2 EAC – Encoded Archival Context

As teorias ligadas aos Arquivos já há muito reconheceram a importância do contexto em que os arquivos, manuscritos ou outras fontes de materiais foram criados. Desta forma, o contexto é considerado como um elemento crítico para uma interpretação rigorosa como prova.

Durante muitos anos, a informação contextual foi considerada parte da metodologia de pesquisa, sendo presente na fase de preparação dos investigadores para o seu contacto com as suas fontes. Acontece também, que a informação contextual é uma ferramenta poderosa para a descoberta das fontes primárias.

As descrições e bibliografias arquivísticas, tradicionalmente não continham informação contextual. Esta teria de ser obtida através da leitura de passagens narrativas nas quais se tornava difícil a obtenção de informação devidamente estruturada.

Para colmatar estas lacunas, o desenvolvimento de descrições estruturadas e normalizadas do contexto, como o EAC («EAC-CPF», 2019) ou a adoção do standard ISAAR(CPF), foram marcos importantes para permitir a definição mais explícita e estruturada e conseqüentemente permitir o seu registo e troca entre diferentes sistemas (Szary, 2006).

**Tabela 10: Resumo Descritivo da definição EAC**

O quê? Definição de Formato para arquivo e intercâmbio de informação contextual	
Quem? Grupo Informal de arquivistas com o apoio da Digital Library Federation	Como? Definindo um DTD – Document Type Definition para um formato XML
Quando? 1998	Porquê? Devido à dificuldade de troca de informação contextual por existirem diferentes representações.

### 2.2.5.3 CIDOC Conceptual Reference Model

O CIDOC CRM («CIDOC CRM», 2019) é um Modelo Conceptual, como tal, ao contrário de outras normas que são muito mais concretos, define os conceitos e as principais entidades.

O CIDOC CRM tem objetivos diferentes de normas como o ISAD(G), na medida em que em vez de definir a estrutura, relações e campos de meta-dados, foca-se mais em definir um conjunto de conceitos num domínio bem como os relacionamentos entre eles, constituindo-se assim como uma Ontologia (Runa et al., 2018).

O CIDOC CRM surge derivado do grande conjunto de normas centradas nos meta-dados. Além disso, tem um objetivo mais lato do que apenas a informação de arquivo, estendendo-se a áreas como os museus ou arqueologia promovendo a interoperabilidade entre informação heterogénea dos diferentes domínios e(ou) standards (Bruseker, Carboni, & Guillem, 2017).

**Tabela 11: Resumo Descritivo do CIDOC CRM**

O quê? Ontologia para conceitos e informações relacionados com a herança cultural e documentação museológica	
Quem? Grupo de Trabalho CIDOC do Conselho Internacional de Museus	Como? Definindo os conceitos, entidades e respetivas relações num domínio
Quando? 2006	Porquê? Para promover a interoperabilidade e reutilização entre conceitos relacionados com a herança cultural

## 2.2.6 Preservação Digital

Atualmente, existe uma crescente preocupação quanto à preservação de documentos digitais. Cada vez mais documentos não têm qualquer existência física e não há qualquer certeza de que no futuro se conseguirá aceder aos conteúdos, seja por obsolescência do formato dos ficheiros, seja por impossibilidade de executar os programas que os conseguem manusear (no futuro poderão não existir programas capazes de aceder ao conteúdo de um ficheiro gerado hoje). Neste momento existem muitos casos em que não é possível aceder a documentos de texto criados nos anos 90 pois não existem programas capazes de ler aquelas versões dos formatos. Também os suportes físicos não garantem que os documentos serão acessíveis durante um prazo razoável, seja pela sua deterioração, seja pela inexistência de leitores adequados. Observe-se por exemplo as muitas instituições lutam com dificuldades de leitura de documentos que existem em *disquetes* de 5,25 polegadas ou a dificuldade existente em aceder a vídeos suportados no formato Betamax (Ferreira, 2006). Para mitigar estas dificuldades foram criadas técnicas, complementadas com modelos claramente definidos que ajudam no aumento da probabilidade de garantia de acesso futuro à informação (Thomaz & Soares, 2009).

A Informação Digital, ao contrário de outros suportes de informação, de que o papel é um exemplo, carece de dispositivos intermédios para que a informação seja acedida e consumida.

Estes dispositivos e formatos têm a característica de estarem em constante evolução tornando-se fácil e rapidamente obsoletos (Ferreira, 2006).

Desafios importantes se colocam quanto à Preservação da Informação Digital, desde os formatos, a garantia de autenticidade, a meta-informação de preservação, obrigando por isso a uma clara definição de estratégias de preservação. Vários autores alertam atualmente para a necessidade de políticas que conduzam à implementação de estratégias de preservação de informação digital, sob pena de estarmos em face de perigos graves de perda de informação importante, provavelmente perigos maiores do que a perda de informação existente noutros suportes tangíveis (Arellano, 2004).

Estas preocupações derivaram no conceito de preservação digital, que se pode definir como “o conjunto de atividades ou processos responsáveis por garantir o acesso continuado a longo-prazo à informação e restante património cultural existente em formatos digitais” (Thibodeau, 2002).

Acompanhando as necessidades, durante as últimas décadas, ocorreu uma proliferação de sistemas, cada um com especificidades próprias, dificultando a transferência, migração e(ou) ligação entre a informação. Felizmente, tanto o aparecimento dos normativos, como a implementação de Sistemas Livres ou de Código Aberto está, não só a obrigar à evolução de tais sistemas, mas também a derivar na emergência de Sistemas que, no futuro, poderão permitir Preservar e Gerir Documentos de forma eficaz e de acordo com os principais normativos (Lopes, 2008).

De entre as várias normalizações, destaca-se o modelo OAIS (CCSDS, 2012), que nos próximos pontos se descreve.

### **2.2.6.1 Modelo de Referência OAIS**

O modelo *Open Archiving Information System* (OAIS), é resultado de um esforço conjunto do *Consultative Committee for Space Data Systems* (CCSDS) e da *International Organization for Standardization* (ISO), a partir de 1990, culminando com a sua publicação em 2003, posteriormente atualizado com a versão 2 em, 2012. Este modelo de referência veio a tornar-se standard ISO 14721:2012 (Lee, Slattery, Lu, Tang, & McCrary, 2002). O seu objetivo foi desenvolver um “conjunto de normas capazes de regular o armazenamento a longo-prazo de informação digital produzida no âmbito de missões espaciais” (Ferreira, 2006).

Desta forma, o Modelo Conceptual OAIS tornou-se uma referência no que se refere à preservação de Informação Digital, contribuindo notavelmente com: a identificação de componentes funcionais que deverão fazer parte de um sistema de informação dedicado à preservação digital; a descrição das interfaces do sistema; os objetos de informação que são manipulados no seu interior; a definição de terminologia própria que facilita a comunicação entre os intervenientes envolvidos na preservação de objetos digitais (Ferreira, 2006).

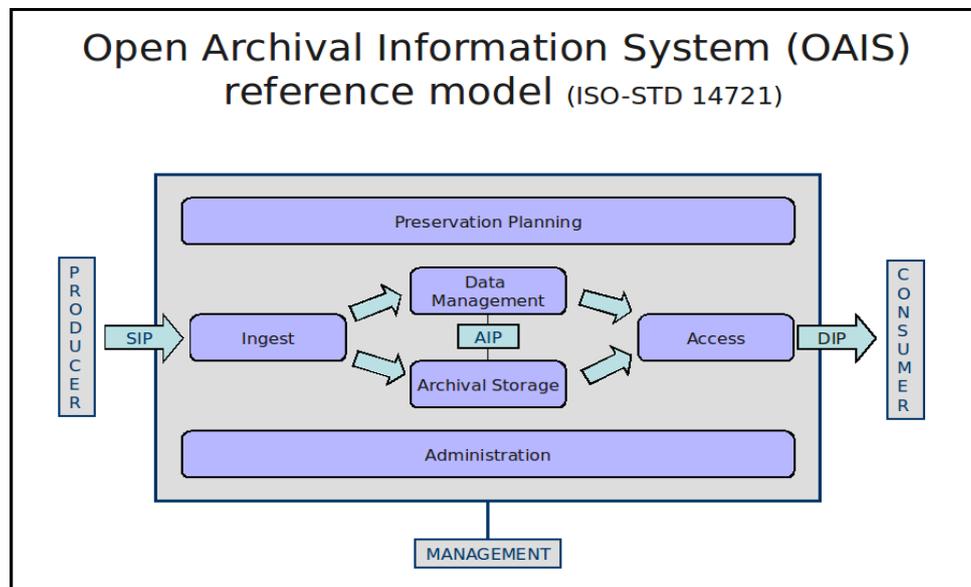


Figura 4: Modelo de Referência "Open Archival Information System (OAIS)(«Archivemática», 2019)"

Da figura anterior, que descreve esquematicamente o modelo OAIS, é importante referir os seguintes conceitos (CCSDS, 2012):

- OAIS – Open Archival Information System: sistema arquivo de uma uma organização responsável por preservar e disponibilizar informação a uma comunidade;
- Produtor: Entidade externa responsável pela submissão da informação;
- Consumidor: Entidade externa a quem é entregue a informação para disseminação preservada;
- SIP – Submission Information Packet: Conjunto de informação, agregada num pacote coerente, para submissão a um sistema OAIS. O Modelo de Referência não especifica o formato concreto nem do pacote, nem do seu conteúdo, pelo que cabe aos OAIS acordarem no formato a utilizar;
- AIP – Archival Information Packet. Pacote de Informação organizada com objetivo de ser armazenada;

- DIP – Dissemination Information Packet. Versão do pacote de informação cujo objetivo é a entrega aos consumidores em resposta a um pedido de acesso (Lavoie, 2008).

**Tabela 12: Resumo Descritivo do Modelo de Referência OAIS**

O quê? Modelo <i>Open Archival Information System</i> para Preservação Digital	
Quem? Comité Consultivo para os Sistemas de Dados Espaciais	Como? Definindo um modelo conceptual
Quando? 1999	Porquê? Para aumentar a probabilidade de preservação da informação relacionada com a industria espacial

### 2.2.6.2 METS

O Acrónimo METS (Library of Congress, 2015) significa *Metadata Encoding & Transmission Standard*. É um esquema que define um standard para a codificação de metadados descritivos, administrativos e estruturais de objetos de bibliotecas digitais (Dappert & Enders, 2008).

Trata-se de um esquema definido usando a linguagem *XML Schema*. É mantido pela Biblioteca do Congresso dos Estados Unidos da América («Library of Congress», 2019) (*Network Development and MARC Standards Office*) e desenvolvido pela *Digital Library Federation* («Digital Library Federation», 2019).

Um documento METS - ficheiro XML que descreve um Objeto Digital - pode ser utilizado no contexto da Preservação Digital, como Pacote de Informação, nomeadamente Pacote de Submissão de Informação – SIP. É usado como parte integrante da definição de um SIP do projeto E-Ark («E-Ark Project», 2014).

Um documento METS é composto pelas seguintes secções:

- Cabeçalho – Meta-dados que descrevem o próprio documento METS, ou seja, as secções que se seguem;
- Meta-dados Descritivos - Permitem a descoberta identificação e seleção nomeadamente através da identificação do assunto e do conteúdo do objeto;
- Meta-dados Administrativos – Descrição técnica das características físicas do objeto, meta-dados sobre gestão de direitos de autor;

- Ficheiros – Lista de todos os ficheiros e alguma informação estrutural de cada um deles (ex: localização do ficheiro, mimetype), que constituem as versões eletrónicas do objeto digital;
- Ligações Estruturais – Descreve a existência de ligações entre ficheiros que compõem o objeto digital (ex: um ficheiro *html* que referencia uma imagem);
- Comportamento – Serve para descrever a existência de comportamentos executáveis com o conteúdo no objeto METS, por exemplo a execução de um serviço web sobre um conteúdo.

**Tabela 13: Resumo Descritivo do esquema de meta-dados METS**

O quê? Standard XML de Meta-dados sobretudo relacionados com a preservação digital	
Quem? Biblioteca do Congresso dos EUA	Como? Definindo uma definição XML (DTD) e estrutura de pastas
Quando? 2001	Porquê? Para definir um formato standard de arquivo de objetos digitais

### 2.2.6.3 Projeto E-Ark

No âmbito da Preservação Digital, não poderia deixar de se referir o Projeto E-Ark («E-Ark Project», 2014). Acrónimo de *European Archival Records and Knowledge*, foi uma iniciativa Europeia cujo objetivo foi a definição de uma aproximação, consubstanciada sob a forma de projetos piloto e definições que demonstrem ser possível a implementação de sistemas robustos, escaláveis, capazes de cumprir os requisitos necessários em instituições públicas ou privadas, no que se refere ao arquivo, aquisição, preservação e reutilização de informação.

Em concreto, os pilotos abertos irão demonstrar sistemas funcionais de arquivos eletrónicos que cumpram o Modelo de Referência OAIS, nomeadamente a *ingestão, arquivo, e reutilização de dados estruturados e não estruturados*.

O projeto iniciou em 1 de Fevereiro de 2014, tendo terminada a sua fase de investigação em 2 de Fevereiro de 2017. Participaram no projeto um grupo de arquivos nacionais europeus, agências de investigação governamentais, fornecedores privados de soluções informáticas na área dos arquivos digitais e organizações não governamentais internacionais (Alföldi & Réthy, 2017).

Da criação dos projetos piloto, derivaram na definição de especificações, das quais se destacam

- Especificação comum, baseada no METS (Bredenberg et al., 2016);
- Pacote de Submissão de Informação – SIP;
- Pacote de Arquivo de Informação – AIP;
- Pacote de Disseminação de Informação DIP, definida no modelo OAIS para a entrega de informação aos consumidores após solicitação.

As especificações dos *Pacotes* concretos (AIP, DIP e SIP), baseiam-se na especificação comum. Nesta, um Pacote de Informação IP, existe sob a forma de uma “árvore” de pastas existente no sistema de ficheiros do dispositivo onde reside. A estrutura da “árvore” de pastas e ficheiros é fixa tal como de seguida se descreve:

- Nome IP – Pasta base do IP. O seu nome deve corresponder ao ID do IP no sistema de preservação. Todas as sub-pastas contêm a informação do pacote (Ex: IP\_577ER\_223);
  - METS.XML – Ficheiro descritor;
  - metadata – Pasta que contém todos os ficheiros de meta-dados organizados em sub-pastas;
    - descriptive – Pasta que contém ficheiros de meta-dados de descrição dos objetos a preservar;
    - preservation – Contém ficheiros de descrição relativos à preservação em si (ex: PREMIS.xml);
    - Other – Podem ser criadas tantas pastas relativas a outros tipos de meta-dados, por exemplo relativos a uma área específica do negócio da organização;
  - Representations – Pasta que contém as representações digitais dos ficheiros a ser preservados. Dentro de cada representação poderão existir as pastas de metadata, relativas à representação correspondente;
    - rep\_id1 – Pasta base de uma das representações dos objetos do pacote (ex: pdf);
      - data – Ficheiros, numa das representações do pacote;
        - file1.pdf, file2.pdf, file\_n.pdf;
      - documentation;

- DOC-format-spec.pdf;
- rep\_id (n) – As pastas anteriores (sub-pastas de rep\_id1) repetem-se tantas vezes quantas as representações (ex: pdf, doc, entre outros) existentes no pacote.

**Tabela 14: Resumo Descritivo de Enquadramento do Projeto E-Ark**

O quê? Definição concreta de formatos, guidelines, implementações	
Quem? E-Ark Project Team («E-Ark Project», 2014)	Como? Com financiamento da União Europeia
Quando? 2014	Porquê? Para obter maior consistência nos arquivos europeus

### 2.2.6.4 PREMIS

A sigla PREMIS deriva do grupo de trabalho com o “PREservation Metadata: Implementation Strategies”. Esse grupo criou um relatório chamado “PREMIS data Dictionary for Preservation Metadata”.

O relatório define um dicionário de dados para preservação de meta-dados, mas também um conjunto de boas práticas acerca da preservação de meta-dados (Caplan, 2009).

O dicionário de dados consiste na lista e relações de conjuntos de meta-dados e respetivos elementos (chamados de unidades semânticas) que os repositórios deverão implementar, por forma a cumprir as suas funções de preservação.

As funções de preservação podem variar consoante o repositório, mas em geral incluem ações que garantem que os objetos digitais sem mantêm viáveis, isto é: que podem ser lidos por programas; que são desenháveis, isto é: podem ser apresentados (sob a forma correspondente: som, imagem parada, imagem em movimento); que não foram inadvertidamente alterados e que quaisquer modificações (por exemplo derivadas da mudança de formatos) são devidamente documentadas (Congress & Committee, 2015).

Apesar de ser um dicionário de dados vocacionado para a preservação, no entanto o PREMIS não se foca em questões relacionadas com os meta-dados específicos de formatos, nem de informação da implementação e(ou) de permissões.

Assim, o PREMIS foca-se sobretudo nos meta-dados específicos às funções de preservação como por exemplo as ações de preservação ou as diferentes representações da informação.

O PREMIS e outros são parcialmente sobreposições ao Modelo de Referência OAIS. Existem várias tentativas de compatibilização entre os vários dicionários de dados, mas também aqui se nota o problema da fragmentação e multiplicação de definições, formatos e normas (Dappert & Enders, 2008).

**Tabela 15: Resumo Descritivo do Relatório PREMIS**

O quê? Esquema XML e lista de boas práticas para implementação de repositórios de preservação	
Quem? Grupo de trabalho PREMIS do Online Computer Library Center	Como? Implementando um esquema XML (XML Schema) e documentos com descrição de boas práticas
Quando? 2005	Porquê? Para aumentar a probabilidade de implementação de Preservação Digital de uma forma standardizada

## 2.3 Tecnologia

A investigação descrita nos capítulos anteriores, relativas ao estado da arte da Gestão de Documentos de Arquivo levantou a interrogação acerca de que implementações de Software de Gestão de Documentos (ou) de Arquivo existem, disponíveis com licenças que as permitam ser consideradas Software Livre ou de Código Aberto? Ao investigar para responder à questão anterior, tornou-se possível compreender que existem implementações que se adequam mais a alguns normativos e sub-áreas da Gestão Documental, e outras que resolvem problemas muito específicos e limitados no seu âmbito. Foi por isso importante complementar a questão de quais os Sistemas de Software Livre ou de Código Aberto que resolvem problemas da Gestão Documental, com uma terceira questão, sobre qual o grau de completude dos normativos que implementam? As respostas às questões anteriores levaram à percepção de que um sistema para Gestão de Documentos de Arquivo teria de ser implementado usando produtos de suporte, que implementam conceitos múltiplos que deveriam interagir de forma coerente. Esta assunção derivou na investigação acerca do estado da arte das Arquiteturas de Sistemas de Informação que permitam a cooperação e um correto funcionamento dos vários produtos que como um todo e que irão implementar um Sistema para a Gestão de Documentos de Arquivo permitindo assim a definição dos capítulos 4, que apresenta uma solução conceptual e capítulo 5 onde se concretiza a solução.

## 2.3.1 Arquiteturas de Sistemas de Informação

Tendo em conta os resultados da investigação, que apontam para:

- A inexistência de um produto que cubra todas as áreas da gestão de Documentos (ou) de Arquivo;
- A proliferação de formatos não só para os objetos digitais, mas também para os metadados descritores;
- A necessidade de que um Sistema desta natureza responda às necessidades dos utilizadores e seja capaz de gerir tanto pequenas como grandes quantidades de informação em diferentes formatos.

Tornou-se claro que deveria ser efetuada uma investigação mais profunda acerca do estado da arte de arquiteturas que sejam de tal forma versáteis que permitam a adaptabilidade de recursos mantendo um desempenho coerente com poucos ou muitos utilizadores, com pequena ou grande massa documental.

O presente capítulo descreve a investigação relacionada com os pontos chave das mais modernas arquiteturas de Sistemas de Informação, nomeadamente quanto à sua escalabilidade e integração.

### 2.3.1.1 Escalabilidade

*Escalabilidade* é uma característica dos sistemas, que indica a capacidade de as suas características poderem variar de forma a suportar um aumento de carga sem que perca as suas principais funcionalidades.

A investigação acerca de escalabilidade, no contexto desta dissertação ocorre, uma vez que os sistemas de Gestão de Documentos de Arquivo poderão ter enormes variações de dimensão, tanto quanto ao número de utilizadores que a ele acedem como à quantidade de documentos que gerem. De facto poderá ir de um sistema mono-utilizador até ao acesso simultâneo de muitos milhares de utilizadores. Da mesma forma poderá gerir poucas centenas de documentos ou muitas centenas de milhões, tornando assim importante que estes sistemas tenham capacidade de dimensionamento e adaptação para que seja possível funcionar nas várias realidades.

Não existem métricas concretas aceites para medição do grau de escalabilidade, muito embora tenham sido efetuadas várias tentativas para a sua definição (Hill, 1990).

Existem no entanto duas definições importantes quanto à arquitetura montada (Agerwala & Chatterjee, 2005):

- Escalabilidade Horizontal (também chamada *scale out/in*) – Consiste numa coleção de computadores (normalmente de baixa capacidade), cooperando de forma apresentar-se como uma única entidade. São exemplos de escalabilidade horizontal *clusters*; sistemas *blade* de elevada densidade; sistemas massivamente paralelos.
- Escalabilidade Vertical (também conhecida por *scale up/down*) – Consiste numa arquitetura mais convencional de um único sistema a que são adicionados ou retirados mais CPU(s) (por exemplo com SMP), memória ou periféricos mais rápidos à medida que são necessários.

Uma arquitetura que escala horizontalmente apresenta mais versatilidade pois os investimentos podem ser realizados gradualmente, à medida que um sistema necessita de escalar já que, no caso de uma arquitetura com *scale up*, um grande investimento em equipamento é necessário efetuar inicialmente (Hill, 1990).

Com o crescimento da utilização do *cloud computing*, seja em nuvens públicas, privadas ou híbridas, também a escalabilidade ganhou uma adaptação. Assim, uma arquitetura escala verticalmente quando se adicionam recursos à máquina virtual ou contentor (por exemplo, mais cpu ou mais memória disponível) e escala horizontalmente quando se acrescentam ou máquinas virtuais ou contentores de acordo com as necessidades do sistema (Liu, Shie, Lee, Lin, & Lai, 2014).

Acresce que, a virtualização dos ambientes, ao ser intangível, permite que a escalabilidade dos sistemas possa ocorrer de forma automatizada. Concretamente, sobre um sistema que funcione em máquinas virtuais ou contentores pode ser executado um programa que avalie a sua resposta perante a carga num momento, e, perante os resultados modifique os recursos disponíveis para a sua execução, sejam estas modificações efetuadas sobre uma única VM (*scale up*) ou através da criação e execução ou remoção de máquinas virtuais adicionais (*scale out/in*). A estas características, de adaptação do(s) sistemas consoante a carga dá-se o nome de *auto-scaling* (Joy, 2015).

A escolha da arquitetura de escalabilidade tem influência na arquitetura da implementação do software, como veremos nos pontos que se seguem.

### 2.3.1.2 Sistemas baseados em Micro Serviços

Uma vez que a investigação permitiu compreender que as características de um sistema de gestão de Documentos de Arquivo implica a implementação de um sistema versátil quanto à sua escalabilidade, e que esta obrigava à existência de especiais preocupações no que se refere à arquitetura do software, foi efetuada uma investigação acerca de qual o seu estado da arte para aferir a sua aplicabilidade na solução estudada nesta dissertação.

Nas últimas décadas, o paradigma de criação de sistema de informação tem sido caracterizado pelas seguintes tendências:

- Distribuição dos programas em vários processos e(ou) máquinas físicas ou virtuais;
- Modularização dos sistemas em blocos funcionais;
- Baixa dependência (*Loose Coupling*) garantindo que módulos conseguem ativar-se ou desativar-se de acordo com as necessidades

Estas evoluções aumentam a probabilidade de reutilização do código e também a sua robustez, derivando numa aumento global da sua qualidade (Dragoni et al., 2018).

Algumas aproximações, como o SOA – *Service Oriented Architectures* foram passos firmes nas direções de distribuição, modularização e baixa dependência, na medida em que trouxeram a ideia de serviço como uma entidade de software que interage com outras entidades de software através da troca de mensagens utilizando formatos (ex: XML), protocolos (ex: SOAP, HTTP) e interface bem definidos.

Os micro-serviços podem ser considerados como uma nova evolução neste sentido, passando as arquiteturas orientadas aos serviços do âmbito da Integração de Sistemas para o seu Desenho, Desenvolvimento e *Deployment*.

O termo “micro-serviços” começou a ser usado em Maio de 2011 na sequência de de um seminário de Arquitetos de Software que se focou em ideias comuns de Padrões de Arquiteturas de Software (Fowler, 2014). Esta ocorrência, apenas veio atribuir um nome a uma aproximação já em utilização pratica com nomes variados. A *Netflix*, por exemplo, usava já uma arquitetura muito similar com o nome *fine-grained SOA* (Netflix Technology, 2013).

Os Micro-serviços são agora uma tendência consolidada na arquitetura de software focando-se no desenho e desenvolvimento de software que garanta grande facilidade de manutenção e um elevado grau de escalabilidade, baseando-se nos seguintes princípios:

- *Bounded Context*, que significa que funcionalidades relacionadas são combinadas numa única funcionalidade de negócio, sendo esta implementada por um micro-

serviço. Desta forma mantém-se grande alinhamento entre as funcionalidades do negócio e a estrutura do sistema;

- Tamanho. Nos micro-serviços a pequena dimensão de cada um é uma característica base (ao contrário do que se passava no SOA). Sugere-se que, quando um serviço é demasiado grande, deve ser refinado em dois ou mais novos serviços;
- Independência. Encoraja a baixa dependência definindo que cada micro-serviço. deve ser operacionalmente independente.

Por oposição aos micro-serviços, chama-se “sistema de software monolítico”, a um sistema que é construído numa única unidade de execução (Garriga, 2018).

### **2.3.1.3 Integração**

Os sistemas informáticos só em casos muito especiais existem isolados. Por exemplo um software de vendas tem de interligar a uma aplicação de inventário (Fowler, 2002).

No âmbito investigação efetuada nesta dissertação compreendeu-se que um Sistema de Gestão de Documentos de Arquivo é normalmente constituído por vários módulos. Em concreto, as melhores soluções implementadas sob a forma de Software Livre ou de Código Aberto, existem sob a forma de sistemas quase autónomos que, para conseguir apresentar uma solução para a questão de investigação, terão de saber cooperar e integrar entre elas.

As principais soluções de integração tipicamente têm de lidar com os seguintes desafios fundamentais:

- As redes são pouco fiáveis;
- As redes são lentas;
- As aplicações Informáticas são diferentes;
- A mudança/evolução do software é inevitável.

Ao longo do tempo, os engenheiros de software resolveram estes problemas através de várias abordagens:

- Transferência de Ficheiros: Uma aplicação escreve e a outra lê ficheiros com nomes e localizações acordadas;
- Bases de Dados partilhadas: As várias aplicações partilham o esquema de base de dados;

- Execução Remota de Procedimentos ou Funções: Uma aplicação expõe algum da sua funcionalidade de forma a poder ser executada por outra, usando recursos de execução remota. A comunicação ocorre de forma síncrona;
- Envio de Mensagens: Uma aplicação publica mensagens num canal. A outra aplicação lê as mensagens que chegam ao canal, assim que a ele se conseguir ligar. As duas aplicações acordam no formato da mensagem. A comunicação ocorre de forma assíncrona.

Todas estas quatro aproximações, embora resolvam o mesmo problema de integração, têm vantagens e desvantagens. Concretamente, as aplicações normalmente integram usando o método que melhor se adequa a uma integração concreta (Hohpe & Woolf, 2004).

É comum, a utilização de peças de software, chamadas *Message Oriented Middleware* que implementam múltiplos protocolos de comunicação, alguns standards, normalmente permitindo mais do que uma das abordagens de comunicação atrás definidas (Hohpe & Woolf, 2003).

O problema da integração entre aplicações é especialmente relevante no caso de uma arquitetura de micro-serviços, dado que esta se baseia essencialmente na comunicação entre componentes. Apesar de existir a percepção de que nestas arquiteturas, os serviços comunicam exclusivamente usando o padrão REST, não só a sua definição o obriga como vários autores defendem a utilização de outros mecanismos (Xiao, Wijegunaratne, & Qiang, 2016), especialmente de filas de mensagens com o padrão *Publish/Subscribe* e ainda as implementações concretas de grandes sistemas baseados em micro-serviços usa com frequência ligações ponto a ponto com REST mas também sistemas de transferência de ficheiros e(ou) de envio de mensagens (Shadija, Rezai, & Hill, 2017).

### **2.3.2 Tecnologia de Base e *Commodities***

Nos sistemas informáticos, a exemplo do que tem vindo a suceder com produtos de outras atividades que não as Tecnologias de Informação, tem vindo a tornar-se possível uma relativamente fácil troca de algumas componentes, sem que tal implique a completa modificação de um software que neles se suporte. Por exemplo, com fornecimento de eletricidade tornou-se possível a mudança de operador que a fornece, pois todos os dispositivos que a ela se ligam continuam a funcionar corretamente. Analogamente no caso das Bases de Dados Relacionais, é atualmente possível substituir a sua utilização num sistema

por outra implementação diferente. Se bem que possa não ser tão simples como a mudança de operador de fornecimento de eletricidade, é possível com relativa facilidade se existirem alguns cuidados no desenvolvimento do software (Davenport, 2005). Esta *commoditization* de algum software, em que a marca deixa de ser o mais importante valor de um produto, passando a ser as funcionalidades e compatibilidade com o produto incumbente tem sido especialmente bem “aproveitada” pelos *Softwares* desenvolvidos com Licenças Livres ou de Código Aberto (DiBona, Stone, & Cooper, 2005).

Por estas razões, grande parte do software de base, transformou-se em *commodities*, permitindo cada vez mais que a sua troca seja efetuada com menor esforço.

Exemplos claros de programas (ou conjuntos de programas) que se podem atualmente considerar *commodities* são: sistemas operativos, bases de dados relacionais, máquinas virtuais, plataformas de funcionamento aplicacional (tipicamente servidores aplicacionais).

É especialmente no software horizontal de infraestrutura, aquele que serve de suporte ao funcionamento de uma grande quantidade de outros sistemas mais especializados (Fitzgerald, 2006), que o *Software Livre* e de Código Aberto tem conseguido maior penetração.

Os sistemas de Gestão Documental suportam-se claramente em várias categorias de programas de infraestrutura: Sistema Operativo; Sistema de Gestão de Base de Dados; Servidor de Indexação; Plataforma de Desenvolvimento e Execução; Servidor Aplicacional, existindo em todas estas áreas grande oferta de sistemas com licença de *Software Livre* e de Código Aberto.

As plataformas (conjunto de programas de infraestrutura horizontal) onde o *Software* de Gestão Documental pode ser executado tornam-se especialmente importantes pois cada vez é mais referido nos normativos a importância de implementar sistemas dos quais, no futuro, seja possível migrar, para novos sistemas que entretanto venham a ser desenvolvidos (DLM Forum, 2010). Ao usar componentes Livres ou de Código Aberto como infraestrutura horizontal um sistema está também a melhorar a sua conformidade, pois os sistemas livres em última instância, ao ter o seu código disponível, permitem maior abertura por parte dos sistemas neles construídos.

### **2.3.3 Software aplicado à Gestão de Documentos e de Arquivos**

O trabalho de investigação efetuado, relativo aos conceitos e normativos específicos da Gestão de Documentos de Arquivo, permitiu compreender que em geral existem

implementações para diversas áreas da Gestão de Documentos de Arquivo, que implementam a resolução de problemas normalmente de uma forma parcelar. Existem programas que se focam no problema do Armazenamento e descrição genérica com meta-dados, expansíveis ou não (na medida em que permitem criar modelos adicionais de meta-dados para além daqueles que disponibilizam de base), outros que resolvem o problema dos Documentos de Arquivo Histórico, com ou sem existência digital, outros que se viram para a questão da preservação dos objetos digitais, alguns que, apesar de não haverem consensos normativos, se focam na digitalização dos ativos tangíveis, mas também de ferramentas de âmbito mais genérico que podem ser integradas por forma a resolver problemas cada vez mais importantes nesta área, são exemplos os programas de OCR – Reconhecimento de Caracteres Óticos ou de Visão Computacional<sup>1</sup> (Huang, 1996), que poderão, de forma com maior ou menos grau de automatização, permitir melhorar significativamente a classificação dos documentos e o preenchimento dos meta-dados de indexação.

Para o problema concreto da Gestão de Documentos de Arquivo existem vários tipos de software (Tyrväinen, Päivärinta, Salminen, & Iivari, 2006). As tipologias nem sempre são consistentes com as normas e os standards. Acontece ao nível dos sistemas informáticos, o mesmo que se encontra nos normativos: uma grande fragmentação e dificuldade de balizagem das funcionalidades por tipologia. Não obstante, são dignos de referência os seguintes tipos de sistemas informáticos:

- ECM – Enterprise Content Management;
- RM – Records Management;
- Archival Description System;
- Digital Preservation.

---

<sup>1</sup> Campo interdisciplinar que gere a forma como os computadores podem ser usados para obter compreensão de alto nível de imagens digitais ou vídeos

### 2.3.3.1 Sistema de Gestão Documental

Já existem há décadas implementações de Software de Gestão de Documentos de Arquivo. Alguns fabricantes optaram por criar sistemas muito abrangentes, que se posicionavam em todas as vidas (de acordo com a teoria das três idades do arquivo) deste o arquivo corrente até à fase de descrição arquivística. De entre as várias denominações, sobressaiu no mercado, a designação *Enterprise Content Management*. Infelizmente, trata-se de uma designação que não descreve estes sistemas de uma forma muito precisa, existindo mesmo alguma confusão acerca da definição do âmbito e funções de um sistema de ECM (Kaplan, 2002).

A abrangência do termo leva à adoção da seguinte definição (Smith & McKeen, 2003):

“Estratégias, ferramentas, processos e competências de que uma organização necessita para gerir os seus ativos de informação (independentemente do tipo) ao longo do seu ciclo de vida”.

Destes sistemas, alguns são mais ou menos abertos, no que se refere não só à forma como armazenam e indexam a informação, facilitando ou não o acesso ao conteúdo, mas também à possibilidade da sua expansão, e também utilitários de exportação e importação dos ativos que gerem.

A Gestão Documental, termo aceite em Portugal para a designação Anglo Saxónica *Records Management* (Alberts et al., 2010), é uma área que pode ser vista como um módulo da designação abrangente dos sistemas ECM. De facto, grande parte dos sistemas de ECM, contêm mecanismos que são geralmente conhecidos como de Gestão de Documental (Ribeiro, 2018).

Um sistema de Gestão Documental, implementa funcionalidades específicas, nomeadamente (Dikopoulou & Mihiotis, 2012):

- A gestão dos próprios registos que a organização cria e mantém;
- Um plano de classificação para os registos;
- Um sistema de classificação de acessos (categorias de utilizadores, termos e condições para aceder e utilizar os registos);
- Os meta-dados para descrição e pesquisa;
- Um calendário de destruição, conservação ou arquivo para todos os registos;
- Procedimentos e políticas para gestão dos registos;
- Um programa de formação para as partes interessadas dentro da organização;

- Um programa de auditoria para avaliação dos procedimentos de Gestão de Registos;
- Uma associação entre a classificação e o calendário de conservação/destruição.

A gestão dos registos consiste nos mecanismos que permitem a inserção, obtenção, modificação e manutenção não só da informação relativa aos registos, mas também em relação ao objeto registado, caso este possa ter uma representação digital.

Por plano de classificação entende-se um sistema de taxonomia, normalmente hierárquico e aberto – que pode ser alterado e configurado por utilizadores com permissões específicas – a que cada registo fica associado.

Os meta-dados, palavra normalmente definida como dados acerca dos dados, são definições de esquemas (tipificação de objetos) que descrevem um item concreto de dados (Garshol, 2004). Um bom sistema de ECM implementa mecanismos de meta-dados que são expansíveis. Tal significa que utilizadores com permissões para tal, poderão definir especificamente para a área de negócio da organização, os conjuntos de informação que descrevem os seus registos e(ou) objetos digitais.

Em resumo, um sistema de gestão documental permite aos utilizadores inserir, descrever e aceder aos registos e documentos com a devida segurança. Existem implementações fechadas e proprietárias, bem como implementações de *software* livre e de código aberto.

### **2.3.3.2 Descrição Arquivística**

Com o surgimento de normativos específicos para a descrição arquivística como a ISAD(G), foram desenvolvidos sistemas específicos, endereçando o problema específico da descrição arquivística de forma separada e muitas vezes autónoma da Gestão Documental (arquivo corrente e intermédio) (Ribeiro, 2018)

No mundo do software proprietário existe um grande número de implementações de software para resolver este problema, no entanto foram também desenvolvidos sistemas específicos com licenciamento de *Software* Livre e de Código Aberto. De entre as várias opções, destaca-se o software AtoM, dado que foi desenvolvido em estreita colaboração com o Conselho Internacional dos Arquivos e com a Unesco (D. Flores & Hedlund, 2014).

Um Sistema de descrição Arquivística, em primeira instância permite a criação da taxonomia definida pela norma de descrição arquivística, normalmente a ISAD(G), nomeadamente a existência de fundos, séries, sub-fundos e (ou) sub-séries e os conjuntos de meta-dados

definidos na norma, mas também normalmente implementam os meta-dados e ligações inerentes à estrutura da norma ISAAR(CPF), para autoridades.

### **2.3.3.3 Preservação Digital**

A investigação efetuada permitiu perceber a existência estudos e normativos que se preocupam com a questão da preservação a longo termo dos conteúdos digitais. Era por isso expectável a existência de programas capazes de ajudar na execução de tal tarefa. Existem ferramentas que o fazem de uma forma bastante estruturada, mas também se percebe que muitas organizações usem essencialmente utilitários de âmbito limitado (por exemplo ferramentas de conversão de ficheiros com som ou imagem em movimento) integrando-os através de métodos próprios mais ou menos alinhados com o modelo de referência OAIS.

A pesquisa efetuada levou-nos a concluir que existe um grande número de implementações, notando-se algum alinhamento com o modelo de referência mas estando ainda longe de existir normalização suficiente para que estes produtos possam facilmente ser comutados entre eles (Madalli et al., 2012). Nota-se sobretudo a existência de soluções implementadas sobretudo ao nível das organizações que em cada país são responsáveis pela gestão e normalização dos arquivos.

Em Portugal, a Direção Geral do Livro, Arquivos e Bibliotecas participou ativamente na definição e implementação do Repositório de Objetos Digitais Autênticos. O RODA é “um projecto lançado pelo Instituto dos Arquivos nacionais/Torre do Tombo que pretende abordar de forma sistemática estas questões no intuito de vir a colmatar um vazio atualmente existente relativamente à gestão continuada de objetos digitais” (Barbedo et al., 2007, p. 1).

### 2.3.4 Pesquisa

Com o aumento da quantidade de informação armazenada ou descrita, a pesquisa tornou-se uma área fundamental para permitir a descoberta da informação relevante para os utilizadores e sistemas. Os sistemas de informação estruturados historicamente implementam formas de pesquisa diretamente obtendo a informação nas fontes específicas, o que significa que como normalmente a informação está estruturada de acordo com o modelo relacional, serão efetuadas pesquisas às colunas de tabelas específicas obtendo e devolvendo as coleções aos utilizadores. Estas pesquisas são normalmente implementadas sob a forma de relatórios aos quais se enviam parâmetros para as pesquisas devolvendo a informação que foi solicitada especificamente (Goldman, 2008).

A evolução da WWW começou por incidir sobre informação não estruturada. Desta forma, os sistemas de pesquisa efetuam indexações sobre estes conteúdos, permitindo depois que se realizem pesquisas por palavras e(ou) expressões. Motores de pesquisa como o Google («Google», 2019), colecionam em bases de dados enormes quantidades de informação sobre as quais são efetuadas pesquisas que devolvem a informação de forma extremamente rápida. Posteriores evoluções nestes motores de pesquisa vieram permitir pesquisas não apenas em documentos não estruturados, mas também em bases de dados relacionais ou outras.

Apesar de os sistemas de indexação públicos serem muito eficazes, evoluídos tecnologicamente e a indexação ser efetuada sem custos diretos, coloca as instituições num dilema de *vendor lock-in*, provavelmente ainda mais grave do que no caso do software, já que a sua informação pode também ser usada com efeitos comerciais e(ou) colocada fora das jurisdições locais onde devem, muitas vezes por lei, existir (Kulathuramaiyer & Balke, 2006).

Desta forma, a existência de mecanismos de indexação e pesquisa são funcionalidades fundamentais num Sistema de Gestão Documental. Acontece que, pelas razões já atrás descritas, cada vez mais se assiste a uma fragmentação e especialização dos sistemas de informação, cada qual contendo o seu sistema de pesquisa. Esta realidade, acrescido da necessidade de prover os utilizadores (ou sistemas) com um único ponto para efetuar pesquisas já que não seria aceitável para nenhum utilizador ter de repetir a mesma pesquisa nos vários sistemas da sua organização.

### 2.3.4.1 Pesquisa Federada vs Pesquisa Centralizada

Os chamados motores de pesquisa modernos, efetuam as consultas dos utilizadores sobre conjuntos de dados, chamados índices, organizados de forma específica para permitir executar as pesquisas de forma eficaz.

Os índices podem ser centralizados num sistema único da organização ou federados, que significa que, quando o utilizador solicita uma consulta, esta é efetuada sobre um conjunto de repositórios que se encontram registados num diretório que implementa os protocolos de pesquisa sobre todos eles (Shokouhi & Si, 2011)

No âmbito da área documental, o Protocolo Federado OAI:PMH, que se descreve no próximo ponto é bastante aceite pelas comunidades.

### 2.3.4.2 OAI: PMH

O OAI:PMH, *Protocol for Metadata Harvesting* é um protocolo publicado pela *Open Archives Initiative* («Open Archives Initiative», 2019), organização cuja missão é promover e aplicar standards de interoperabilidade técnica entre arquivos, por forma a partilhar meta-informação entre eles.

O objetivo do *OAI:PMH* é coletar descrições de meta-dados de arquivos, ficando estes disponíveis em repositórios sobre o qual se efetuam as pesquisas, disponibilizando não só os resultados como também o repositório onde a informação é disponibilizada. É um protocolo que utiliza XML transferido por HTTP.

O OAI:PMH é um protocolo usado por sistemas que suportam serviços importantes de que são exemplos: Open Journal Systems («Open Journal Systems», 2019); Dspace («DSpace», 2019); Pangaea («Data Publisher for Earth & Environmental Science», 2019).

No OAI:PMH existem dois tipos de sistemas intervenientes:

- O Data Provider, que acede ao repositório e que disponibiliza informação para ser recolhida pelo coletor;
- O Service Provider, opera os coletores sobre os vários repositórios e indexa a informação recolhida.

Os principais conceitos do OAI:PMH são:

- Coletor – executado pelo *Service Provider*, recolhe informação dos *Data Providers*;

- Repositório – Sistema onde a informação é arquivada e indexada sendo disponibilizada por um Data Provider que a ele está ligado;
- Registo - Conjunto de meta-dados coletados de um *Data Provider*.

Para a utilização do OAI:PMH como protocolo de transferência entre os repositórios, um indexador e servidor de pesquisa é fundamental para conseguir criar um sistema que use as melhores características de cada um dos outros sub-sistemas (Van de Sompel, Nelson, Lagoze, & Warner, 2004).

### **2.3.5 Optical Character Recognition - OCR**

Grande parte dos documentos geridos em Sistemas de Gestão de documentos, ou não nasceram de forma digital, ou foram impressos em papel e assim transferidos entre as organizações. Nesse caso, muitas vezes, as organizações procedem à sua digitalização em equipamentos próprios, que, na maioria dos casos fazem-no para formatos de imagens em que o texto não existe como tal, mas sim representado na imagem sob a forma de pontos (*pixels*). Desta forma, torna-se impossível efetuar pesquisas textuais.

Quando um documento é ingerido por um sistema de Gestão Documental, por forma a ser permitida a sua pesquisa por texto, o seu conteúdo e os seus meta-dados são indexados pelo sistema, sendo as pesquisas efetuadas através da utilização destes índices.

Desta forma, documentos em que o texto esteja representado sob a forma de imagem são mais dificilmente localizáveis já que o seu texto não é indexado.

Existe *software* que permite converter as letras contidas nas imagens em representação do carácter que a letra representa, tornando-se assim palavras e textos. Formatos como o PDF permitem mostrar a página digitalizada, tal como ela existe na imagem, mas adicionando uma camada que contém o texto nela contido (normalmente sem informação acerca do tipo de letra usado) (Mori, Suen, & Yamamoto, 1992).

### **2.3.6 Transformação**

Tanto os sistemas de Gestão Documental como as implementações de Preservação Digital, e até a aquisição dos conteúdos, necessitam de efetuar transformações sobre os objetos digitais que armazenam.

A existência de um grande número de ferramentas de conversão entre formatos, cada uma especializada num tipo de conteúdo (por exemplo: imagens, documentos de texto e(ou) vídeo ou som), faz com esta tarefa seja sobretudo uma tarefa de integração. Alguns produtos de Gestão Documental integram as transformações no seu cerne, eventualmente usando as ferramentas externas.

No entanto, este tipo de tarefa pode não só necessitar de utilizar muitos recursos computacionais, como o facto de se poder usar várias ferramentas e estas dependerem de versões específicas tanto do Sistema Operativo como de outras ferramentas, fazem desta tarefa um candidato a arquiteturas distribuídas e elásticas. De facto podem ocorrer momentos em que as necessidades de processamento sejam tão elevadas que o único servidor local não seja suficiente para conseguir executar esta tarefa, podendo não se justificar o seu *upgrade* pois esses momentos de maior processamento podem ocorrer apenas em temporariamente (por exemplo após uma migração de documentos ou se se definir converter os ficheiros para uma nova versão de um formato).

Assim, seria benéfico a existência clara na arquitetura de um servidor de transformação, permitindo a sua distribuição dinâmica em várias máquinas físicas ou virtuais, ou apenas em contentores, bem como a especialização de conversores (por exemplo a existência de um nó apenas para conversões de vídeo).

Infelizmente, este tipo de componente externo não está claramente definido e passível de ser usado pelos sistemas de gestão documental, como por exemplo se de um sistema de gestão de bases de dados se tratasse, sendo escolhida uma implementação monolítica e interna ao sistema de Gestão Documental (Shallcross, 2016).

## **2.4 Software Livre e de Código Aberto**

É comum assumir que o conceito de Software Livre e de Código Aberto como o conhecemos atualmente, tem a sua génese no aparecimento da licença GPL – GNU Public Licence, em 1989, por parte do projeto GNU (que entretanto evoluiu para uma fundação) («Free Software Foundation», 2019), liderado por Richard Stallman. No entanto, de uma forma menos estruturada, já antes existiram tendências de partilha de algoritmos, conceitos e código. Entre os anos 50 e 80, não só as comunidades científicas desenvolveram a computação, tipicamente permitindo o livre acesso às ideias, algoritmos e implementações, como as leis anti-monopólio derivaram na partilha de programas como o Unix (que, devido às limitações à

AT&T foi partilhado com as principais universidades). Nessa época, as grandes empresas centravam as suas vendas no hardware e nos serviços. Durante os anos 70, com o advento da micro-computação, organizações como o *Homebrew Computer Club* promoviam largamente a partilha, mas já se notavam grandes movimentações do conceito de software proprietário com propriedade intelectual licenciada. A famosa carta de Bill Gates à comunidade do clube é exemplo deste movimento de encerramento e redução de partilha (Söderberg, 2015).

Em 1989, com a adesão dos Estados Unidos à Convenção de Berna, ocorre uma mudança radical: Até 1 de Março de 1989, nos EUA, um trabalho não gravado num bem tangível, para ser protegido deveria ter associado uma notícia de copyright, a partir desse momento, o trabalho passou a estar protegido salvo a existência de uma licença em contrário.

Todas estas ocorrências derivaram no final dos anos 70 do século passado, num enorme crescimento de um mercado de Software cuja utilização era licenciada mediante pagamentos (Andrew Laurent, 2004).

Richard Stallman, um programador do laboratório de Inteligência Artificial do MIT («MIT CSAIL», 2019), habituado ao ambiente de partilha existente até ao final dos anos 70, começou a sentir as novas vicissitudes de trabalhar com software licenciado ao tentar modificar um *driver* de uma impressora e ao trabalhar com sistemas operativos que nesse momento já eram disponibilizados com uma licença proprietária. Este facto levou a que, em 1984 se demitisse do MIT para fundar o movimento GNU («gnu.org», 2019) cujo objetivo era a criação de um Sistema Operativo como o Unix, mas que não era Unix (daí o acrónimo recursivo GNU: GNU is Not Unix) e que poderia ser utilizado livremente. A primeira dificuldade em que esbarraram foi a necessidade de especificar que as suas criações poderiam ser utilizadas por todos e em vez de proteger a propriedade intelectual, ou seja: como conseguiriam que as suas criações não fossem protegidas. Para tal teriam de escrever um texto, que, em linguagem legal, garantisse a sua manutenção como criação livre (Stallman, 1998). A escrita deste texto criou a GPL – GNU General Public Licence e o conceito de copyleft, na qual se define exatamente o conceito de software livre e os termos da sua disponibilização. É neste texto que se definem as principais características do software livre, as quatro liberdades (Free Software Foundation, 1989):

- Liberdade de executar o programa para quaisquer propósitos;
- Liberdade de estudar como o programa trabalha e adaptá-lo às suas necessidades. O acesso ao código fonte é uma pré-condição;
- Liberdade de distribuir cópias para ajudar os outros;

- Liberdade de distribuir a outros cópias das versões modificadas. Desta forma, toda a comunidade beneficia das modificações. Acesso ao código fonte é uma pré-condição.

Para enquadrar os desenvolvimentos e o licenciamento, Stallman criou a *Free Software Foundation*. Por forma a conseguir implementar um novo Sistema Operativo, a FSF começou por desenvolver as ferramentas de base como compiladores da linguagem C, compiladores de compiladores (Flex e Bison) ou editores de texto entre outros. Em 1990 as principais componentes para a criação do Sistema Operativo GNU estavam terminadas, com exceção do núcleo, já que a FSF, em vez de desenvolver o núcleo com os mesmos princípios do *kernel* do Unix, tentou inovar e criar um sistema operativo com o conceito de múltiplos pequenos núcleos executados no âmbito de um microkernel. Esta abordagem veio a demonstrar-se um erro já que nunca foi possível uma versão estável ou mesmo funcional (Walfield & Brinkmann, 2007).

Em 1991, Linus Torvalds, estudante na Universidade de Helsínquia, desenvolveu um *kernel* similar ao Unix, com base em ideias no *kernel* Minix («Minix3», 2019), um sistema operativo muito limitado com objetivos didáticos, similar ao Unix desenvolvido pelo professor Andrew Tanenbaum. A disponibilização do código do Linux com licença GPL, bem como a abordagem tradicional, rapidamente trouxe programadores ao projeto, tornando-se uma alternativa viável em poucos anos (Love, 2010).

Em 1997, Eric Raymond, programador do projeto *fetchmail* escreveu o livro *A Cathedral e o Bazar* (Raymond, 1998) que se foca na comparação entre os métodos tradicionais de desenvolvimento de software e a forma de desenvolver na comunidade de código partilhado. É com base no sucesso do livro que funda, em 1982, com Bruce Perens, a organização *Open Source Initiative*, que tenta separar a forte componente ideológica do movimento do software livre, dos resultados obtidos no desenvolvimento em que o código é partilhado, a que chamou Software de Código Aberto. A discordância entre os dois movimentos é pública: O Movimento do Software Livre foca-se na ética e liberdade enquanto que o movimento do Open Source tem preocupações de índole mais prática defendendo basicamente o modelo de desenvolvimento em rede com partilha dos programas fonte. Este movimento, liderado pela OSI vai mesmo mais longe, defendendo que o Software de Código Aberto não deve ser imposto através de licenciamento viral (como na GPL), que obriga a que todo o software derivado seja também livre, mas que se impõe por ser um modelo que tanto em resultados como economicamente é mais viável. Em termos ideológicos há quem defenda que a visão do movimento do Software Livre se aproxima de uma visão mais socialista e a do *Open Source*

mais do liberalismo e do funcionamento do mercado (Söderberg, 2015). Na verdade, o desenvolvimento com um modelo de Código Aberto, seja com uma licença de Software Livre ou uma licença mais liberal, demonstrou não só ser viável, como capaz de gerar sistemas de uma imensa qualidade. Prova disso é o facto de sistemas desenvolvidos com este modelo serem a base do funcionamento da esmagadora maioria das infraestruturas mundiais, como a *cloud computing*, o *IOT*, os super computadores, os telemóveis, os servidores empresariais (Crowston, Wei, Howison, & Wiggins, 2008) O software desenvolvido em modelo Open Source, seja ou não protegido com uma licença de Software Livre expandiu-se não apenas como software de infraestrutura, mas existindo implementações para todas as áreas de negócio, desde os ERP(s), CRM ou Gestão de Documentos (Morgan & Finnegan, 2014).

### 2.4.1 Metodologia de Avaliação de Software Open Source

A escolha de produtos de Software Livre e de Código Aberto pode tornar-se especialmente difícil. Porque existe atualmente muita escolha e porque é difícil avaliar de forma clara os vários fatores importantes para o investimento na sua utilização. As funcionalidades que implementam são claramente importantes, no entanto não são suficientes. Existem literalmente milhares e milhares de projetos de Software Livre ou de Código Aberto, que resolvem problemas muito variados mas que não são mantidos por nenhuma equipa ou que não estão implementados em tecnologias atuais. Investir na utilização de um produto pode ser muito consumidor de tempo e dinheiro e pode levar facilmente a caminhos sem saída após um enorme esforço (Umm-e-Laila, Zahoor, Mehboob, & Natha, 2017). Por estas razões, é muito importante seleccionar corretamente o projeto para resolver o nosso problema. Esta seleção é, muito frequentemente realizada através de técnicas ad hoc, no entanto, nos últimos anos, têm emergido metodologias para avaliação e seleção de produtos de Software Livre ou Código Aberto. Estas metodologias têm como objetivo a escolha por entre uma lista de produtos, daquele que melhor se adequa a um contexto. As principais metodologias adequam-se aos objetivos pretendidos, mas não existem estudos que permitam concluir acerca da melhor utilização de acordo com diferentes objetivos (Adewumi, Misra, & Omoregbe, 2013)

**Tabela 16: Metodologias de Avaliação de Software Open Source**

Nome	Descrição	Modelo Origem
QSOS	Quatro estágios iterativos. Suportado por ferramentas. Permite uma avaliação objetiva. Boa comunidade de utilizadores («QSOS», 2018).	ISO/IEC9126

Qualipso/QualOSS	Reduz a subjetividade automatizando as medidas. Site indisponível.	Capgemini Maturity Model
OpenBRR	Aberta e configurável, podendo ser aplicável a qualquer solução de negócio («OSSPAL», 2019).	ISO/IEC 9126
OMM	Escala de três valores. Simples e disponível. Pouco usada	Capability Maturity Model

No âmbito desta dissertação decidiu-se usar a metodologia QSOS («QSOS», 2018) para a seleção dos produtos que se adequam melhor a cumprir o seu papel na arquitetura de referência desenhada. O acrónimo QSOS significa *Qualification and Selection of Open Source software*. A seleção foi baseada na análise de várias publicações, mas pesou muito o facto de esta disponibilizar ferramentas on-line, de poder ser usada mediante um licenciamento GNU Public Licence e de ser bastante prática (Deprez & Alexandre, 2008). O QSOS foi desenvolvido e é suportado pela empresa Francesa Atos para qualificar, selecionar e comparar solução de software disponíveis com Licenças de Software Livre ou de Código Aberto. Permite uma análise estruturada com vista a permitir uma decisão racional baseada em dados objetivos (Deprez & Alexandre, 2008).

O QSOS tem os seguintes objetivos:

- Analisar Objetivamente os vários produto de Software Livre ou Open Source candidatos;
- Que solução cumpre melhor as necessidades funcionais atuais e futuras;
- Que solução melhor integra com a infraestrutura técnica e a arquitetura aplicacional;
- Levar em conta constrangimentos e quantificar os riscos relacionados com o facto de ser uma solução Open Source;
- Clarificar a maturidade da solução, através da perceção da dimensão e dinamismo da comunidade de utilizadores e programadores;
- Analisar os riscos de um *fork*. Definir como antecipar e gerir esse facto, caso venha a acontecer;
- Validar quais os níveis de suporte necessários e disponíveis;
- Analisar que licenciamentos se aplicam;
- Compreender se é possível influenciar o projeto (*roadmap*, desenvolvimento de novas funcionalidades);
- Produzir resultados objetivos, motivados e personalizados.

O QSOS é composto por quatro passos iterativos e interdependentes:

- Definição e organização do que será a base da avaliação, concretamente documento de critérios de avaliação estruturados como árvores, com dois grandes eixos:
  - Maturidade do Projeto, definido de forma global, para qualquer tipo de projeto.
  - Cobertura Funcional, específica para cada área de avaliação.
- Avaliação dos produtos concretos de Software Livre e de Código Aberto, sobre a base avaliação, qualificando cada produto em cada critério.
- Qualificação do contexto de utilização específico através dos pesos dos critérios.
- Seleção das soluções mais relevantes através de comparações baseadas nas valorizações dos critérios.



Figura 5: Metodologia QSOS (Moreau, 2013)

Uma avaliação QSOS é efetuada usando os seguintes instrumentos:

- *Template* contendo os critérios concretos de avaliação. No caso desta dissertação, foi criado um template para cada tipo de sistema avaliado: Gestão Documental, Descrição Arquivística, Preservação Digital, sendo avaliados mediante vários tipos de critérios (maturidade, funcionalidades, arquitetura, entre outros);
- Avaliações de cada produto, baseadas no *template* de cada tipo de sistema avaliado. Por exemplo, no caso da Gestão Documental foram avaliados: Alfresco, LogicalDoc, Nuxeo e OpenKM;
- Comparações entre produtos, através de gráficos de quadrantes, de radar e de pontuações.

Para efetuar uma avaliação com QSOS usam-se as seguintes ferramentas:

- Ferramenta de criação de mapas mentais, Freemind, com a qual se criam os *templates* específicos de cada tipo de software, se obtém uma representação gráfica das valorizações de cada software concreto;

- Editor de valorizações, que usando os *templates* definidos disponibiliza um interface prático para a valorização de cada critério definido no *template*. No final, permite salvar cada avaliação em ficheiro próprio;
- Com o comparador e gerador de resultados é possível carregar as valorizações específicas de cada produto e gerar relatórios comparativos, nomeadamente gráficos de quadrante, de radar, apresentações, textos e tabelas em formatos ODT, ODF e ODC.

## 2.4.2 Economia de Serviços e Dependência de Fabricantes

Verifica-se atualmente uma tendência de mudança na economia. Se anteriormente era frequente a venda de produtos, sendo estes definidos como bens, normalmente tangíveis, cuja propriedade é transferida na transação, entre o fornecedor e o cliente, atualmente nota-se uma tendência em transacionar o resultado final para o consumidor, ou seja o serviço. Esta tendência nota-se em áreas tangíveis como nos transportes em que se nota por exemplo a tendência para transacionar a utilização completa de uma viatura (caso do *renting*) em vez da venda da viatura física ao consumidor que depois a utilizaria com o fim dos serviços. Na área do software, com o sucesso de fenómenos como o do Software Livre ou de Código Aberto, em que a transferência da propriedade ou direito de utilização, deixa de ser significativa, esta tendência está a transformar todo o mercado (Morgan & Finnegan, 2014). Empresas que no passado foram líderes da venda de direitos de utilização de software como produto, transformaram os seus negócios em vendas de serviços. Caso paradigmático é o da Microsoft com o seu principal produto Microsoft Office que é atualmente comercializado sob a forma de serviço, tendo já neste caso, este modelo ultrapassado em receitas o modelo da tradicional venda como produto (Cusumano, 2008).

## 2.5 Metodologia de Descrição de Arquitetura

É objetivo final desta dissertação propor, se possível, uma solução concreta para implementar um sistema para Gestão de Documentos de Arquivo. Para tal, dada a diversidade de soluções, bem como de produtos de suporte, o que implica um elevado grau de complexidade, será desenhado em primeiro lugar uma arquitetura conceptual de referência, selecionados os produtos com a metodologia QSOS atrás descrita, e finalmente desenhada uma arquitetura concreta que incluirá os produtos de infraestrutura e aqueles que concretamente implementam

os requisitos específicos da Gestão de Documentos de Arquivo. Por forma a suportar a descrição da arquitetura, escolheu-se estudar o estado da arte das formalizações para descrição de arquiteturas empresariais, tendo como objetivo descrever tanto a Arquitetura Conceptual de Referência como a Arquitetura Concreta, utilizando formalismos já bem estudados e implementados. A figura que se segue descreve o contexto em que as descrições de arquitetura se inserem («ISO/IEC/IEEE 42010», 2011).

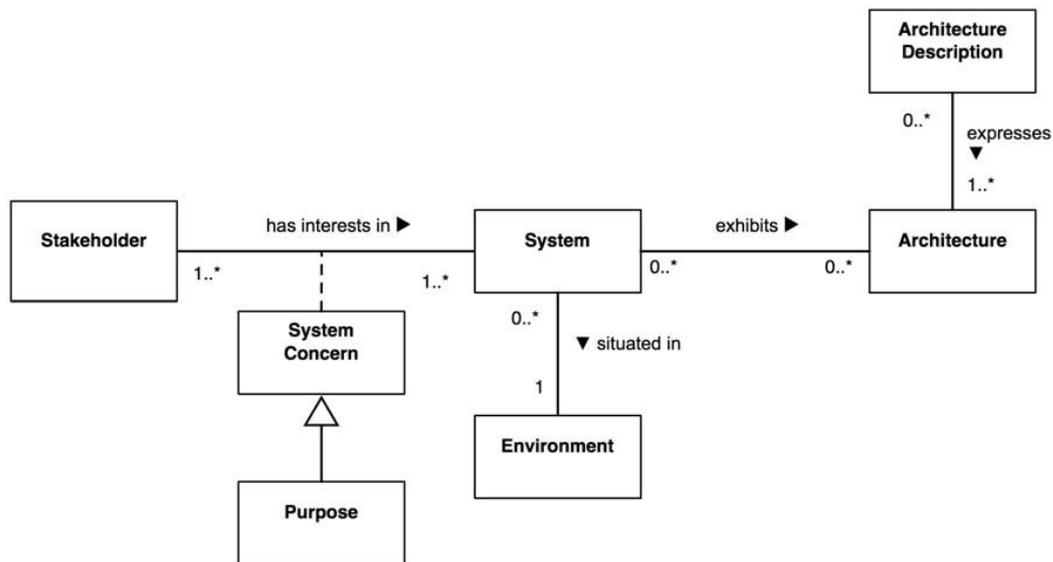


Figura 6: Contexto de Descrições de Arquitetura

Existem várias definições de arquitetura empresarial. Uma definição alinhada com a norma ISO 42010 define arquitetura empresarial como “*fundamental concepts or properties of an enterprise in its environment embodied in its elements, relationships, and in the principles of its design and evolution.*” («ISO/IEC/IEEE 42010», 2011, p.1). Uma arquitetura empresarial é tipicamente descrita utilizando modelos. A descrição da arquitetura empresarial é uma ferramenta útil e necessária para compreender a complexidade e gerir a mudança dos sistemas que descrevem (Hinkelmann et al., 2016).

Derivado à complexidade de uma descrição da arquitetura da empresa, foram desenvolvidas várias frameworks e(ou) Linguagens de Descrição de Arquiteturas (ADLs) com o objetivo de assistir a tarefa de descrição da arquitetura empresarial. A norma ISO 42010 define *framework* para descrição de arquitetura empresarial como “*conventions, principles and practices for the description of architectures established within a specific domain of application and/or community of stakeholders*” («ISO/IEC/IEEE 42010», 2011, p.3) e *Linguagem de Descrição de Arquitetura* como “*any form of expression for use in architecture descriptions*” («ISO/IEC/IEEE 42010», 2011, p. 10)

Existem dezenas de *frameworks* e de ADLs para descrição de arquiteturas de sistemas empresariais. Para efeitos da descrição das Arquiteturas de Referência e Concreta, escolheu-se usar uma Linguagem de Descrição de Arquiteturas (Hussain, 2013). A escolha da ADL (Björnander, 2019) foi efetuada de acordo com os trabalhos científicos que as descrevem e comparam. Finalmente, das ADLs consideradas que se apresentam no quadro que se segue, usa-se a ArchiMate (The Open Group, 2017). As razões prendem-se com o facto de ser considerada uma das mais adequadas para descrever Arquiteturas Empresariais, por ser baseada no UML («Unified Modeling Language Specification», 2017), por ser um standard do *Open Group* e por existirem várias ferramentas, algumas com Licença de Software Livre para desenhar a arquitetura.

**Tabela 17: Linguagens de Descrição de Arquitetura (Hussain, 2013)**

Nome	Descrição
ArchiMate	Desenvolvida pelo Open Group, baseada na norma ISO 1471 (que foi substituída pela norma ISO 42010)
UML	Unified Modeling Language. Usada extensivamente na indústria e academia. Mais apropriada a um nível de detalhe abaixo da arquitetura.
XADL	Linguagem de Meta-modelização implementada sob a forma de XML schemas. Muito usada na indústria e em meios académicos. Desenvolvida pela Universidade da Califórnia.
Wright	Desenvolvido pela Universidade de Carnegie Mellon

O Archimate é uma Linguagem de Descrição de Arquitetura. Foi desenvolvido como parte de um projeto colaborativo, financiado por instituições Holandesas. Foi mais tarde transferido para o Open Group, no qual se tornou no standard de descrição de arquiteturas inserido na *Framework* de Arquitetura do Open Group (TOGAF) (Lankhorst, Proper, & Jonkers, 2009).

Consiste nos seguintes componentes primários:

- Uma Framework conceptual que permite a conceção e classificação de arquiteturas;
- Uma sintaxe abstrata que contém a definição formal da linguagem em termos de meta modelo;
- Conceitos de Modelação que permitem a descrição de aspetos relevantes das empresas. Estruturam a sintaxe abstrata, focando-se no significado de cada conceito;
- Semântica da linguagem que definem o significado de cada artefacto da linguagem e tipos de relações;
- Sintaxe de Notação Visual que definem de que forma os artefactos da linguagem se representam visualmente;

- Mecanismo de Pontos de Vista correspondendo aos diferentes diagramas da linguagem UML. Permitem visualizar a arquitetura de diferentes prismas.

A especificação Archimate refere a existência de uma linguagem *core mais* reduzida, que é estendida sendo disponibilizada sob a forma de uma versão mais completa (*Full Framework*). Para efeitos desta dissertação será utilizada a versão estendida.

A linguagem Archimate estrutura-se em duas dimensões: Camadas e Aspectos. As camadas, que de seguida se enumeram, representam os níveis em que o sistema pode ser modelado:

- *Strategy Layer* – Inclui elementos que suportam a estratégia de modelação, como recursos, capacidades ou plano de ação;
- *Business Layer* – Serviços de negócio oferecidos aos clientes, que podem ser realizados sob a forma de processos de negócio executados pelos atores que interagem com o sistema.;
- *Application layer* – Serviços Aplicacionais que suportam o negócio e as aplicações informáticas que implementam os serviços;
- *Technology Layer* – Serviços tecnológicos (ex: armazenamento, comunicações), necessários para executar as aplicações;
- *Physical Layer* – Inclui a definição de equipamentos, redes e outros materiais;
- *Implementation and Migration Layer* – Inclui o suporte à implementação e técnicas para a migração de arquiteturas, como oportunidades, soluções, planos de migração ou governança da implementação.

Os aspectos definidos pelo Archimate são:

- *Active Structure Aspect* – Representa os elementos estruturais, por exemplo os atores, componentes aplicacionais ou dispositivos, isto é “os sujeitos da atividade”;
- *Behaviour Aspect* – Representa o comportamento, como processos, funções, eventos e serviços que são executados pelos atores;
- *Passive Structure Aspect* – Representam os objetos em que os comportamentos são executados. São normalmente os objetos de informação na camada de negócios, ou os objetos de dados na camada aplicacional.
- *Motivational Aspect* – Permite aos diferentes *stakeholders* descrever a motivação de atores ou domínios específicos.

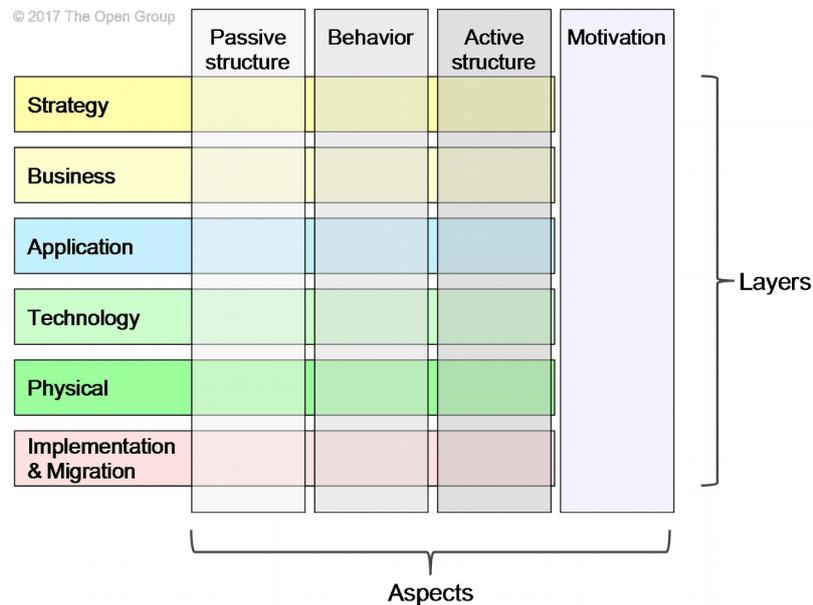


Figura 7: Representação das Dimensões do Archimate (The Open Group, 2017)

A estrutura da *framework* permite a modelação do sistema a partir de diferentes pontos de vista. Desta forma cada célula refere-se a diferentes preocupações dos *stakeholders*.

A notação gráfica do Archimate (The Open Group, 2017) foi construída de forma a ser familiar aos arquitetos habituados a técnicas mais antigas como o ERD (Chen, 1976), UML («Unified Modeling Language Specification», 2017) ou BPMN («BPMN Specification», 2011). A notação standard para a maioria dos elementos consiste num retângulo contendo um icon no canto superior direito. Mesmo se o Archimate não é completamente rígido na utilização da simbologia standard, esta deve ser preferida sempre que possível.

O meta-modelo do standard também define que devem ser usadas diferentes sombras de cinzento para distinguir elementos que pertencem a diferentes aspetos:

- Branco para conceitos abstratos;
- Cinzento claro para estruturas passivas;
- Cinzento médio para comportamentos;
- Cinzento escuro para estruturas ativas.

O standard não define obrigatoriedade de significado das cores dos objetos visuais, no entanto as cores são normalmente usadas para distinguir as camadas:

- Amarelo para a camada de negócio;
- Azul para a camada aplicacional;
- Verde para a camada tecnológica.

Adicionalmente, uma letra no canto superior esquerdo identifica a camada a que um objeto pertence: M-Motivation, S-Strategy, B-Business, A-Application, T-Technology, P-Physical.

Também a estrutura dos cantos de cada símbolo tem um significado:

- Cantos a noventa graus representam elementos de estrutura;
- Cantos Arredondados representam elementos comportamentais;
- Cantos em diagonal representam elementos motivacionais.

## 2.6 Síntese

A Gestão de Documentos de Arquivo é uma atividade teorizada nos últimos séculos. Desde o final do século XX foram definidas um grande conjunto de normas e standards. Em muitos casos normaliza-se em relação aos mesmos assuntos, mas é clara a divisão entre as áreas da Gestão Documental, que inclui os arquivos correntes e intermédios, a Descrição Arquivística para o caso dos arquivos históricos ou definitivos e da Preservação Digital que se foca na problemática relacionada com a possibilidade de acesso e utilização futura aos ativos digitais correntes.

O Software Livre e de Código Aberto demonstrou ser uma alternativa viável para a implementação de programas não só de infraestrutura mas também de utilização específica às atividades concretas.

As arquiteturas mais modernas, que incluem conceitos fundamentais que permitem o correto funcionamento dos programas variando com a escala de utilização, concretamente quantidade de utilizadores ou massa documental, estão teorizadas e estudadas atingindo um bom grau de maturidade. A descrição das arquiteturas pode ser efetuada com recurso a metodologias estandardizadas. Suportado nestes métodos e respetivas técnicas melhora-se a abordagem da definição de uma arquitetura conceptual ou concreta.

Seja software de infraestrutura ou específico para a área da Gestão Documental, a tarefa da seleção da melhor solução não é trivial, no entanto existe trabalho publicado que definem várias técnicas sistemáticas que ajudam a escolher os programas ou conjunto de programas de Código Aberto que melhor resolvem os problemas estudados.

O estado da arte da Gestão de Documentos de Arquivo, do Software Livre ou de Código Aberto, da sua seleção, da descrição formal de arquiteturas é tal que permite a elaboração dos pontos seguintes da Dissertação: Proposta de Solução Conceptual e da Proposta de uma Solução como Trabalho Empírico.

## 3 Proposta de Solução Conceptual

Neste capítulo, de acordo com a investigação efetuada, é proposto um modelo de referência onde estão identificados os principais processos/tipos de Gestão de Documentos de Arquivo, que depois se materializa com uma proposta de solução para um Sistema concreto de Gestão de Documentos de Arquivo, demonstrada sob a forma de um protótipo funcional.

Sendo uma Solução Conceptual, não propõe qualquer produto ou tecnologia concreta, descrevendo a arquitetura e os métodos de seleção das tecnologias e produtos concretos que implementam a solução de Gestão de Documentos de Arquivo. Nesta fase de definição da solução, o modelo descreve uma Arquitetura de Referência, Abstrata na medida em que define o tipo de blocos arquiteturais sem os concretizar sob o ponto de vista de produto. No capítulo seguinte, onde se descreve a construção da solução será depois efetuada a seleção dos produtos mais adequados a aplicar sobre a arquitetura de referência dando origem a uma arquitetura concreta que soluciona o problema estudado, da criação de um Sistema de Gestão de Documentos de Arquivo com software Livre ou de Código Aberto.

### 3.1 Arquitetura de referência

Na arquitetura de referência identificaram-se como principais blocos centrais do problema a resolver, as três áreas da Gestão de Conteúdos de Arquivo identificadas na revisão de literatura: Gestão Documental; Descrição Arquivística; Preservação Digital. Cada qual, por sua vez será desdobrada em áreas acessórias acerca das quais se encontraram referências nos trabalhos científicos estudados (OCR, Detecção de Layouts, Pesquisa). Seguidamente referem-se todas as questões de infraestrutura sejam elas de base como os Sistemas Operativos ou acessórias como Sistemas de Gestão de Mensagens e(ou) Sistemas de Gestão de Bases de Dados. Em resumo, o modelo conceptual de referência tipifica e identifica as várias componentes de um Sistema de Gestão de Documentos de Arquivo bem como a sua localização num modelo descrito por camadas e interligações. O diagrama que de seguida se apresenta, que de seguida se apresenta, foi criado usando a Linguagem de Descrição de Arquiteturas *Archimate* (*The Open Group, 2017*), que foi estudada e descrita na revisão de literatura, mais concretamente com o produto *Archi*, disponibilizado sob a forma de Software Livre («Open Source ArchiMate Modelling», 2019).



Figura 8: Arquitetura Conceptual

Da leitura do diagrama destacam-se a existência de vários níveis:

- Infraestrutura (Física ou Virtual) – Nós onde as diferentes componentes do Sistema serão instalados. Escalável e Distribuído (verde na figura);
- Serviços – Serviços Específicos disponíveis para a camada de integração ou utilizadores (cor de laranja na figura);
- Integração – Integração via API(s) ou Filas de Mensagens entre Sistemas (azul na figura);
- Interface com o Utilizador – Interfaces específicos ou partilhados para permitir aos utilizadores interagir com as componentes do Sistema (cinzento na figura).

### 3.2 Decisões Arquiteturais

De acordo com o investigado, nomeadamente a flexibilidade esperada num sistema que se pretende resolver o problema da gestão de Documentos de Arquivo de forma completa e versátil, concluiu-se que o Sistema a criar deve cumprir os seguintes requisitos arquiteturais (não funcionais):

- Deve ser implementado de forma a escalar o mais possível horizontalmente;
- Deve permitir um elevado grau de integração;

- Deve ser possível estender as funcionalidades através da introdução de novos componentes.

Assim, baseado nos requisitos acima definidos, tomaram-se as seguintes decisões arquiteturais:

- O sistema deve poder ser implementado numa arquitetura Física ou Virtualizada;
- Cada componente deve poder funcionar sob a forma de Máquina Física, Máquina Virtual ou Contentor;
- Sempre que possível deve ser exequível a adição de novas instâncias de cada componente;
- Novos componentes devem poder ser facilmente criados, adicionados de forma autónoma, comunicando e integrando com os existentes.

### 3.3 Componentes Arquiteturais

Para efeitos de descrição da arquitetura, o quadro que se segue apresenta os blocos arquiteturais identificados devidamente integrados numa taxonomia que assim os permite classificar quanto às suas funções no sistema.

**Tabela 18: Taxonomia de Componentes**

<b>Tipo</b>	<b>Sub-Tipo</b>	<b>Descrição</b>
Gestão de Documentos de Arquivo	Gestão Documental	Records Management ou Gestão de Documentos de Arquivo seguindo as recomendações do MoReq2010 ou do DoD5015.2.
	Descrição Arquivística	Descrição e Gestão dos Objetos com Interesse para Arquivo de acordo com a norma ISAD(G).
	Preservação Digital	Preservação dos Ativos digitais de acordo com o o modelo de referência OAIS.
	Assinatura Digital	Assinatura Digital Qualificada a apor nos documentos.
	Reconhecimento de Layouts	Reconhecimento do tipo de documento através do seu aspeto e arranjo gráfico.
	Optical Character Recognition	Reconhecimento de caracteres em imagens.
	Transformação de Formatos	Transformação de ficheiros num formato diferente mas de objetivos similares (ex: texto em texto, áudio em áudio).

<b>Tipo</b>	<b>Sub-Tipo</b>	<b>Descrição</b>
	Pesquisa	Pesquisa de informação nos vários componentes do sistema.
Infraestrutura	Sistema Operativo	Interface entre programas e computador físico ou virtual.
	Virtualização do servidor	Programa que cria uma versão virtual (não física) do hardware no qual o Sistema Operativo é executado.
	Contentor	Isolamento e virtualização da execução de programas do restante Sistema Operativo.
	Sistema de Gestão de Base de Dados	Servidor que permite armazenar e aceder a uma base de dados relacional.
	Plataforma de compilação e execução	Compiladores dos vários serviços adicionais, bem como respetivos run-times (ex: Java, PHP)
	Diretório de Utilizadores e Autenticação	Registo e autenticação de utilizadores e grupos.
	Integração	Message Oriented Middleware que, através da troca de mensagens permite uma integração organizada entre componentes.

### 3.4 Componentes para a Gestão de Documentos de Arquivo

Às três grandes áreas da Gestão de Documentos de Arquivo, cada uma instalando a sua componente (normalmente uma aplicação informática) concreta, juntam-se um conjunto de componentes auxiliares. Estas incluem, por exemplo, o reconhecimento ótico de caracteres e(ou) deteção de Layouts.

#### 3.4.1 Gestão Documental

Para a Gestão Documental (termo anglo-saxónico *Records Management*) foi selecionado um sistema que implementa da forma mais completa possível as principais normas relacionadas com este tema, nomeadamente ISO15489 (ISO, 2016), DoD 5015.2 («DoD Standard 5015.2», 2007) e(ou) MoReq 2010 (DLM Forum, 2010).

O sistema de Gestão Documental:

- Implementa
  - Plano de Classificação;
  - Calendário de Conservação;

- Autos de Destruição.
- Tem uma arquitetura:
  - Modular, nomeadamente implementando os serviços internos definidos pelo MoReq2010;
  - Expansível, no que se refere à possibilidade de implementar modelos de meta-dados;
  - Modular, permitindo a adição de novos módulos funcionais;
  - Integrável, sendo assim possível a integração com as outras componentes.

### **3.4.2 Descrição Arquivística**

A componente de Descrição Arquivística implementa a principal norma em vigor: ISAD(G) (ICA, 2011), bem como as normas acessórias como por exemplo a ISAAR(CPF) (Comitê de Normas de Descrição, ICA, 2004), nomeadamente:

- Os blocos de meta-dados da ISAD(G)
- Fundos;
- Séries

Além disso , em termos de arquitetura:

- É integrável, através de integração via base de dados ou execução de comandos externos;
- É expansível já que permite a adição de módulos adicionais, embora separados.

### **3.4.3 Preservação Digital**

O módulo de Preservação Digital implementa o Modelo de Referência OAIS (CCSDS, 2012) e usa formatos como os definidos pelo esquema METS (Library of Congress, 2015).

Em termos de Arquitetura:

- É integrável sob a forma de micro-serviços;
- É expansível permitindo a adição de programas adicionais para diferentes formatos.

### **3.4.4 Assinatura Digital**

Num sistema de Gestão de Documentos de Arquivo, sobretudo no que se refere à Gestão Documental (*Records Management*) deve existir a possibilidade de os utilizadores assinarem digitalmente os documentos que gerem. Estas componentes devem cumprir os seguintes requisitos:

- Estar disponível com uma Licença de Software Livre ou de Código Aberto;
- Suportar as principais tecnologias de assinatura e de encriptação;
- Suportar os principais formatos de suporte a armazenamento da chave privada, nomeadamente o Cartão do Cidadão Português («O Cartão de Cidadão», 2019).

### **3.4.5 OCR**

Não foi identificada qualquer norma no que se refere ao reconhecimento de caracteres após a sua digitalização, no entanto, para permitir melhor indexação e efetuar pesquisas sobre os conteúdos, trata-se de uma ferramenta fundamental que deverá estar disponível no sistemas.

Preferencialmente, esta componente deve poder ser implementada sob a forma de componente servidora, distribuída e escalável. Só desta forma, será possível a criação de sistemas que aloquem ou removam recursos sempre que a carga o impuser. De facto, trata-se de uma tarefa que necessita de muita capacidade de processamento e que não é usada de forma constante no tempo, o que significa que poderão ocorrer picos de utilização, nomeadamente em momentos de ingestão massiva de grandes massas documentais.

### **3.4.6 Deteção de Layouts**

Tal como no caso do OCR, também no caso de Deteção de Layouts, não foram identificadas normas relacionadas.

Trata-se no entanto de uma funcionalidade que, principalmente no caso da Gestão Documental é importante, na medida em que irá permitir automatizar a deteção de um tipo de documento, podendo inclusivamente vir a ajudar à sua classificação e(ou) ao reconhecimento de meta-dados em zonas específicas dos documentos.

Um exemplo de utilização desta tecnologia é a deteção das faturas por fornecedor e a extração do(s) valores que nela são referidos (através da utilização de OCR em zonas específicas do documento).

### **3.4.7 Indexação e Pesquisa**

Apesar de ser previsível a existência de um sistema de indexação e(ou) pesquisa em cada um dos módulos principais da Gestão de Documentos de Arquivo (Gestão Documental, Descrição Arquivística e Preservação Digital), um sistema da natureza global que se pretende implementar, necessita que as pesquisas sejam federadas, obrigando por isso à existência de um sistema de indexação mais global. Este sistema deve, se possível, implementar o protocolo OAI:PMH («Open Archives Initiative», 2019), ligando-se às três componentes (que, se possível, devem funcionar sob a forma de Data Provider de OAI:PMH), atuando como Service Provider.

### **3.4.8 Transformação de Formatos**

Para efeitos tanto de Preservação Digital, como de visualização na Gestão Documental, deve existir uma componente capaz de transformar um ficheiro em formatos alternativos, da mesma natureza (ex: doc em pdf, jpg em png).

Esta componente, se possível deve poder ser executada sob a forma de serviço, eventualmente assíncrono, libertando desta forma, as componentes principais enquanto a(s) transformações ocorrem.

Deverá cumprir os seguintes requisitos:

- Estar disponível com uma Licença de Software Livre ou de Código Aberto;
- Funcionar como serviço adicional;
- Efetuar transformações entre os principais formatos.

## **3.5 Componentes de infraestrutura**

As componentes de infraestrutura servem de suporte ao funcionamento dos módulos mais específicos da Gestão de Documentos de Arquivo. Sem elas, estes não poderiam funcionar, mas o seu âmbito é muito mais genérico, podendo por isso serem usadas em sistemas de inúmeras naturezas (ERP, CRM, IOT, entre outros).

### 3.5.1 Sistema(s) Operativo(s)

As restantes componentes, sejam elas de infraestrutura ou específicas do Sistema de Gestão de Documentos de Arquivo funcionarão em máquinas físicas ou virtuais. Cada um destes nós permitirá o funcionamento das diferentes componentes através da instalação do Sistema Operativo que efetua a gestão e interface dos ativos físicos (computadores) e as componentes, constituídas por conjuntos de programas e(ou) configurações, cada qual com o seu objetivo específico.

Deverá cumprir os seguintes requisitos:

- Ser disponibilizado com uma licença de Software Livre ou de Código Aberto;
- Permitir o funcionamento de todas as componentes do Sistema.

De notar que poderão ser escolhidos Sistemas Operativos diferentes para suportar diferentes componentes, caso seja necessário, no entanto esta diversidade deve ser evitada para não reduzir a produtividade da operação do Sistema.

### 3.5.2 Virtualização

Existem várias razões para a utilização de virtualização neste tipos de sistemas, especialmente a necessidade de usar diferentes versões de software de base que suportem diferentes componentes, evitando assim a *dependency hell* (problema derivado de vários softwares instalados dependerem de diferentes versões do mesmo software de base (Abate, Di Cosmo, Treinen, & Zacchiroli, 2012) ). O software de virtualização escolhido deve:

- Funcionar no Sistema Operativo escolhido como anfitrião;
- Permitir a instalação de todos os sistemas operativos “convidados”;
- Ser disponibilizado com licença de Software Livre ou de Código Aberto.

### 3.5.3 Orquestração de Contentores

Dados os objetivos de escalabilidade e a decisão de usar uma arquitetura com micro-serviços, a utilização de contentores para cada micro-serviço será uma decisão arquitetural. De alguns dos micro-serviços (como por exemplo o de OCR), devem poder existir várias instâncias (vários programas que implementam o micro-serviço em execução simultânea em qualquer nó do sistema). Desta forma, deverá existir uma componente que efetue a gestão das várias

instâncias dos vários contentores, nomeadamente a sua execução, criação, destruição, coordenação e(ou) distribuição de carga.

### **3.5.4 Autenticação e Diretório de Utilizadores**

Tanto os módulos principais usados para a Gestão de Documentos de Arquivo (Gestão Documental, Descrição Arquivística e Preservação Digital) estão implementados com gestão de utilizadores, grupos e respetivas permissões para efetuar as operações específicas.

Para reduzir a obrigatoriedade de gestão de utilizadores nos vários sistemas, um Sistema de Gestão de Documentos de Arquivo deve conter um mecanismo único, partilhado ou replicado para o registo e autenticação de utilizadores e grupos.

Este deve cumprir os seguintes requisitos:

- Implementar o protocolo LDAP;
- Conter interface de gestão para os gestores do sistema;
- Ser disponibilizado com Licença de Software Livre ou de Código Aberto.

### **3.5.5 Sistemas de Gestão de Bases de Dados Relacionais**

Os dois módulos, Gestão Documental e Descrição Arquivística persistem a informação em Sistemas de Bases de Dados Relacionais. Usualmente, através da utilização de bibliotecas que virtualizam o acesso às bases de dados, são suportados vários produtos.

Os requisitos dos Sistemas de Gestão de Bases de Dados são:

- Ser disponibilizados sob a forma de Software Livre ou de Código Aberto;
- Ser suportadas pelos módulos que as utilizam.

### **3.5.6 Filas de Mensagens Assíncronas**

Num sistema que conta com várias instâncias, sejam eles programas, máquinas físicas ou virtuais ou contentores, não existe garantia de que todas as componentes estão permanentemente em funcionamento. No entanto existe informação que, ao ser modificada num sistema, essa informação também deve ocorrer em outros, por exemplo quando uma análise de reconhecimento ótico terminar. Estas mensagens, tipicamente assíncronas devem ser geridas por sistemas de Filas de Mensagens Assíncronas. Estes sistemas permitem que um

emissor efetue o envio de mensagens e o recetor as receba em momentos diferentes, assim que estiver disponível para a receber. Além disso, são sistemas normalmente escaláveis horizontalmente e com persistência das mensagens que entretanto não foram consumidas.

O Sistema de Filas de mensagens deve cumprir os seguintes requisitos:

- Ser disponibilizado com licença de Software Livre ou de Código Aberto;
- Implementar os principais standards de Filas de Mensagens, concretamente AMQP, JMS;
- Permitir a implementação de troca de mensagens de acordo com os principais Padrões de Integração Empresariais (EAI).

### **3.5.7 Micro-serviços**

Estando perante um sistema bastante heterogéneo, surge a necessidade de implementação de programas tanto de integração como para implementar funcionalidades adicionais. Dada a necessidade de escalabilidade horizontal e heterogeneidade de tecnologias, a implementação de micro-serviços é uma decisão clara. Algumas das vantagens dos micro-serviços são: o facto de poder ser implementado em diferentes linguagens ou tecnologias; poderem ser lançadas instâncias múltiplas promovendo a escalabilidade horizontal; a simplicidade de implementações autónomas e focadas. Apesar de poderem ser seleccionadas várias tecnologias para a implementação de micro-serviços, estas devem, no entanto, cumprir um conjunto de requisitos como:

- Ser disponibilizados com Licenciamento Livre ou de Código Aberto;
- Implementar micro-serviços REST;
- Poderem ser executados em contentores;
- Poderem ser executadas múltiplas instâncias.

### **3.5.8 Indexação**

O sistema de indexação, apesar de já acima descrito na área específica da Gestão de Documentos de Arquivo, será suportado por um Sistema de Indexação. Este sistema deve:

- Ser escalável horizontalmente funcionando de forma distribuída;
- Poder ser integrável usando micro-serviços;
- Ser de elevado desempenho;

- Poder ser criado Interface(s) específico(s) com o utilizador.

### **3.6 Visão Global do Modelo Conceptual de Referência**

Em resumo, o Modelo de Referência identifica não só os principais processos/tipos de Componentes e Sub-Componentes necessários para implementar um Sistema de Gestão de Documentos de Arquivo, como também requisitos arquiteturais, sem os quais tal Sistema não funcionaria de forma correta (uma vez que poderia até implementar os requisitos específicos dos tipos de sistema, mas não os de integração ou escalabilidade de tal sistema). Este modelo, enfim, permite a construção de Sistemas de Gestão de Documentos de Arquivo, baseados em tecnologias disponibilizadas com licença de Software Livre ou de Código Aberto uma vez que define:

- Os tipos de Componentes específicos da Gestão de Documentos de Arquivo;
- As sub-componentes sejam elas acessórias (por exemplo OCR), de operação (por exemplo Sistemas Operativos), de Integração (por exemplo Sistema de mensagens ou Diretório de Utilizadores) ou de Escalabilidade (por exemplo os contentores e sua orquestração e exemplo de micro-serviços);
- A forma como as componentes devem interoperar;
- A metodologia para análise e seleção dos produtos que serão componentes do Sistema;
- Um modelo de escalabilidade.

Desta forma, na imagem que se segue apresenta-se uma visualização gráfica do modelo como tal sistema pode ser implementado.

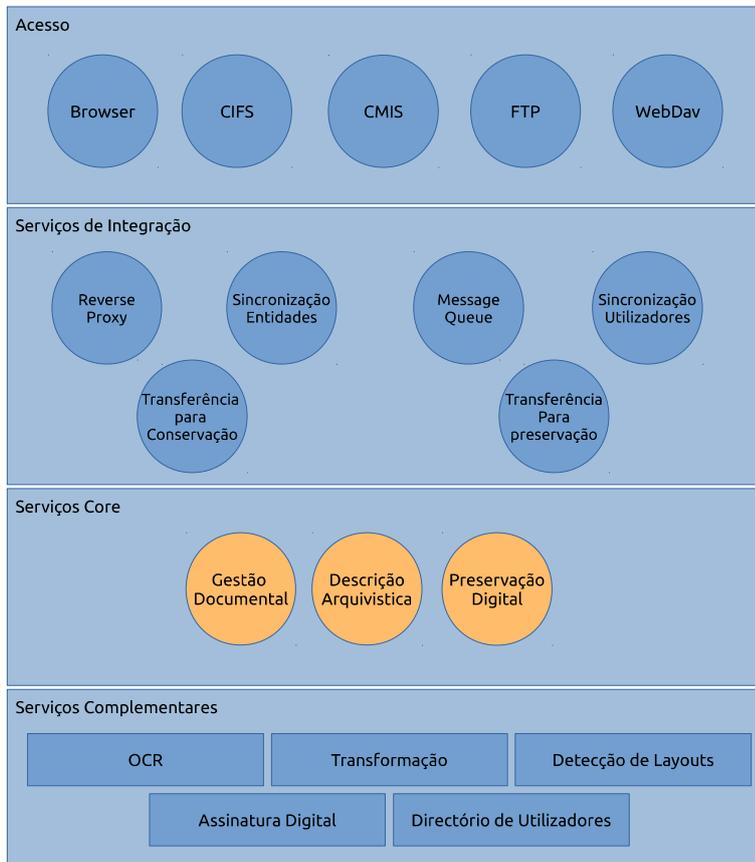


Figura 9: Visão Global do Modelo Conceptual

## 4 Trabalho Empírico

Na sequência da arquitetura conceptual desenhada, foi implementado um protótipo que incluiu todas as principais componentes necessárias para o correto funcionamento de um Sistema de Gestão de Documentos de Arquivo.

Todo o software usado, seja ele de Infraestrutura, de Integração ou de Gestão de Documentos de arquivo é distribuído com licenciamento de Software Livre ou de Código Aberto, embora com diferentes Licenças.

Neste capítulo descreve-se o processo de seleção das várias componentes que compõe o sistema, mas também a arquitetura que com eles se constrói uma possível solução e cada uma das componentes e também o seu papel na solução de Sistema para Gestão de Documentos de Arquivo. Desta forma, as componentes são enumeradas incluindo o seu mapeamento relativamente a uma taxonomia de tipo de componente, ao produto concreto e finalmente à função que este ocupa.

De acordo com as metodologias e respetivas técnicas estudadas na revisão de literatura, foi realizada uma pesquisa e análise de vários produtos de *software* já existentes e disponibilizados com licenças de Software Livre ou de Código Aberto, e, de entre eles, para cada área, foi selecionado o produto que melhor se adequa. A sua adequação está relacionada com a sua maturidade, arquitetura, funcionalidades implementadas, e tecnologia de infraestrutura usada. Para efeito de seleção dos produtos de negócio, foi usada a metodologia QSOS, estudada e descrita na revisão de literatura e respetivas ferramentas associadas. De notar que a metodologia QSOS foi usada apenas para as componentes principais e mais específicas da Gestão de Documentos de Arquivo: Gestão Documental; Descrição Arquivística; Preservação Digital. Todas as componentes foram primeiramente selecionadas através de pesquisas na Internet de acordo com a sua adequação à inclusão na arquitetura e o seu licenciamento.

Assim, na revisão de literatura foi estudado o estado da arte da área da Gestão de Documentos de Arquivo, dos normativos existentes para a área, da forma de avaliação da adequação de Software com Licenciamento livre ou de Código Aberto e das linguagens de descrição de arquiteturas. No segundo capítulo, já com base na investigação até ali efetuada, foi realizada uma proposta de conceito de solução com o qual, agora se vai concretizar na definição de uma

arquitetura concreta, escolha concreta de componentes e finalmente a definição da implementação.

## 4.1 Protótipo

O protótipo criado, demonstra ser possível a implementação de um Sistema de Gestão de Documentos de Arquivo com Software Livre ou de Código Aberto.

Para a sua criação foram efetuadas atividades de avaliação, seleção e implementação aplicando o Modelo Conceptual a uma realidade concreta. A atividade de implementação em concreto, consistiu em tarefas de:

- Definição concreta da localização dos componentes na arquitetura;
- Instalação de cada componente, tentando que esta possa ser reproduzida (usando para tal, sempre possível contentores);
- Configuração específica de cada componente, que, infelizmente não é trivial dada a complexidade do sistema, mas que se tentou, sempre possível automatizar através da utilização de scripting;
- Programação de Serviços de Integração, escaláveis e de acordo com uma arquitetura de micro-serviços;

No final, o protótipo criado disponibiliza um conjunto de artefactos, cuja ordenação integração e orquestração não é trivial, não estando por essa razão - da sua complexidade e dificuldade de integração e garantia de escalabilidade – disponíveis. Os artefactos que constituem o protótipo são:

- Máquina Virtual base – Imagem de um Linux, devidamente configurado para receber todas as componentes;
- Sistema de Gestão Documental – Contentor concreto instalado, configurado e integrado, incluindo um script de instalação que permite replicar a sua instalação/configuração;
- Sistema de Descrição Arquivística - Contentor concreto instalado, configurado e integrado, incluindo um script de instalação que permite replicar a sua instalação/configuração;

- Sistema de Preservação Digital - Contentor concreto instalado, configurado e integrado, incluindo um script de instalação que permite replicar a sua instalação/configuração;
- Micro-Serviços para replicação de autoridades/entidades e utilizadores – conjunto de programas em Java com Spring Boot («Spring Boot», 2019) que demonstram como integrar entre os sistemas usando micro-serviços instalados em contentores escaláveis;
- *Script* global de replicação da solução;
- Todas as configurações dos sub-sistemas acessórios, necessários para a Gestão de Documentos de Arquivo.

## 4.2 Seleção das Componentes

A seleção das Componentes Concretas que compõe o Sistema de Gestão de Documentos de Arquivo foi realizada de duas formas:

- Utilizando a metodologia QSOS no caso dos módulos core;
- Através de uma Investigação Sistemática das componentes mais adequadas no caso das componentes acessórias ou de infraestrutura.

Neste capítulo, descreve-se o processo de seleção das componentes concretas que irão constituir a proposta de solução concreta, objetivo desta dissertação, tanto em termos de ferramentas core do negócio da gestão de conteúdos como ferramentas auxiliares. A seleção baseou-se sempre na investigação, para cada uma das componentes identificadas na Proposta de Solução Conceptual, de que implementações concretas existem cujo Licenciamento seja de Software Livre e de Código Aberto. Para cada uma das componentes pesquisaram-se as comunidades onde se enquadram, as referências em artigos de investigação e interdependências (por exemplo em que sistema operativo pode ser executado o módulo de Descrição Arquivística). Finalmente, no caso dos três módulos principais: Gestão Documental, Descrição Arquivística e Preservação Digital, foram selecionados três produtos que poderiam funcionar como componentes, e, sobre eles aplicados a metodologia de seleção QSOS.

Pela sua complexidade e especificidade, foram efetuadas análises que usaram a metodologia QSOS, sobre o software mais específico das três áreas em estudo:

- Gestão Documental – Pacotes de repositório documental que implementam mecanismos de *Records Management*, nomeadamente a implementação dos requisitos do MoReq2010 e das definições/recomendações da ISO15489:1.

- Descrição Arquivística – Sistema que implementam a norma ISAD(G)/ISAAR(CPF) e(ou) CIDOC CRM.
- Preservação Digital – Software focado na implementação do Modelo de Referência OAIS.

A metodologia QSOS, foi investigada e descrita no capítulo de revisão de literatura, e tem como objectivos: analisar, avaliar e comparar as soluções *Open Source* por forma a tornar possível decidir quais os sistemas a seleccionar. Neste anexo publicam-se os resultados obtidos pela avaliação aplicada sobre os três principais sub-sistemas.

## **4.2.1 Gestão Documental**

Gestão Documental (cujo termo anglo-saxónico é Records Management) é um termo que descreve a Gestão de Documentos de Arquivo normalmente efetuada sobretudo na fase de arquivo corrente, mas também intermédio. Deve permitir implementar mecanismos diversos como planos de classificação, tabelas de seleção, meta-dados expansíveis entre outros, de acordo com normas como a ISO1482 ou o MoReq.

### **4.2.1.1 Modelo Utilizado para Análise**

O primeiro passo para realizar a avaliação do software de Gestão Documental consistiu na definição do modelo de análise que mais não é que a definição dos eixos de avaliação e em cada um deles, um conjunto de eixos em que cada um contem um conjunto de critérios valoráveis.

Os critérios usados para avaliar as soluções descritas foram divididos nos seguintes eixos:

- ♦ Maturity: Maturidade do projeto em termos do desenvolvimento e manutenção do produto
- ♦ Architecture: Conceitos arquiteturais
- ♦ System Services: Serviços Implementados pelo Software
- ♦ Functionality: Funcionalidades específicas de um serviço de Gestão Documental (RM)

### **4.2.1.2 Maturidade**

A maturidade define de que forma um projeto tem historia e organização que lhe permita ser considerado suficientemente confiável para ser escolhido como solução com futuro.

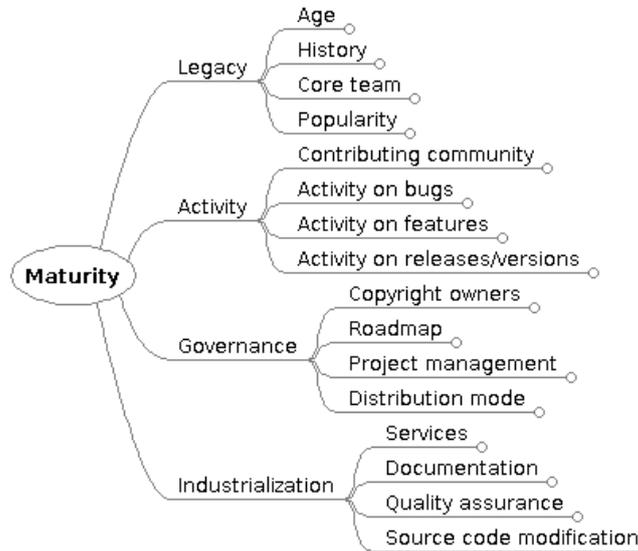


Figura 10: Critérios de Avaliação da Maturidade de Software de Gestão Documental para aplicação da metodologia QSOS

A avaliação da maturidade é fixa na metodologia QSOS, não podendo ser alterada. Avalia as seguintes dimensões:

- ♦ Legacy: História e herança do Projeto;
- ♦ Activity: Atividade interna e à volta do Projeto;
- ♦ Governance: Estratégia do Projeto;
- ♦ Industrialization: Nível de organização do Projeto de acordo com as boas práticas de desenvolvimento e implementação de Sistemas Informáticos.

### 4.2.1.3 Arquitetura

A arquitetura avalia se em termos de definição da sua implementação, esta é definida de acordo com uma estrutura correta e viável.

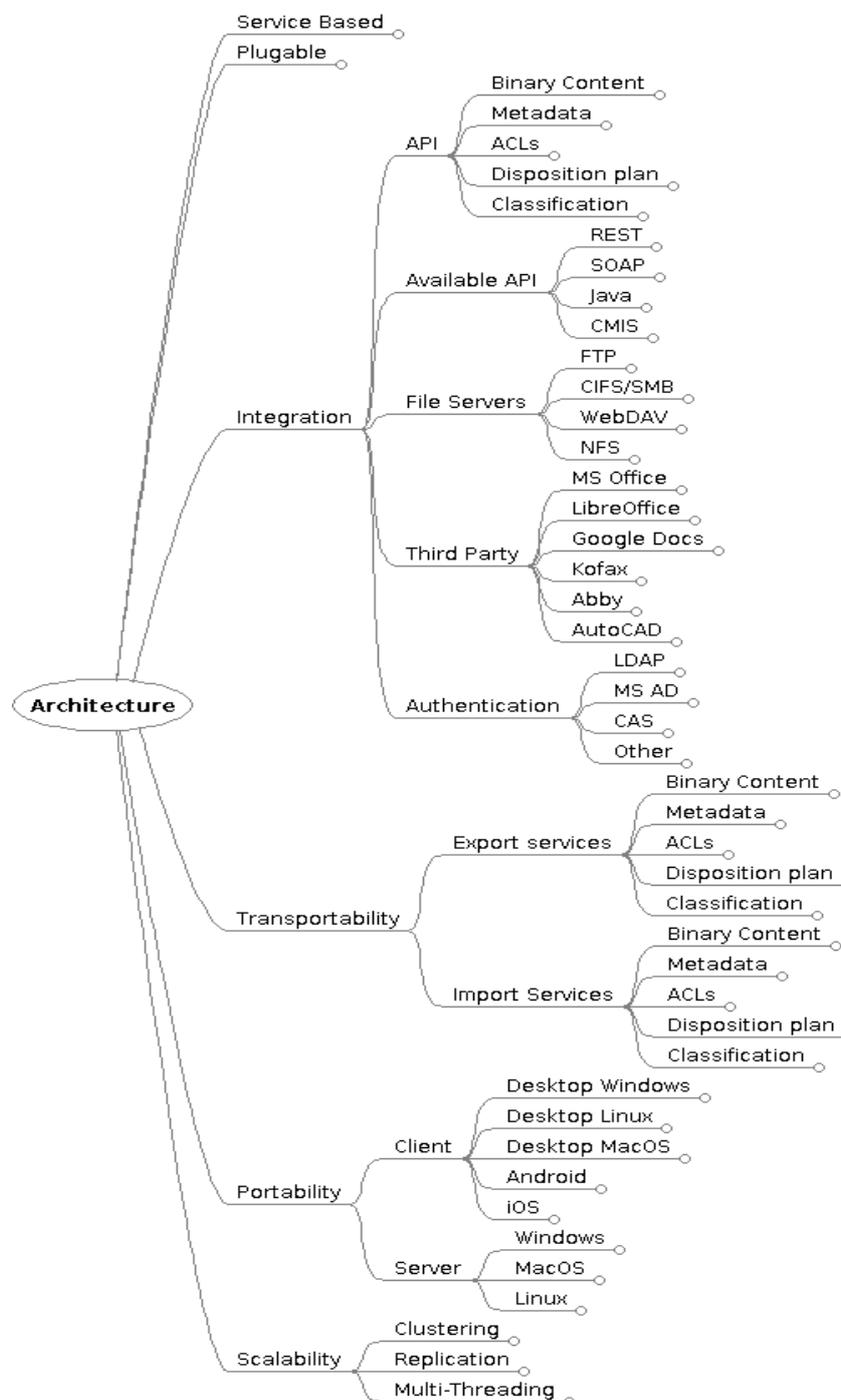


Figura 11: Critérios de Avaliação da Arquitetura de Software de Gestão Documental para aplicação da metodologia QSOS

As questões relacionadas com os blocos funcionais da solução, concretamente questões arquiteturais, são avaliados de acordo com os seguintes eixos:

- ◆ Service Based: A arquitetura está implementada de acordo com uma divisão clara por serviços autónomos embora cooperantes;
- ◆ Plugable: A arquitetura permite a adição de módulos autónomos independentes e não fundamentais à operação de base;

- ♦ Integration: A arquitetura leva em conta necessidades de integração com outros sistemas;
- ♦ Transportability: A arquitetura está montada por forma a permitir o transporte dos conteúdos de ou para outros sistemas;
- ♦ Portability: A implementação é suficientemente genérica para permitir a sua operação em diferentes ambientes operativos;
- ♦ Scalability: A arquitetura está criada por forma a garantir que é possível o sistema escalar o seu desempenho à medida que a sua utilização aumenta.

#### 4.2.1.4 Serviços de Sistema

Os critérios avaliados em relação aos serviços estão diretamente alinhados com normas como o MoReq e(ou) a ISO15482.

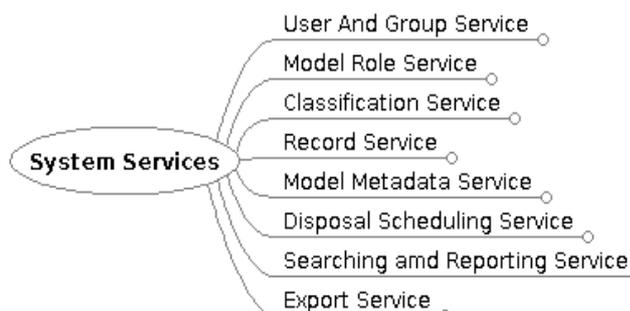


Figura 12: Critérios de Avaliação da Serviços Disponíveis em Software de Gestão Documental para aplicação da metodologia QSOS

A relação com a definição dos serviços, nomeadamente do MoReq2010 definiu a avaliação de:

- ♦ User And Group Service: Existência de Serviço de Gestão de Utilizadores e Grupos;
- ♦ Model Role Service: Papeis que podem ser desempenhados pelos utilizadores e(ou) grupos e permissões a estes aplicadas;
- ♦ Classification Service: Existência de um Serviço de Classificação dos Registos de acordo com um Plano de Classificação;
- ♦ Record Service: O sistema implementa um serviço para a Gestão de Registos (RM);
- ♦ Model Metadata Service: Existe um serviço que permite a criação e utilização de modelos de meta-dados
- ♦ Disposal Scheduling Service: O sistema implementa serviços que permitem a definição de calendários de retenção, conservação e(ou) destruição;
- ♦ Searching and Reporting Service: Existe um sistema de pesquisa e emissão de relatórios;
- ♦ Export Service: Existe(m) serviço(s) que permitem a exportação de registos e conteúdos.

## 4.2.1.5 Funcionalidades

A avaliação das principais funcionalidades do Sistema foram também criadas com base nas principais normas como o MoReq e(ou) ISO1482.

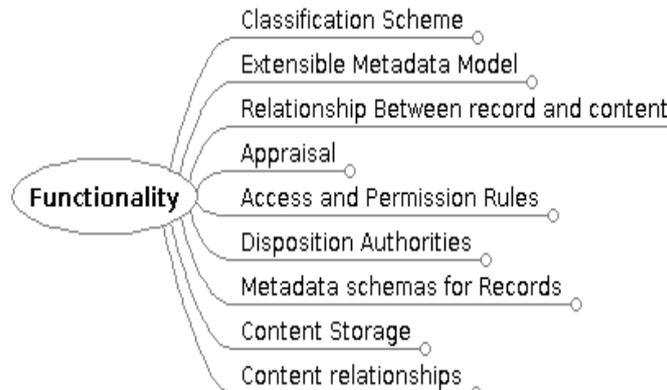


Figura 13: Critérios de Avaliação das Funcionalidades de Software de Gestão Documental para aplicação da metodologia QSOS

No que se refere às principais funcionalidades do sistema, avaliou-se de acordo com os seguintes critérios:

- ♦ Classification Scheme: Estão implementadas funcionalidades que permitem a existência de um serviço de classificação;
- ♦ Extensible Metadata Model: É possível estender o modelo de meta-dados;
- ♦ Relationship Between record and content: Estão implementadas funcionalidades que permitem a relacionar registos e conteúdos;
- ♦ Appraisal: Existe forma de construir uma tabela de seleção;
- ♦ Access and Permission Rules: O software contém possibilidade de criar regras de acesso e de permissões;
- ♦ Disposition Authorities: O sistema implementa a geração de autos de eliminação;
- ♦ Metadata Schemas for Records: É possível a criação de modelos de meta-dados para os registos;
- ♦ Content Storage: O sistema implementa armazenamento de conteúdos;
- ♦ Content relationships: O sistema permite a existência de Relação entre Conteúdos.

## 4.2.1.6 Soluções Identificadas

A identificação das várias soluções foi efetuada através de uma pesquisa sistemática realizada através da identificação de palavras chave, pesquisando referência a produtos e os respetivos projetos.

### 4.2.1.6.1 Universo de Soluções

As solução seleccionadas e analisadas foram:

- ♦ Alfresco Community (<https://community.alfresco.com>): Foi avaliado o software Alfresco Community acrescido do módulo de extensão Records Management.  
O Alfresco foi desenvolvido pela empresa Alfresco, Inc em Maidenhead, UK. Os seus fundadores e líderes têm muita experiência na implementação do Documentum que venderam à EMC. O Alfresco é desenvolvido em Java, com interface para o utilizador a funcionar num browser (aplicação web).
- ♦ LogicalDoc 7.7.5 (<https://www.logicaldoc.com>):  
O LogicalDoc foi desenvolvido, é mantido e suportado pela LogicalDoc, Inc sediada em Itália. É um ECM desenvolvido em Java com um cliente rico desenvolvido em GWT.
- ♦ Nuxeo 10.1 (<http://www.nuxeo.com>):  
O Nuxeo é um Enterprise Content Management, criado, financiado e desenvolvido pela empresa Francesa com o mesmo nome. Apesar de ser um negócio empresarial, todo o código fonte para todas as versões está disponível. O Nuxeo está desenvolvido em Java e tem interface web.
- ♦ OpenKM 6.3.6 (<https://openkm.com>):  
O software OpenKM foi desenvolvido e é mantido por uma companhia sediada em Palma de Maiorca, Espanha. O seu principal produto é o OpenKM, desenvolvido em Java com um interface Web.

O ECM é uma área onde existem muitas soluções implementadas sob a forma de Software Livre ou de Código Aberto. Os produtos avaliados foram selecionados com base na existência de mais referências encontradas nas pesquisas efetuadas.

#### **4.2.1.7 Alfresco Community**

O Alfresco, é um produto já maduro, dado que já conta com dezenas de versões que são usadas por milhares de organizações. Está implementado com uma arquitetura bem estruturada, com um *back-end* em Java e um *front-end web*. Além dos principais programadores, colaboradores remunerados da Alfresco, conta com uma comunidade importante, grande parte dela organizada formalmente em torno da associação *Order Of the Bee*. A sua maior fraqueza é o fato de a empresa ter sido adquirida por um fundo já não sendo detida nem pelos fundadores nem pelos engenheiros originais, já pode desfocar a estratégia para longe do desenvolvimento do produto.

##### **4.2.1.7.1 Maturidade do Projeto**



Figura 14: Análise QSOS - Maturidade do Alfresco Community

Objetivamente, tal como ilustrado na figura anterior, podemos facilmente compreender que cumpre de forma muito completa os critérios definidos para a sua avaliação.

#### 4.2.1.7.2 *Princípios Operativos*

O Alfresco tem o seu núcleo desenvolvido em Java, usando Spring como principal *Framework*, devidamente estruturado em serviços autónomos. O núcleo é uma *web application Java* à qual os clientes se podem conectar por serviços REST, CMIS, WebDAV, Sharepoint protocol entre outros.

Implementa vários *front-end(s)*, nomeadamente um interface web, chamado Alfresco Share, que funciona em *Browsers* e é expansível. É possível conexão a partir de qualquer cliente CMIS (por exemplo a partir do LibreOffice). Disponibiliza cliente Android e MacOS para dispositivos móveis.

#### 4.2.1.7.3 *Arquitetura*

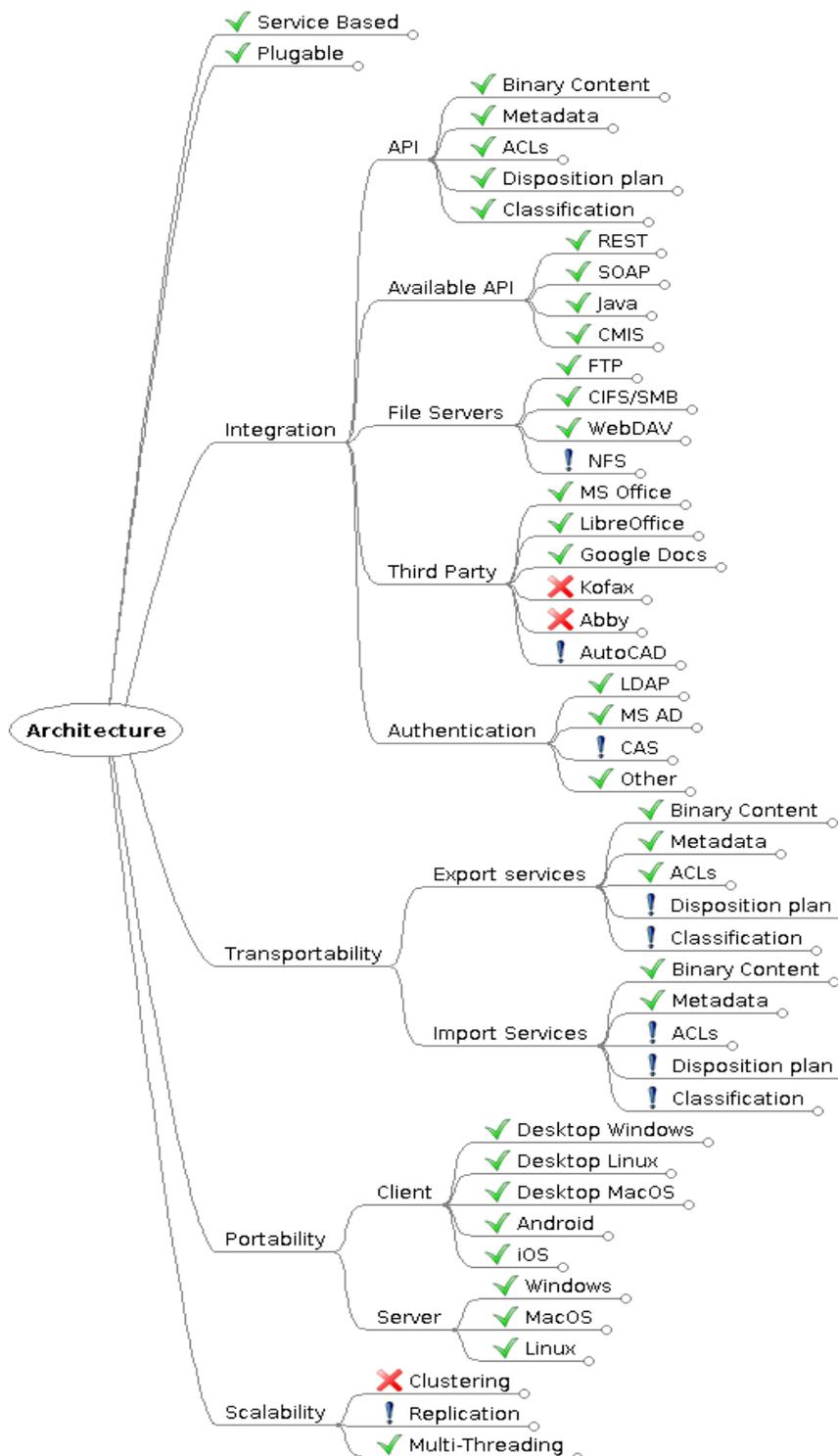


Figura 15: Arquitetura do Alfresco Community

O Alfresco é implementado com tecnologia recente, que a empresa tenta sempre atualizar, na medida do possível (com muito cuidado para que a base instalada se mantenha sempre a operar). As suas ligações a sistemas externos tornam-no extremamente versátil e capaz de se

adaptar a ambientes empresariais. Implementações importantes, como o funcionamento em *clustering*, fundamental numa arquitetura que se pretende escalável só esta disponível na versão empresarial. A partir da versão 6.0 a escalabilidade baseia-se em contentores devidamente orquestrados pelo *Kubernetes*.

### Serviços do Sistema



Figura 16: Análise QSOS - Serviços de Sistema do Alfresco Community

O Alfresco implementa todas a principais funcionalidades definidas como requisitos arquiteturais pelo MoReq2010.

### Funcionalidades



Figura 17: Análise QSOS - Funcionalidades do Alfresco Community

As funcionalidades implementadas pelo Alfresco, acrescidas pelo módulo de *Records Management* permitem a implementação de um Sistema de Gestão Documental muito completo e que cumpre os principais requisitos definidos por normativos e boas práticas, como a ISO15482, MoReq e DoD5015.2.

## 4.2.1.8 LogicalDoc

O LogicalDoc é um produto maduro que já conta com bastantes versões, apesar de não ser dos mais usados. Não se encontraram evidências acerca da existência de uma estratégia clara de evolução ou de um processo de desenvolvimento evoluído.

### 4.2.1.8.1 Maturidade do Projeto

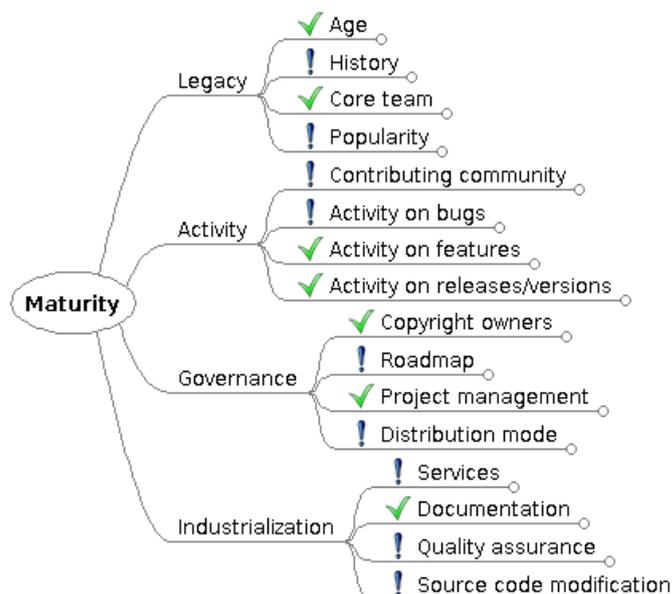


Figura 18: Análise QSOS - Maturidade do Logical Doc

Embora seja um produto com história, o LogicalDoc não parece ser um produto com uma estratégia, comunidade e organização clara e transparente.

#### 4.2.1.8.2 *Princípios Operativos*

O LogicalDoc está desenvolvido com uma arquitetura evoluída, disponibilizando uma componente servidora à qual os interfaces com o utilizador se ligam. O *front-end* funciona em *browsers* e é desenvolvido com o GWT – Google Web Toolkit. A utilização desta tecnologia que não é considerada como de grande futuro, e o facto de o front-end ser disponibilizado com a componente servidora faz que não seja claro qual a dimensão da evolução do produto para tecnologias mais usadas globalmente.

#### 4.2.1.8.3 *Arquitetura*

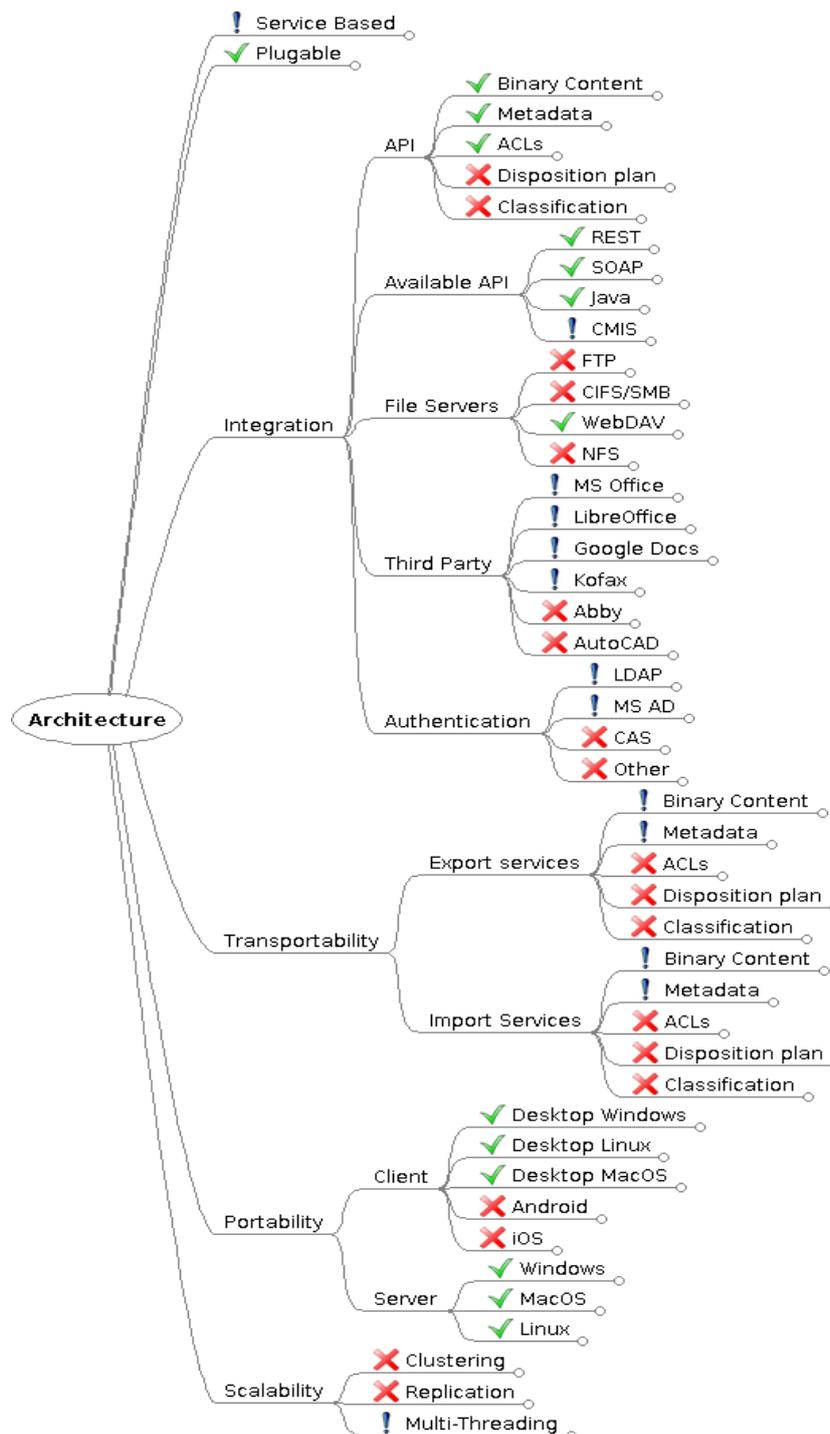


Figura 19: Análise QSOS - Maturidade do LogicalDoc

O LogicalDoc, ao não implementar a maioria das funcionalidades de *Records Management*, é um produto de ECM incompleto. Além disso não tem versão para dispositivos móveis nem apresenta evidências de ser escalável.

### Serviços de Sistema



Figura 20: Análise QSOS - Serviços de Sistema do LogicalDoc

A inexistência de funcionalidades de Gestão de Registos torna o LogicalDoc pouco cumpridor no que se refere aos requisitos definidos pelo MoReq2010.

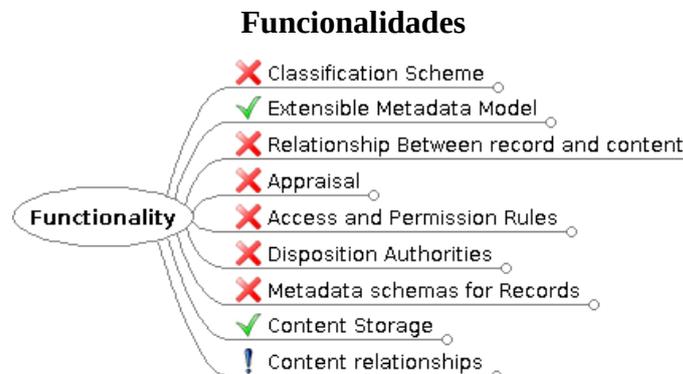


Figura 21: Análise QSOS - Funcionalidades do LogicalDoc

O LogicalDoc não implementa funcionalidades importantes para implementar um sistema de Gestão de Documentos de Arquivo, muito embora, possa ser uma boa solução no que se refere a gestão de conteúdos.

## 4.2.1.9 Nuxeo

O Nuxeo é um produto maduro: Conta com dezenas de versões a funcionar em dezenas de clientes ao longo de vários anos; Tem uma equipa de desenvolvimento sólida, bem liderada e com um caminho e visão claramente definidos.

### 4.2.1.9.1 Maturidade do Projeto



Figura 22: Análise QSOS - Maturidade do Nuxeo

#### 4.2.1.9.2 Princípios Operativos

O Nuxeo funciona com uma arquitetura que se baseia na existência de um *backend* ao qual os serviços de apresentação se ligam. Está desenvolvido em Java, usando *frameworks* populares, muito produtivas e funcionais. Disponibiliza um grande conjunto de módulos, para resolução de diferentes tipos de problemas. O *frontend* funciona em *Browsers* e está desenvolvido em GWT.

#### 4.2.1.9.3 Arquitetura

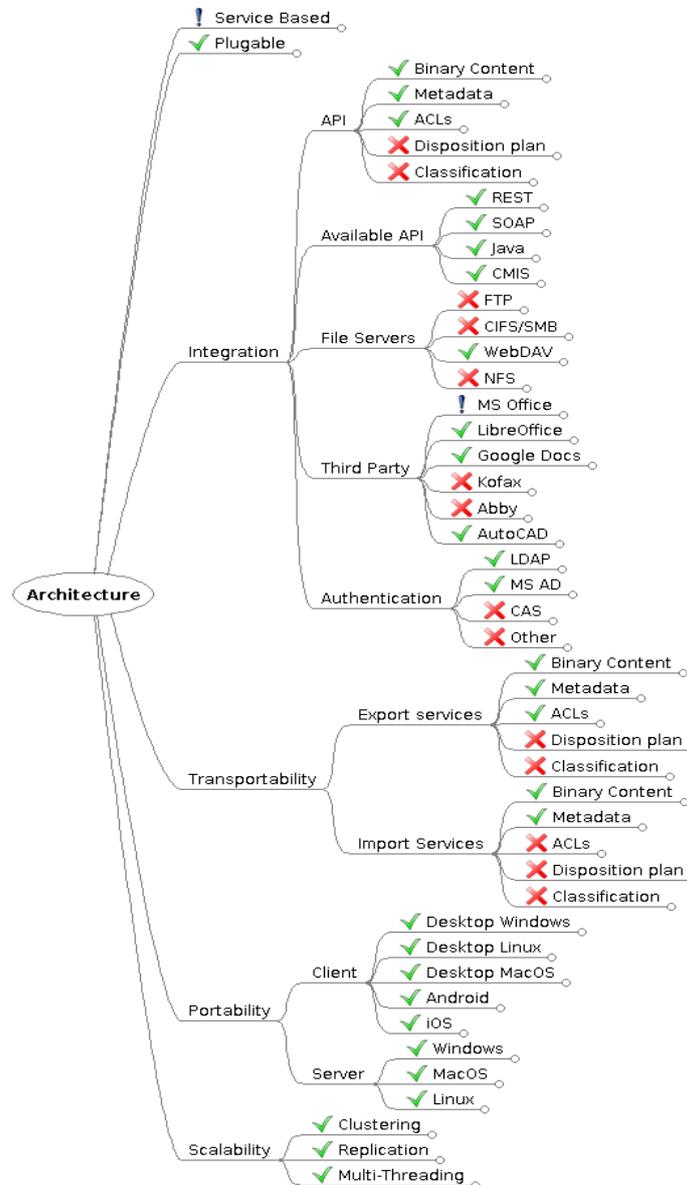


Figura 23: Análise QSOS - Maturidade do Nuxeo

O Nuxeo é uma plataforma de ECM muito completa, mas não implementa funcionalidades de *Records Management*. Apesar de se notar a implementação de muitos tipos de funcionalidades em tecnologias diferentes, esta diversidade poderá estar a causar desperdício de recursos.

### Serviços de Sistema

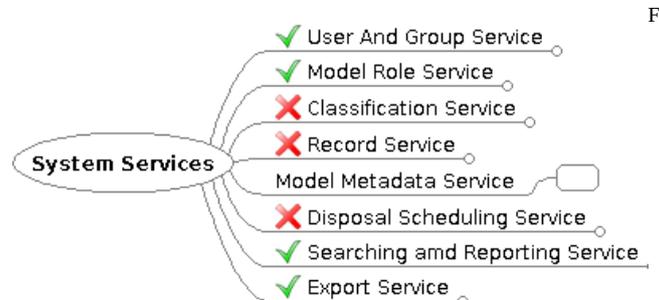


Figura 24: Análise QSOS - Serviços de Sistema do Nuxeo

Apesar de estar estruturado sob a forma de serviços e de implementar serviços importantes para a gestão de conteúdos, não implementa funcionalidades fundamentais de um sistema de Gestão Documental.

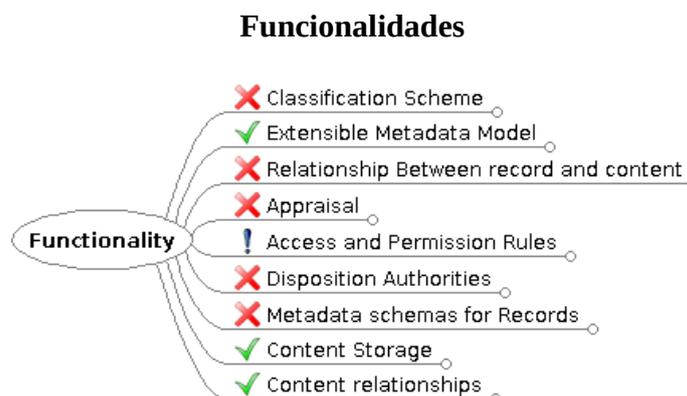


Figura 25: Análise QSOS - Funcionalidades do Nuxeo

Ao nível das principais funcionalidades de um ECM, o Nuxeo é um produto muito versátil e completo, no entanto, a inexistência de um módulo de *Records Management* torna-o limitado no que se refere à implementação dos principais requisitos do MoReq2010 e(ou) da ISO15482.

#### 4.2.1.10 OpenKM

O software OpenKM foi desenvolvido e é mantido por uma companhia sediada em Palma de Maiorca, Espanha. É desenvolvido em Java com um interface Web. É disponibilizado com uma Licença *GNU General Public Licence* (versão 2).

##### 4.2.1.10.1 Maturidade do Projeto

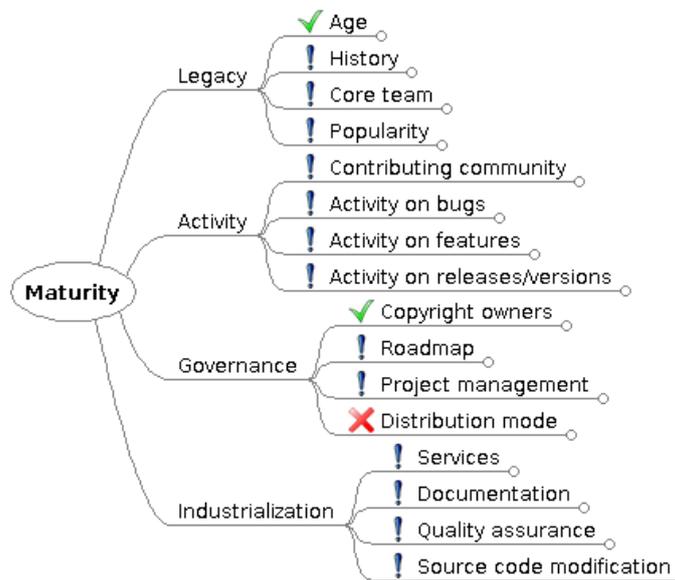


Figura 26: Análise QSOS - Maturidade do OpenKM

O OpenKM não apresenta evidências de disponibilizar estratégia clara, comunidade ativa ou equipa coesa que derive numa visão estratégica e organização eficaz do projeto.

#### 4.2.1.10.2 Princípios Operativos

O OpenKM é uma aplicação Web desenvolvida em Java com algumas das *frameworks* mais populares. O interface com o utilizador está desenvolvido em GWT.

#### 4.2.1.10.3 Arquitetura

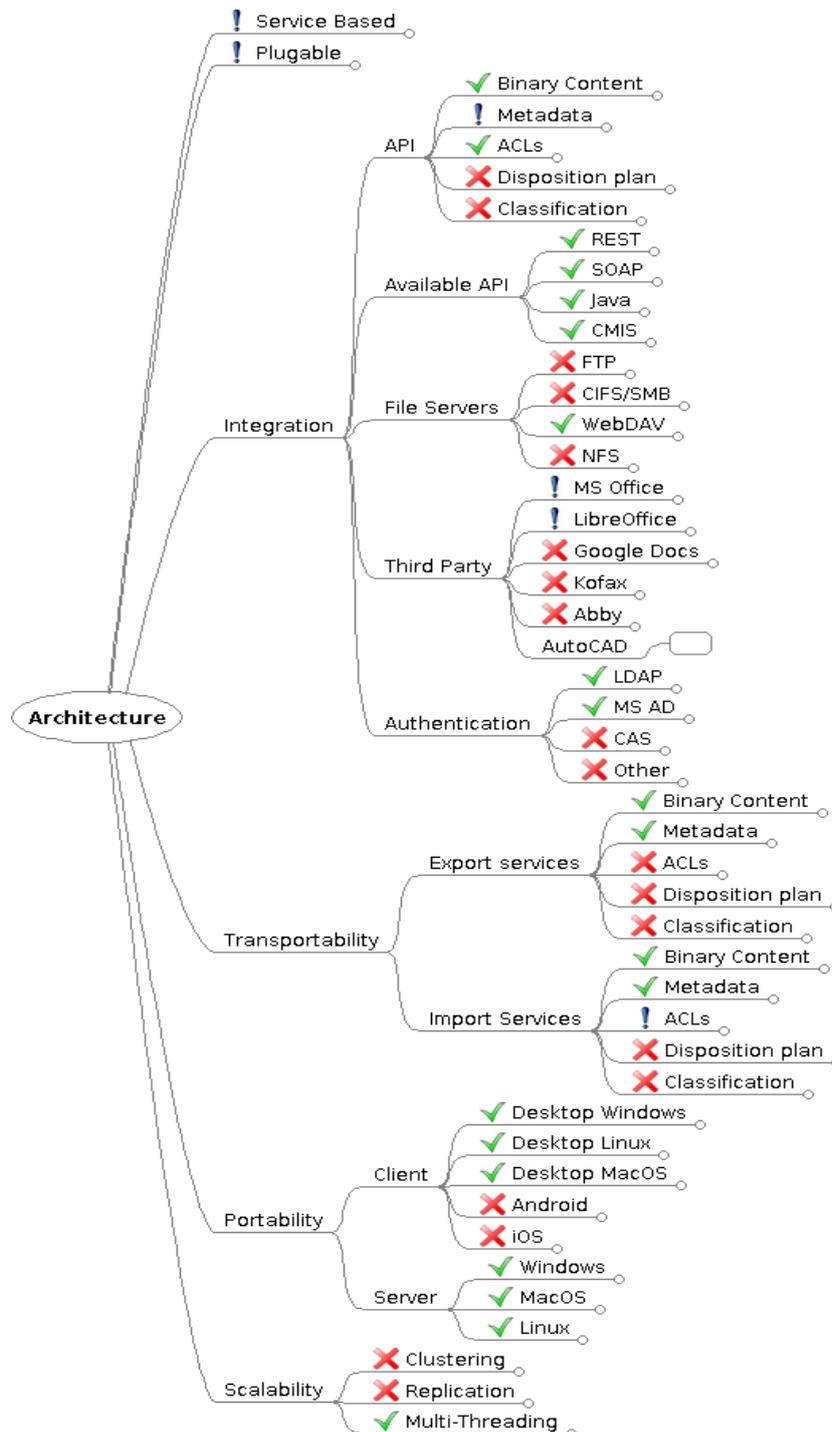


Figura 27: Análise QSOS - Maturidade do OpenKM

O OpenKM implementa as principais funcionalidades esperadas num ECM, no entanto não apresenta evidências de ser escalável. A sua base de código é algo confusa.

### Serviços do Sistema



Figura 28: Análise QSOS - Serviços de Sistema do OpenKM

Apesar de estar estruturado sob a forma de serviços e de implementar serviços importantes para a gestão de conteúdos, não implementa funcionalidades fundamentais de um sistema de Gestão Documental.

### Funcionalidades

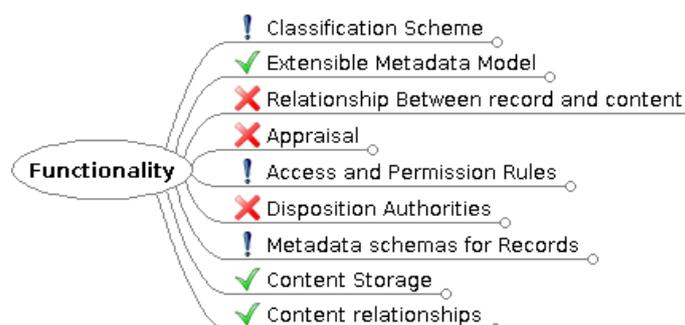


Figura 29: Análise QSOS - Funcionalidades do OpenKM

O OpenKM não implementa as principais funcionalidades de *Records Management*, e por isso não cumpre os principais requisitos arquiteturais definidos pelo MoReq2010 e(ou) pela ISO 15482.

## 4.2.1.11 Síntese

### 4.2.1.11.1 Comparação das Soluções

#### Maturidade

Alfresco Community 5.2 LogicalDoc 7.7.5 Nuxeo 10.1 OpenKM 6.3.6 > Maturity

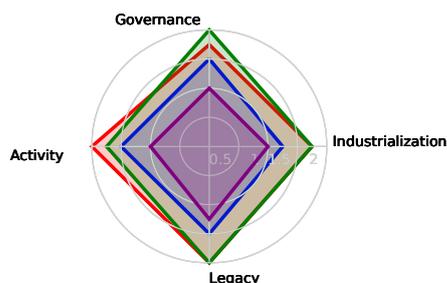


Figura 30: Análise QSOS - Comparação de Maturidades de Sistemas de Gestão Documental

A maturidade do Alfresco destaca-se a todos os níveis. A empresa e a equipa que o desenvolve tem uma noção muito clara acerca do caminho que pretende seguir e todo o processo desde a conceção à comercialização está bem implementado e desenvolvido.

### Arquitetura

Alfresco Community 5.2 LogicalDoc 7.7.5 Nuxeo 10.1 OpenKM 6.3.6 > Architecture

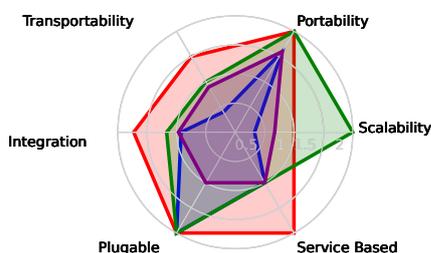


Figura 31: Análise QSOS - Comparação de Arquiteturas de Sistemas de Gestão Documental

O Alfresco Community está muito bem arquitetado o que permite a sua evolução e implementação das funcionalidades chave. Ao nível da escalabilidade, falta à versão comunitária a implementação do *clustering*, disponível na versão empresarial.

### Serviços do Sistema

Alfresco Community 5.2 LogicalDoc 7.7.5 Nuxeo 10.1 OpenKM 6.3.6 > System Services

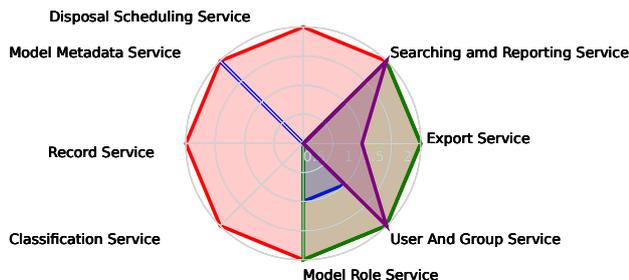


Figura 32: Análise QSOS - Comparação de Serviços dos Sistemas de Gestão Documental

O Alfresco Community destaca-se em termos arquiteturais no que se refere aos requisitos definidos pelo MoReq2010, nomeadamente no que se refere à existência de serviços correspondentes à Gestão Documental (*Records Management*).

### Funcionalidades

Alfresco Community 5.2 LogicalDoc 7.7.5 Nuxeo 10.1 OpenKM 6.3.6 > Functionality

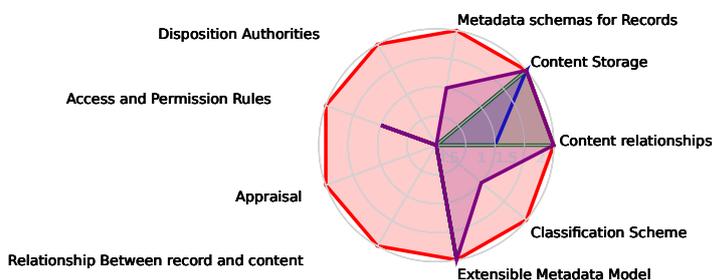


Figura 33: Análise QSOS - Comparação de Funcionalidades de Sistemas de Gestão Documental

Nas funcionalidades específicas de um sistema de Gestão Documental (*Records Management*) o Alfresco destaca-se pelo facto de implementar um módulo específico, atualmente chamado de *Governance* e que está presente na sua estratégia de evolução.

#### 4.2.1.11.2 Conclusão

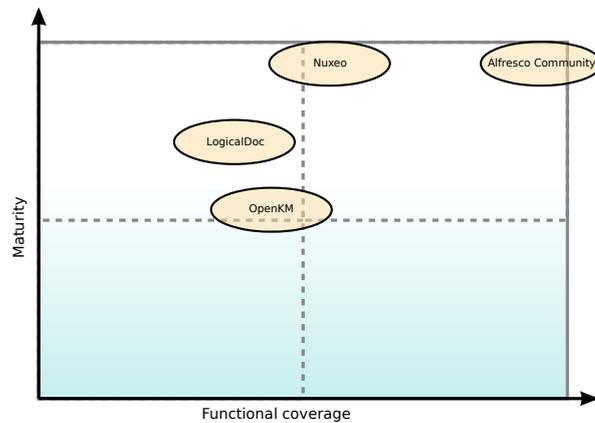


Figura 34: Análise QSOS - Conclusão das comparações de Sistemas de Gestão Documental

O Alfresco Community é claramente o produto de Software Livre ou de Código Aberto que dá mais garantias de melhor implementar um Sistema de Gestão Documental.

## 4.2.2 Descrição Arquivística

Descrição Arquivística é um termo usualmente aplicado a Arquivos Históricos, mesmo se literalmente o seu significado possa ser mais extensivo. Está associado a normas como a ISAD(G) e acessórias como a ISAAR(CPF) ou ao CIDOC-CRM. Um sistema de Descrição Arquivística preocupa-se em cumprir as normas e a implementar funcionalidades acessórias como gestão de utilizadores, exportação e(ou) importação de registos.

### 4.2.2.1 Modelo Utilizado para Análise

Os critérios usados para avaliar as soluções descritas foram divididos nos seguintes eixos:

- ♦ Maturity: Maturidade do projeto em termos do desenvolvimento e manutenção do produto;
- ♦ Architecture: Conceitos arquiteturais;
- ♦ Functionality: Funcionalidades específicas de um serviço de Descrição Arquivística;
- ♦ Standards: Cumprimento dos principais standards da área de Descrição Arquivística.

## 4.2.2.2 Maturidade

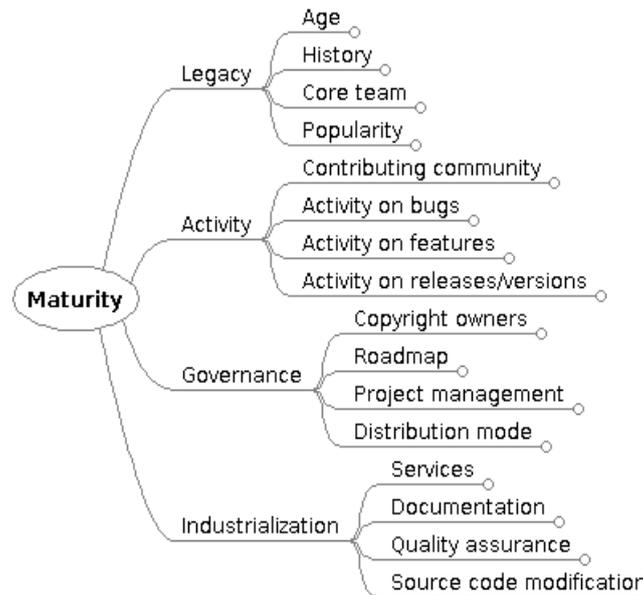


Figura 35: Critérios de Avaliação da Maturidade de Software de Descrição Arquivística para aplicação da metodologia QSOS

A avaliação da maturidade é fixa na metodologia QSOS. Avalia as seguintes dimensões:

- ♦ Legacy: História e herança do Projeto
- ♦ Activity: Atividade interna e à volta do Projeto
- ♦ Governance: Estratégia do Projeto
- ♦ Industrialization: Nível de industrialização do Projeto

### 4.2.2.3 Arquitetura

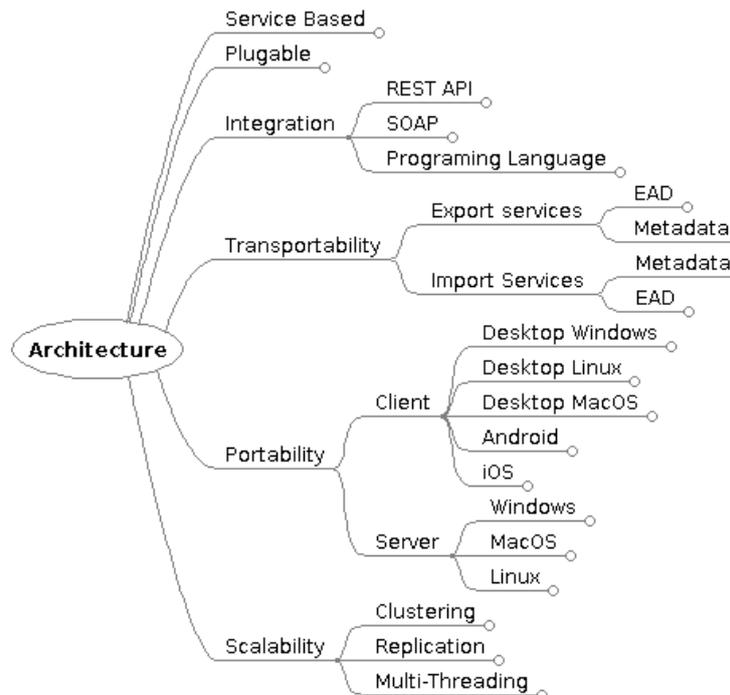


Figura 36: Critérios de Avaliação da Arquitetura de Software de Descrição Arquivística para aplicação da metodologia QSOS

As questões relacionadas com os blocos funcionais da solução, concretamente questões arquiteturais, foram avaliados de acordo com os seguintes eixos:

- ♦ **Service Based:** A arquitetura está implementada de acordo com uma divisão clara por serviços autónomos embora cooperantes.
- ♦ **Plugable:** A arquitetura permite a adição de módulos autónomos independentes e não fundamentais à operação de base.
- ♦ **Integration:** A arquitetura leva em conta necessidades de integração com outros sistemas.
- ♦ **Transportability:** A arquitetura está montada por forma a permitir o transporte dos conteúdos de ou para outros sistemas.
- ♦ **Portability:** A implementação é suficientemente genérica para permitir a sua operação em diferentes ambientes operativos.
- ♦ **Scalability:** A arquitetura está criada por forma a garantir que é possível o sistema escalar o seu desempenho à medida que a sua utilização aumenta.

### 4.2.2.4 Funcionalidades

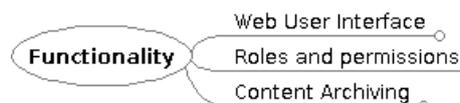


Figura 37: Critérios de Avaliação das Funcionalidades de Software de Descrição Arquivística para aplicação da metodologia QSOS

As funcionalidades avaliadas são as fundamentais em termos de um sistema de Descrição Arquivística já que os standards, avaliados noutro eixo, são determinantes:

- ♦ Web User Interface: O interface funciona num Browser;
- ♦ Roles and permissions: O sistema implementa gestão de utilizadores, os seus papeis e permissões sobre os objetos descritos;
- ♦ Content Archiving: O sistema implementa o arquivo dos conteúdos.

#### 4.2.2.5 Standards



Figura 38: Standards Implementados para análise QSOS

Na revisão de literatura identificaram-se os principais standards, sendo os quatro primeiros definidos pelo ICA:

- ♦ ISAD(G)
- ♦ ISAAR(CPF)
- ♦ ISDIAH
- ♦ ISDF
- ♦ CIDOC-CRM

#### 4.2.2.6 Soluções Identificadas

A identificação das várias soluções foi efetuada através de uma pesquisa sistemática realizada através da identificação de palavras chave, pesquisando referência a produtos e os respetivos projetos.

#### 4.2.2.7 Universo de Soluções

As soluções analisadas foram:

- ♦ AtoM - Access To Memory 2.4.0 (<https://www.accesstomemory.org>): Software desenvolvido pela empresa Artefactual Systems Inc, sediada no Canada, em colaboração com o Conselho Internacional de Arquivos e a Unesco.
- ♦ Collective Access 1.7.6 (<https://www.collectiveaccess.org>): Collective Access é um software livre que funciona em web para gerir e publicar coleções de arquivos e(ou) museus.

Não se encontraram mais projetos de Software Livre ou Código Aberto que estejam ativos e que implementem as normas ISAD(G) e associadas ou as normas CIDOC-CRM.

## 4.2.2.8 AtoM - Access To Memory 2.4.0

Software de Descrição Arquivística desenvolvido pela empresa Artefactual Systems Inc, sediada no Canadá, em colaboração com o Conselho Internacional de Arquivos e a Unesco. É suposto implementar as principais normas associadas à ISAD(G).

### 4.2.2.8.1 Maturidade do Projeto

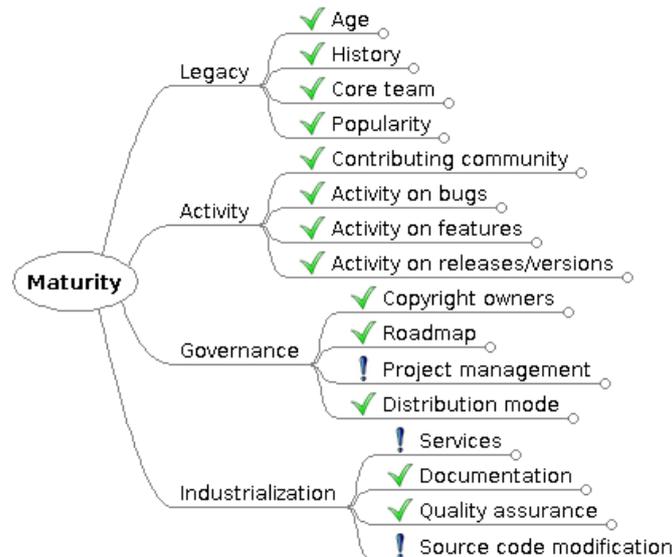


Figura 39: Análise QSOS - Maturidade do AtoM

O AtoM apresenta os principais sinais de um projeto que já é maduro: com existência de várias versões, instalação em bastantes organizações. Dá garantias de evolução já que existe uma visão clara e um *roadmap* com planeamento e uma organização com operação comercial viável que suporta a evolução do projeto.

### 4.2.2.8.2 Princípios Operativos

O AtoM é uma aplicação Web monolítica, implementada em PHP, usando a *framework* Symfony. Disponibiliza uma API REST para integração.

### 4.2.2.8.3 Arquitetura

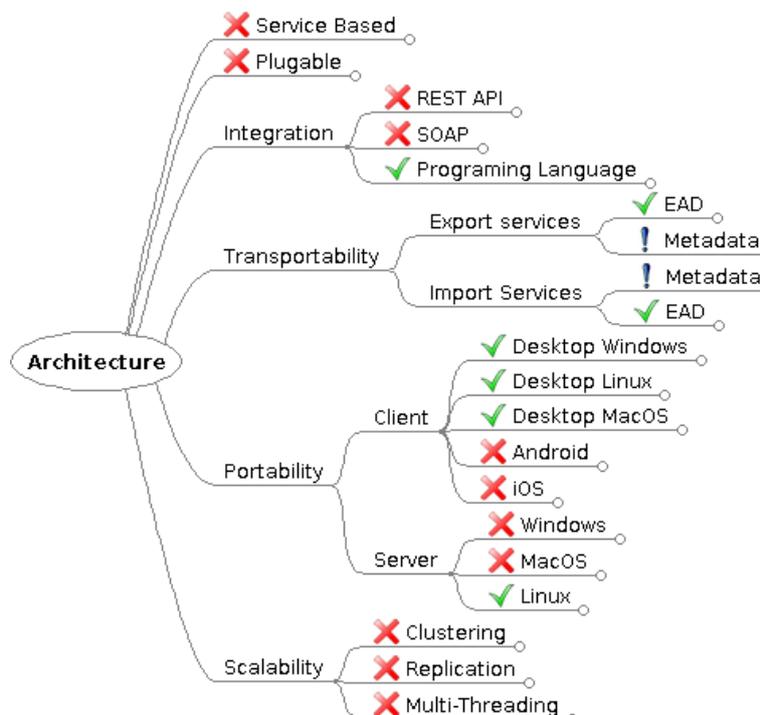


Figura 40: Análise QSOS - Maturidade do AtoM

O AtoM está desenvolvido com uma arquitetura monolítica. Não existe uma clara separação entre o *backend* e o interface web. O seu modelo de meta-dados não é expansível tornando-o limitado a descrever os registos de acordo com o Standard (quando por vezes se torna necessária a adição de informação própria da organização).

### Funcionalidades



Figura 41: Análise QSOS - Funcionalidades do AtoM

O AtoM implementa um interface web, gestão de utilizadores e permissões. O arquivo dos conteúdos funciona mas é algo limitado em termos de escalabilidade e garantia de integridade.

### Standards

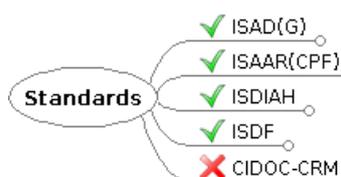


Figura 42: Análise QSOS - Standards implementados pelo AtoM

O AtoM implementa de forma muito completa as normas ISAD(G) e acessórias, definidas pelo Conselho Internacional de Arquivos. Não implementa a norma CIDOC-CRM.

## 4.2.2.9 Collective Access

O Collective Access é uma aplicação de Gestão de Coleções, desenvolvida para museus, arquivos e coleções especiais.

### 4.2.2.9.1 Maturidade

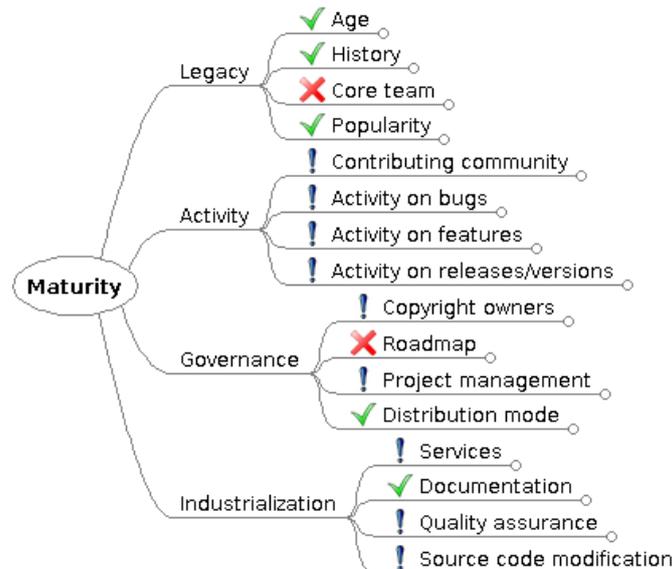


Figura 43: Análise QSOS - Maturidade do Collective Access

O Collective Access é uma aplicação usada por várias organizações sobretudo em Museus. Têm sido apresentadas várias versões sejam de evolução ou de correção de problemas. Apesar de uma comunidade participativa não existem evidências da existência de um *Roadmap* que demonstrem uma visão estratégica clara acerca da evolução do produto.

### 4.2.2.9.2 Princípios Operativos

O Collective Access está desenvolvido em PHP e usa algumas *frameworks*. O seu *front-end* está bastante bem arquitetado fazendo uso de CSS para definir a apresentação. Disponibiliza APIs para integração. O seu código está bastante bem estruturado.

### 4.2.2.9.3 Arquitetura

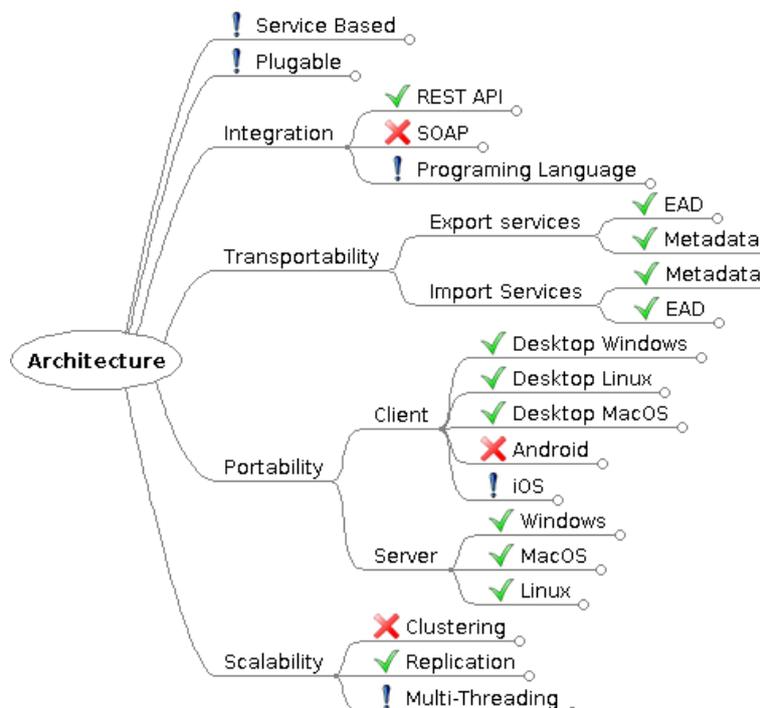


Figura 44: Análise QSOS - Maturidade do Collective Access

O Collective Access está bastante completo em termos de integração mas não existem evidências acerca de uma arquitetura avançada, principalmente em termos de *back-end*.

### Funcionalidades



Figura 45: Análise QSOS - Funcionalidades do Collective Access

O Collective Access implementa as principais funcionalidades de um sistema para Descrição Arquivística.

### Standards

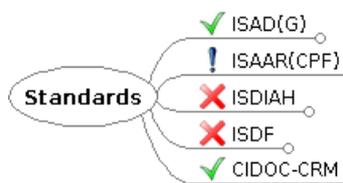


Figura 46: Análise QSOS - Standards implementados pelo Collective Access

O Collective Access é expansível, nomeadamente ao nível do modelo de meta-dados. Este facto torna-o versátil ao nível dos standards de descrição. Implementa a ISAD(G) e o CIDOC-CRM, não existindo, no entanto evidências da existência da implementação dos restantes standards acessórios do ICA.

## 4.2.2.10 Síntese

### 4.2.2.10.1 Comparação das Soluções

#### Maturidade

AtoM - Access To Memory 2.4.0 Collective Access 1.7.6 > Maturity

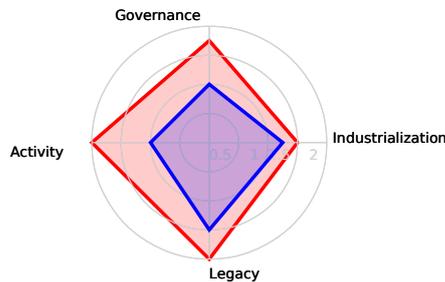


Figura 47: Análise QSOS - Comparação de Maturidades de Sistemas de Descrição Arquivística

O AtoM apresenta-se como uma solução mais madura tanto em termos de história como no que se refere aos processos, organizações e comunidades que possam levar à sua evolução futura.

#### Arquitetura

AtoM - Access To Memory 2.4.0 Collective Access 1.7.6 > Architecture

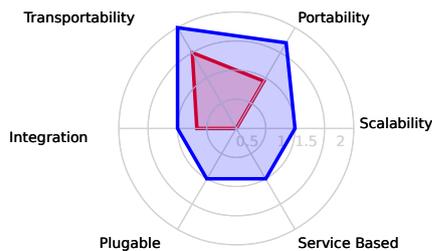


Figura 48: Análise QSOS - Comparação de Arquiteturas de Sistemas de Descrição Arquivística

As evidências apontam para que o AtoM esteja implementado com uma arquitetura mais completa, muito embora as recentes evoluções do Collective Access mostrem que este tem potencial para reduzir a diferença.

#### Funcionalidades

**AtoM - Access To Memory 2.4.0 Collective Access 1.7.6 > Functionality**

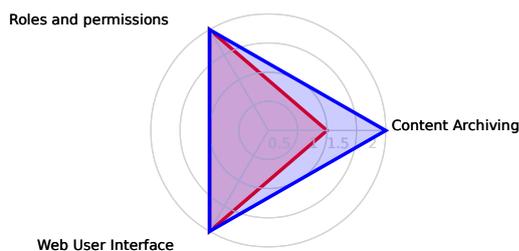


Figura 49: Análise QSOS - Comparação de Funcionalidades de Sistemas de Descrição Arquivística

Dado que as funcionalidades no caso da Descrição Arquivística são sobretudo definidas pelos standards, ambos os produtos implementam bastante bem as três funcionalidades definidas, com uma ligeira vantagem no caso do AtoM pela possibilidade de arquivar conteúdos.

### Standards

**AtoM - Access To Memory 2.4.0 Collective Access 1.7.6 > Standards**

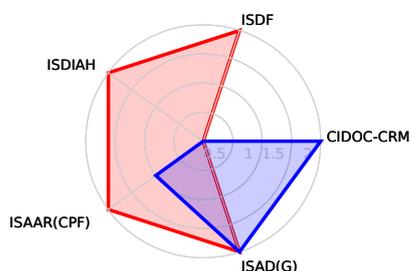


Figura 50: Análise QSOS - Comparação de aderência aos standards de Sistemas de Descrição Arquivística

O AtoM, por estar ligado ao ICA implementa os standards desta organização. O Collective Access implementa bem o CIDOC-CRM e já implementa a ISAD(G), mas não os standards acessórios.

### 4.2.2.11 Conclusão

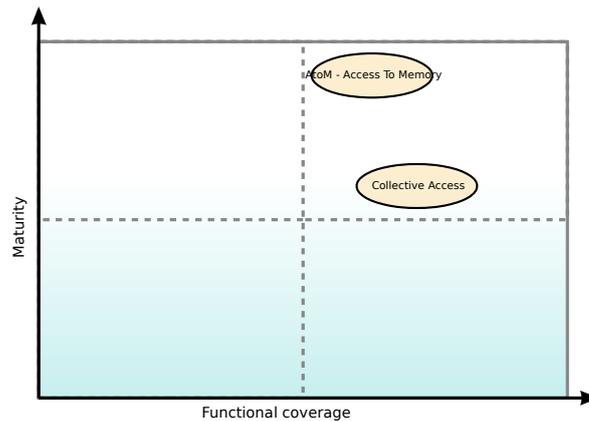


Figura 51: Análise QSOS - Conclusão das comparações de Sistemas de Descrição Arquivística

O AtoM apresenta-se como uma solução mais adequada aos standards do ICA, que são atualmente os mais usados pelos Arquivos.

## 4.2.3 Preservação Digital

A Preservação Digital preocupa-se com a garantia de acesso contínuo e a longo prazo, à informação e património que tem existência em formatos digitais.

Um Sistema de Preservação Digital implementa sobretudo o modelo de referência OAIS desenvolvido na sequencia da corrida espacial dos anos 60 do século passado.

### 4.2.3.1 Modelo Utilizado para Análise

Os critérios usados para avaliar as soluções descritas foram divididos nos seguintes eixos:

- ♦ Maturity: Maturidade do projeto em termos do desenvolvimento e manutenção do produto
- ♦ Architecture: Conceitos arquiteturais
- ♦ Functionality: Funcionalidades específicas de um Sistema de Preservação Digital
- ♦ Standards: Cumprimento das principais recomendações e boas práticas

### 4.2.3.2 Maturidade

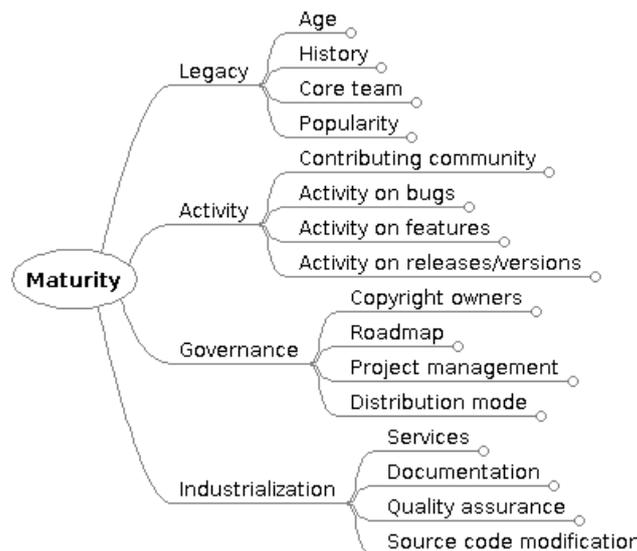


Figura 52: Critérios de Avaliação da Maturidade de Software de Gestão Documental para aplicação da metodologia QSOS

A avaliação da maturidade é fixa na metodologia QSOS. Avalia as seguintes dimensões:

- ♦ Legacy: História e herança do Projeto
- ♦ Activity: Atividade interna e à volta do Projeto
- ♦ Governance: Estratégia do Projeto
- ♦ Industrialization: Nível de industrialização do Projeto

### 4.2.3.3 Arquitetura

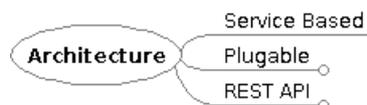


Figura 53: Critérios de Avaliação da Arquitetura de Software de Preservação Digital para aplicação da metodologia QSOS

- ♦ Service Based: O produto foi desenvolvido com o conceito de serviços autónomos cooperativos.
- ♦ Pluggable: O Software permite a inclusão de novos módulos.
- ♦ REST API: É disponibilizada uma API REST para integração.

#### 4.2.3.4 Funcionalidades

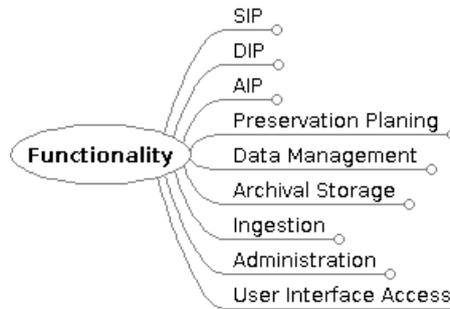


Figura 54: Critérios de Avaliação das Funcionalidades de Software de Preservação Digital

Na revisão de literatura, com base nas definições do Modelo de referência OAIS, identificaram-se as seguintes funcionalidades como importantes de ser avaliadas:

- ♦ SIP: O software implementa o formato de Pacote para Submissão da Informação
- ♦ DIP: É implementado o formato de Pacote para Disseminação da Informação
- ♦ AIP: Implementa o formato de Pacote para Arquivo da Informação
- ♦ Preservation Planing: Existem funcionalidades para Planeamento da Preservação.
- ♦ Data Management: São implementadas funcionalidades para Gestão de Metadados.
- ♦ Archival Storage: Os conteúdos podem ser armazenados no Sistema de Preservação Digital.
- ♦ Ingestion: Existe um processo para ingestão de SIPs
- ♦ Administration: Está implementado um módulo de administração
- ♦ User Interface Access: Além dos serviços, existe interface interativo

#### 4.2.3.5 Standards



Figura 55: Standards Implementados para análise QSOS

Na revisão de literatura identificaram-se como standard determinante o OAIS Reference Model / ISO 14721:2012.

#### 4.2.3.6 Soluções Identificadas

A identificação das várias soluções foi efetuada através de uma pesquisa sistemática realizada através da identificação de palavras chave, pesquisando referência a produtos e os respetivos projetos.

##### 4.2.3.6.1 Universo de Soluções

As soluções analisadas foram:

- ♦ Archivematica 1.7.1 (<https://www.archivematica.org/>)

Archivematica é um Sistema de Preservação Digital, implementado pela Artefactual Systems, Inc. A Artefactual é a empresa que implementou e mantém o AtoM. O Archivematica bebe toda a experiência adquirida e está arquiteturalmente bastante bem pensado. É usado por organizações em vários países.

- ♦ RODA 2.2.9 (<https://roda-community.org>):  
O Roda foi desenvolvido pela empresa Portuguesa Keep Solutions em parceria com a DGLab. É usado por algumas instituições, sobretudo em Portugal.
- ♦ Xena 6.1.0 (<http://xena.sourceforge.net/>)  
O Xena foi desenvolvido pelos arquivos nacionais da Austrália. Não é um software usado de uma forma extensiva.

As alternativas existentes não se mostraram credíveis como projetos que irão evoluir, capazes de conseguir um caminho autónomo, para além das organizações para as quais foram desenvolvidos.

### 4.2.3.7 Archivematica 1.7.1

O Archivematica já conta com algumas versões durante os seus 6 anos de idade. É usado por organizações importantes como por exemplo o Museu de Arte Moderna de New York – MOMA.

#### 4.2.3.7.1 Análise do Projeto



Figura 56: Análise QSOS - Maturidade do Archivematica

O Archivematica é uma solução que se pode considerar madura já que é usada com sucesso em várias instituições. Tem uma empresa de suporte ao seu desenvolvimento e evolução e conta com um *roadmap* que permite compreender que existe uma Visão Estratégica clara quanto à sua evolução.

#### 4.2.3.7.2 Princípios Operativos

O Archivematica está desenvolvido em PHP. A sua arquitetura implementa micro-serviços e integra com ferramentas externas, especializadas para tratar problemas específicos (ex: conversão de formatos de vídeo).

#### 4.2.3.7.3 Arquitetura

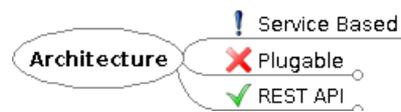


Figura 57: Análise QSOS - Maturidade do Archivematica

A arquitetura, apesar de ser de micro-serviços, não apresenta uma definição clara dos serviços como entidades autónomas, não permitindo a sua expansão com serviços adicionais. Disponibiliza acesso através de APIs REST.

#### Funcionalidade



Figura 58: Análise QSOS - Funcionalidades do Archivematica

O Archivematica implementa todos os principais conceitos envolvidos no modelo de referência OAIS.

#### Standards



Figura 59: Análise QSOS - Standards Implementados pelo Archivematica

O Archivematica implementa o Modelo de Referência OAIS de forma bastante completa.

### 4.2.3.8 RODA 2.2.9

O RODA é usado em produção por várias organizações, com destaque para a Direção Geral do Livro, Arquivos e Bibliotecas. Foi desenvolvido em estreita colaboração entre a empresa Keep Solutions e a DGLab.

#### 4.2.3.8.1 Análise do Projeto

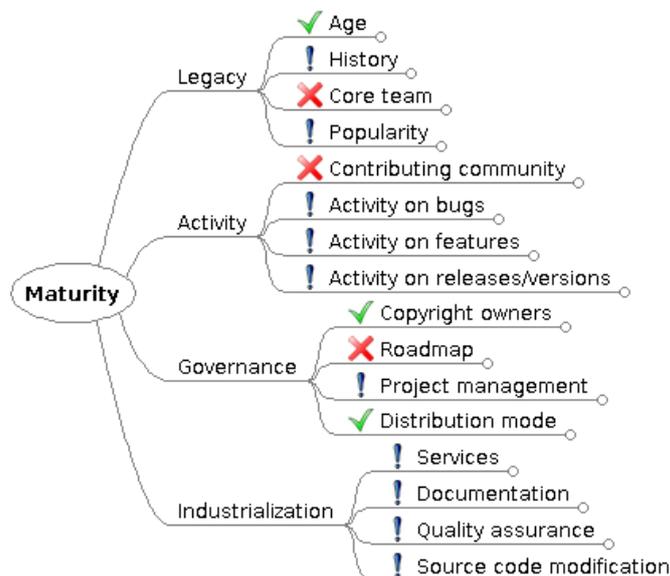


Figura 60: Análise QSOS - Maturidade do RODA

Não existem evidências da existência de uma visão e de um *roadmap* para a sua evolução. A gestão do projeto parece algo difusa. A sua popularidade ocorre sobretudo em Portugal.

#### 4.2.3.8.2 Princípios Operativos

O RODA está sobretudo desenvolvido em Java, com código bastante legível e fácil de compreender. Implementa serviços para integração. Tem um *front-end web* para operação.

#### 4.2.3.8.3 Arquitetura



Figura 61: Análise QSOS - Maturidade do RODA

Embora seja desenvolvido em Java, não se conseguiram identificar de forma clara (através da documentação) a existência do conceito de serviços, muito embora exista a possibilidade limitada de poder ser adicionados módulos de expansão.

### Funcionalidades



Figura 62: Análise QSOS - Funcionalidades do RODA

O RODA implementa todos os principais serviços definidos pelo Modelo de Referência OAIS.

### Standards



Figura 63: Análise QSOS - Standards Implementados pelo RODA

O RODA implementa o Modelo de Referência OAIS.

## 4.2.3.9 Xena 6.1.0

O XENA, apesar de ser Software livre, está desenvolvido sobretudo com o objetivo de ser usado nos arquivos Australianos.

### 4.2.3.9.1 Análise do Projeto

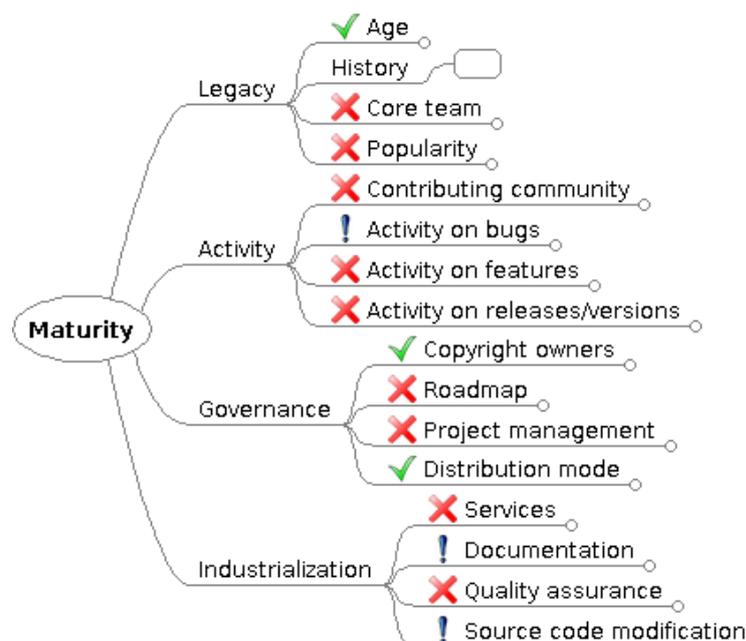


Figura 64: Análise QSOS - Maturidade do Xena

Desta forma, não se vislumbra uma visão, um *roadmap* e uma liderança clara. Não se nota evolução nos últimos anos.

### 4.2.3.9.2 Princípios Operativos

O Xena foi desenvolvido na plataforma Java. O seu código está mal documentado e não se nota uma estruturação clara. Usa um sistema de *build* antigo, não se notando adaptação a versões atuais de software de infraestrutura.

### 4.2.3.9.3 Arquitetura

## Arquitetura

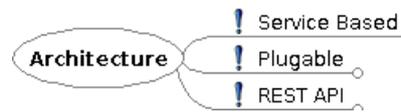


Figura 65: Análise QSOS - Arquitetura do Xena

A arquitetura do Xena foi obtida a partir da análise do código fonte. Parece existir o conceito de serviços, de poder ser expansível e de existir uma API REST para integração.

## Funcionalidade



Figura 66: Análise QSOS - Funcionalidades do Xena

Apesar de ser um produto de preservação digital, não parece implementar de forma completa o Modelo de Referência OAIS. Alguns dos conceitos, como a ingestão e a existência de interface com o utilizador estão presentes,

## Standards



Figura 67: Análise QSOS - Standards Implementados pelo Xena

O Xena não implementa o Modelo de Referência OAIS.

### 4.2.3.10 Síntese

#### 4.2.3.10.1 Comparação das Soluções

## Maturidade

Archivematica 1.7.1 RODA 2.2.9 Xena 6.1.0 > Maturity

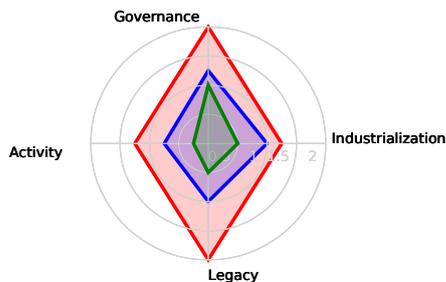


Figura 68: Análise QSOS - Comparação de Maturidades de Sistemas de Preservação Digital

O Archivematica é o Software Livre que maior maturidade apresenta, tanto em termos de quem o suporta, como de utilização ou da forma como está organizado como projeto.

## Arquitetura

Archivematica 1.7.1 RODA 2.2.9 Xena 6.1.0 > Architecture

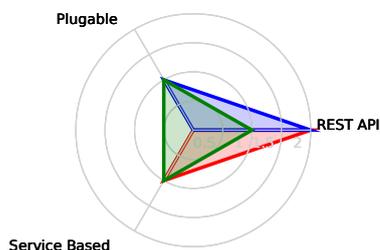


Figura 69: Análise QSOS - Comparação de Arquiteturas de Sistemas de Preservação Digital

A arquitetura do Archivematica, muito embora lhe falte a definição clara de serviço, é a que melhor cumpre os requisitos.

## Funcionalidades

Archivematica 1.7.1 RODA 2.2.9 Xena 6.1.0 > Functionality

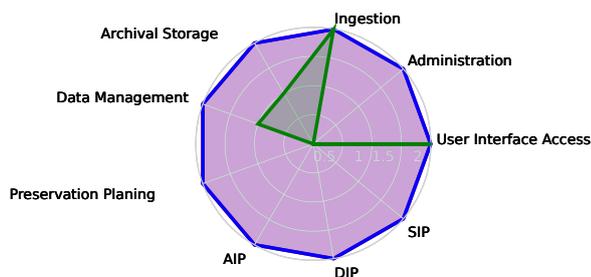


Figura 70: Análise QSOS - Comparação de Funcionalidades de Sistemas de Preservação Digital

Tanto o RODA como o Archivematica implementam as principais funcionalidades de forma completa.

## Standards

Archivematica 1.7.1 RODA 2.2.9 Xena 6.1.0 > Standards

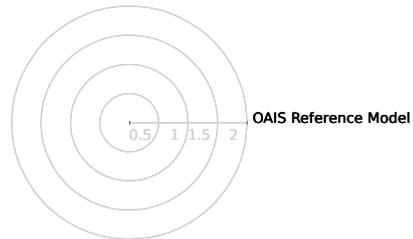


Figura 71: Análise QSOS - Comparação de aderência aos standards de Sistemas de Preservação Digital

O RODA e o Archivematica implementam o Modelo de Referência OAIS.

### 4.2.3.10.2 Conclusão

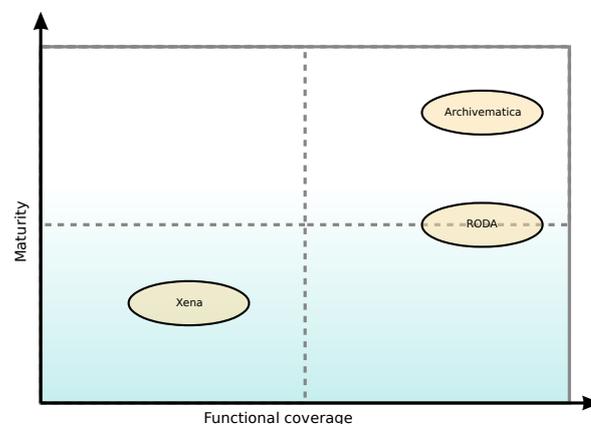


Figura 72: Análise QSOS - Conclusão das comparações de Sistemas de Preservação Digital

O Archivematica é o Software Livre ou de Código Aberto que melhor se adequa a ser implementado num sistema global de Gestão de Conteúdos de Arquivo.

## 4.2.4 Componentes para a Gestão de Documentos de Arquivo

A concretização da seleção dos produtos a utilizar como componentes principais, aplicando a metodologia QSOS, anteriormente documentada.

No quadro que se segue, apresentam-se os produtos sobre os quais incidiu a seleção, tanto para o caso das componentes principais, com a metodologia QSOS, como para os produtos acessórios da Gestão de Documentos de Arquivo.

**Tabela 19: Componentes Aplicacionais para a Gestão de Documentos de Arquivo**

Componente	Produto
Gestão Documental	Alfresco Community («Alfresco Community», 2018)
	LogicalDoc («LogicalDoc», 2019)
	Nuxeo («Nuxeo», 2019)
	OpenKM (OpenKM, 2019)
Descrição Arquivística	AtoM («AtoM: Open Source Archival Description Software», 2019)
	Collective Access («Collective Access», 2019)
Preservação Digital	Archivematica («Archivematica», 2019)
	RODA («RODA - Repositório de Objetos Digitais Autênticos», 2019)
	Xena («Xena - Digital Preservation Software», 2019)
Assinatura Digital	.be eID Sign («.be eID Sign (RedTree)», 2019)
	DocuSign Connector for Alfresco («DocuSign Connector for Alfresco», 2019)
	Esign-cert ( <i>Alfresco client certificate signature (including Chrome)</i> , 2016/2019)
OCR	Open-ocr (Leyden, 2014/2019)
	Tesseract («tesseract-ocr», 2019)
	GOOCR («GOOCR», 2019)
	Cuneiform («OpenOCR», 2019)
Deteção de Layouts	OpenCV + OpenCV-python-web-services («OpenCV», 2019)
	OCROpus (Tom, 2014/2019)
Indexação e Pesquisa	Elasticsearch + Elasticsearch-oai (Prante, 2012/2019)
Transformação de Formatos	Alfresco
	Archivematica

#### 4.2.4.1 Gestão Documental

A investigação realizada para pesquisa de ferramentas que implementem Gestão Documental derivou numa lista relativamente extensa. Com base na análise das tendências e dos seus sites, foram selecionadas para avaliação QSOS: Alfresco Community («Alfresco Community»,

2018); LogicalDoc («LogicalDoc», 2019); Nuxeo («Nuxeo», 2019); OpenKM (OpenKM, 2019).

Finalmente, após a aplicação da metodologia QSOS, o Alfresco Community foi o produto selecionado como componente de Gestão Documental do Sistema de Gestão de Documentos de Arquivo.

O *Alfresco* foi selecionado por implementar as principais funcionalidades de sistemas de ECM, incluindo RM, por ter uma comunidade muito ativa, ser desenvolvido com ferramentas de infraestrutura adequadas, maduras e usadas por um grande número de comunidades em muitos produtos, e por ser suportado por uma empresa que já demonstrou ser capaz de angariar os fundos necessários à evolução do seu sistema.

O Alfresco implementa as principais funcionalidades de um programa para Gestão Documental, nomeadamente: Repositório de Arquivo de Conteúdos, Gestão de Registos, Gestão de Processos. Estas funcionalidades de alto nível derivam de uma visão clara das definições dos principais normativos (ISO, 2016), sendo complementada com uma clara visão de existência de uma camada prestação de serviços e de uma camada de apresentação, que atualmente é disponibilizada sob a forma de *framework* autónoma implementada com uma das referências de software de apresentação.



Figura 73: Arquitetura Alfresco

A versatilidade e abertura de tal arquitetura facilita a integração com o outros sistemas.

O Alfresco é implementado sob a forma de uma aplicação servidora, extensível e com um elevado grau de configuração, sobre a qual é possível instalar módulos que implementam funcionalidades específicas. Esta componente serve base de implementação ao que se chamou de *Content Services*. Os *Content Services* implementam todas as necessárias funcionalidades para o arquivo de conteúdos, nomeadamente: o armazenamento dos ficheiros, a gestão de

utilizadores e grupos, a proteção dos conteúdos com permissões, a sua descrição com modelos de meta-dados que podem ser criados pelos utilizadores (de origem vêm pré-instalados alguns standards como o Dublin Core («DCMI», 2019) ou o Exif («EXIF.org», 2019)). As funcionalidades são expostas através de API(s) permitindo a criação de interfaces com o utilizador ou com outros programas. As API(s) são implementadas em REST (sendo por isso agnósticas quanto ao sistema que as consome), em Java ou em Javascript, sendo desta forma muito abrangentes quanto à integração com outros sistemas ou camadas de apresentação.

O recente desenvolvimento da *Application Developer Framework – ADF*, surge de acordo com o conceito de que sistemas como os de ECM, devem poder ser facilmente integrados nos sistemas verticais das organizações, ao invés de possuírem um interface próprio especializado e autónomo. No passado o Alfresco implementava o seu próprio interface com o utilizador (Alfresco Share), expansível e no qual se podia instalar componentes aplicativos. Atualmente com as APIs e o ADF, as componentes específicas de gestão de documentos podem ser implementadas dentro das aplicações de negócio. Esta visão descreve uma arquitetura moderna e é uma das razões pela qual o Alfresco foi escolhido como implementação de ECM para estudo nesta dissertação.

Além disso, as várias componentes do Alfresco suportam-se em ferramentas que podem ser facilmente consideradas de *Commodities*. O facto de a sua componente servidora ser implementado em Java, torna extremamente fácil a sua utilização em qualquer um dos atuais principais Sistemas Operativos, nomeadamente MS-Windows, Linux, MacOS entre outros. O armazenamento dos meta-dados é implementado recorrendo a uma base de dados relacional, cujo acesso é virtualizado por uma camada de abstração (Apache myBatis («MyBatis», 2019), sendo possível e fácil a troca entre alguns dos principais SGBDs (MariaDB, MariaDB, Postgresql, Oracle, DB2 entre outros).

#### **4.2.4.2 Descrição Arquivística**

Durante a revisão de literatura estudaram-se as normas de Descrição Arquivística. Dada a sua utilização mais abrangente, a norma ISAD(G) do Conselho Internacional de Arquivos é a que foi selecionada para a proposta de solução.

A seleção final da componente de Descrição Arquivística foi realizada entre os produtos AtoM («AtoM: Open Source Archival Description Software», 2019) e Collective Access («Collective Access», 2019).

Dos vários sistemas estudados, escolheu-se o AtoM, devido à associação ao ICA («ICA», 2019) e à Unesco, à lista de referências de instituições que o utilizam, à existência de fortes comunidades em vários países, ao tipo de licenciamento e à existência de uma empresa que desenvolve e suporta o produto, demonstrando maturidade e massa crítica suficiente para garantir que este se irá manter em evolução e manutenção.

#### **4.2.4.3 Preservação Digital**

A investigação de produtos que pudessem ser usados como componentes de Preservação Digital, ao início parecia bastante extensa. Finalmente compreendemos que muitas delas são soluções programadas para uma única organização sem preocupações de se tornarem em soluções replicáveis. No final, a análise QSOS foi efetuada sobre: Archivemática («Archivemática», 2019); RODA («RODA - Repositório de Objetos Digitais Autênticos», 2019) e Xena («Xena - Digital Preservation Software», 2019).

De entre elas, foi selecionado o software Archivemática pela sua arquitetura baseada em micro-serviços, abertura, referências de utilização e existência de uma empresa de suporte bem organizada e presente no mercado global.

#### **4.2.4.4 Assinatura Digital**

A Assinatura Digital tem especial importância no contexto do módulo de Gestão Documental, dado que é aqui que a garantia de autenticidade, integridade e não repúdio tem mais importância.

Assim, a investigação dos produtos foi muito focada em implementações que funcionem no Alfresco. As soluções analisadas foram: a Belga .be eID Sign («be eID Sign (RedTree)», 2019), O conector do DocuSign («DocuSign Connector for Alfresco», 2019) para Alfresco e o projeto esign-cert (*Alfresco client certificate signature (including Chrome)*, 2016/2019), tendo sido selecionado este último por ser o que mais facilmente trabalha com os Certificados existentes no Cartão do Cidadão Português.

#### 4.2.4.5 OCR

Para reconhecimento de caracteres nas imagens de documentos digitalizados escolheu-se primeiro o Tesseract («tesseract-ocr», 2019). A análise incidu sobre os produtos Tesseract; open-ocr («OpenOCR», 2019), GOCR («GOCR», 2019) e cuneiform («OpenOCR», 2019).

O open-ocr foi escolhido porque além de integrar o motor já muito testado e maduro Tesseract, implementa escalabilidade horizontal e micro-serviços tornando-se assim fácil de integrar com as restantes componentes.

#### 4.2.4.6 Detecção de Layouts

Para detetar Layouts foi escolhido o software *OpenCV* («OpenCV», 2019), que usa mecanismos de *Computer Vision* para detetar objetos em imagens estáticas ou em movimento. O *OpenCV* foi originalmente desenvolvido pela Intel, disponibilizado atualmente com licenciamento BSD.

Com o OpenCV é fácil definir uma imagem como base, efetuando comparação com as imagens de que se pretende detetar o Layout, obtendo um grau de similitude quando se efetuar a comparação pela linha de comando. Para efeitos de inserção na arquitetura concreta utilizou-se de uma versão modificada do pacote “OpenCV Python Web Services”.

Desta forma, tornou-se possível a execução sob a forma de serviços REST, da validação de graus de similitude com imagens de referência.

#### 4.2.4.7 Indexação e Pesquisa

Atualmente, os utilizadores de sistemas de informação cada vez mais desejam um ponto único de pesquisa sobre a informação da sua organização (e eventualmente externa). Acontece que a fragmentação em sistemas especializados dificulta esta necessidade, de um ponto único de pesquisa. Assim, existem sistemas, cuja denominação mais comum atualmente é *Enterprise Search*, que indexam e permitem pesquisas sobre os diferentes sistemas existentes.

A investigação levou à seleção do *ElasticSearch* («Open Source Search & Analytics · Elasticsearch», 2019) como motor de indexação e pesquisa, adicionando-lhe componentes *Free e Open Source*, nomeadamente os coletores OAI:PMH do projeto *elasticsearch-oai*, por forma a permitir a utilização de protocolos de pesquisa já implementados nas ferramentas de Gestão Documental e de Arquivo.

O *Elastic Search* está implementado em Java, sob a forma de uma componente web, à qual se pode aceder através de um interface próprio, mas também sendo possível efetuar pesquisas através de uma API REST disponibilizada.

Desta forma, o *ElasticSearch* posiciona-se na arquitetura, como o ponto central de pesquisa dos conteúdos, sendo responsável pela coleta e indexação da informação em cada um dos sistemas especializados.

#### 4.2.4.8 Transformação de Formatos

A transformação de formatos dos ficheiros que constituem os documentos é suportada em primeira instância por ferramentas diversas que existem implementadas no Linux, como o ImageMagic («ImageMagick», 2019), ffmpeg («FFmpeg», 2019), JodConverter (Braconnier, 2016/2019), inkscape («Inkscape», 2019), entre outros.

Tanto dentro do Alfresco, como do Archivemática, as ferramentas respetivamente de Gestão Documental e Preservação Digital, existem serviços que efetuam as transformações entre formatos. Estão disponíveis sob a forma de serviços REST tendo-se optado por não criar qualquer serviço de fachada sobre estes.

#### 4.2.5 Componentes de Infraestrutura

No quadro que se segue, apresentam-se os produtos sobre os quais incidiu a seleção de componentes de infraestrutura.

**Tabela 20: Componentes de Infraestrutura**

Componente	Produto
Sistema Operativo	Linux Debian («Debian -- The Universal Operating System», 2019)
	Linux Ubuntu («Ubuntu - The leading operating system for PCs, IoT devices, servers and the cloud», 2019)
	Linux Fedora («Fedora», 2019)
Virtualização	Kvm («KVM», 2019)
	Oracle VirtualBox («Oracle VM VirtualBox», 2019)
Orquestração de Contentores	Kubernetes («Kubernetes : Production-Grade Container Orchestration», 2019)
	Docker Swarm («Docker Swarm», 2019)
Autenticação e Diretório de	OpenLDAP («OpenLDAP», 2019)

Componente	Produto
Utilizadores	ApacheDS («ApacheDS — Apache Directory», 2019)
Sistema de Gestão de Bases de Dados Relacionais	MariaDB («MariaDB.org», 2019)
	PostgreSQL («PostgreSQL: The world's most advanced open source database», 2019)
	Firebird («Firebird», 2019)
Filas de Mensagens Assíncronas	ActiveMQ («ActiveMQ», 2019)
	RabbitMQ («Messaging that just works — RabbitMQ», 2019)
	Apache Kafka («Apache Kafka», 2019)
Micro-Serviços	Spring Boot («Spring Boot», 2019)
Indexação	ElasticSearch («Open Source Search & Analytics · Elasticsearch», 2019)
	Solr («Apache Solr», 2019)

#### 4.2.5.1 Sistema(s) Operativo(s)

Embora existam vários excelentes Sistemas Operativos disponibilizados com Licença de Software Livre, o eco-sistema de software associado ao Linux é de longe o mais completo. Uma investigação acerca da existência dos pacotes que servirão como componentes do Sistema de Gestão de Documentos de Arquivo para os Sistemas BSD e Solaris, demonstrou que uma grande percentagem teriam de ser compilados e testados nestes ambientes, enquanto que em algumas distribuições de Linux, praticamente todos existem.

Também a questão do suporte se colocou: as empresas de suporte aos produtos que implementam as componentes principais (Gestão Documental, Descrição Arquivística, Preservação Digital), permitem o funcionamento destas plataformas em Linux e Windows.

Assim, o Linux é o Sistema Operativo selecionado. Uma vez que existem muitas distribuições de Linux (empacotamento do Linux com um programa de instalação e um sistema de pacotes que instalam componentes), e após a investigação das componentes necessárias, foram selecionadas como possíveis de usar nas componentes do Sistema (por ordem de prioridade sempre que a componente não identificar qual a distribuição preferencial):

- Debian Linux
- Ubuntu Linux
- Fedora Linux

Os Sistemas Operativos selecionados foram usados como suporte da plataforma de virtualização, ou seja como Sistema Operativo Hospedeiro, mas também como Sistemas Operativos convidados e também como base dos Contentores.

#### **4.2.5.2 Virtualização**

As duas principais tecnologias de virtualização para arquiteturas x86, disponíveis com licença de Software Livre ou de Código Aberto é o KVM/QEMU e o Oracle VirtualBox.

Foi escolhido KVM/QEMU devido à sua facilidade de utilização pela linha de comando e ao facto de o VirtualBox ter opções, nomeadamente de configuração, que apenas são disponibilizadas em versões com suporte pago.

O KVM tem a vantagem de poder ser incluído como módulo do kernel de Linux, sendo disponibilizado um interface através do `/dev/kvm`. Além disso interage com a libvirt, que é uma API que permite a interação com múltiplas soluções de virtualização e ao qual ferramentas como o `virsh` e(ou) o `virt-manager` se ligam para realizar operações sobre as Vms, como ligar, desligar, remover, criar, entre outros.

#### **4.2.5.3 Orquestração de Contentores**

Os contentores são usados em grande medida nas componentes com mais necessidade de escalabilidade horizontal, podendo aumentar ou diminuir a quantidade de instâncias dos micro-serviços sempre que necessário, sendo a carga ser distribuída pelas instâncias que estiverem desocupadas.

O software de Orquestração de Contentores é responsável por operar esta gestão de crescimento ou diminuição sempre que necessário.

Existem várias soluções disponibilizadas com licença de Software Livre ou de Código Aberto, no entanto as que se destacam, pela quantidade de utilizadores e(ou) pela capacidade de evolução das empresas que lhe dão suporte é o Kubernetes , desenvolvido pela Google e o Docker Swarm desenvolvido pela Docker, que é a tecnologia de contentores atualmente mais usada. A seleção do Kubernetes deve-se à muito maior disponibilidade e qualidade de informação acerca da sua configuração e utilização.

#### **4.2.5.4 Autenticação e Diretório de Utilizadores**

Para registrar os utilizadores por forma a que estes não tenham de ser registados em cada um dos módulos, usa-se o protocolo LDAP.

Dos vários produtos que poderiam implementar esta funcionalidade destacam-se o OpenLDAP e o ApacheDS sobre os quais foi realizada a seleção final.

A maior utilização, quantidade de referências e qualidade da documentação ditou a utilização do OpenLDAP.

#### **4.2.5.5 Sistemas de Gestão de Bases de Dados Relacionais**

Também a seleção do Sistema de Gestão de Bases de Dados relacionais foi influenciada pelo suporte por parte dos produtos que implementam as três principais componentes. E neste particular foram escolhidos o MariaDB e o PostgreSQL, ficando o Firebird de fora devido à maior complexidade das ferramentas.

#### **4.2.5.6 Filas de Mensagens Assíncronas**

A comunicação entre componentes, que, sempre que possível é efetuada através da utilização de Sistema de Mensagens que implementa o padrão publish-subscribe e request-reply.

Neste caso a escolha recaiu entre o Apache Kafka, o ActiveMQ e o RabbitMQ. O maior desempenho e escalabilidade do Kafka foi esbatido devido à utilização de um protocolo único, próprio, o que dificulta a integração por obrigar a desenvolver todos os *endpoints*. A questão dos protocolos foi também a razão pela qual o ActiveMQ foi a ferramenta de files de mensagens selecionada uma vez que implementa os principais protocolos como o JMS, AMQP.

#### **4.2.5.7 Micro-Serviços**

Embora no caso da implementação dos micro-serviço não exista propriamente uma seleção, pois tal iria contra a lógica dos micro-serviços de poderem ser desenvolvidos em tecnologias e ferramentas diversas, sempre que possível será usado o Spring Boot, uma vez que simplifica bastante o seu desenvolvimento ao permitir que se consiga desenvolver um micro-serviço com muito poucas linhas de código encarregando-se a ferramenta das tarefas normais como o acesso à filas de mensagens, implementação de um servidor e utilização de um *container web*.

### 4.2.5.8 Indexação

A indexação, sobretudo para a implementação da pesquisa federada é implementada usando o ElasticSearch em vez do também completo SOLR, A sua escalabilidade e ao mesmo tempo simplicidade, aliada à enorme comunidade, documentação disponível mas também a existência de um coletor OAI:PMH fez dele a escolha óbvia.

## 4.3 Avaliação Preliminar

Os resultados da avaliação dos produtos, que irão instanciar o modelo conceptual, constituindo-se como uma possível solução para o problema resumem-se de acordo com as seguintes tabelas:

**Tabela 21: Resumo de Principais Avaliações de Software de Gestão Documental**

Avaliação	Alfresco	LogicalDoc	Nuxeo	OpenKM
Maturidade	Muito Maduro	Maduro	Muito Maduro	Pouco Maduro
Arquitetura	Completa	Pouco escalável	Pouco integrável	Pouco integrável
Serviços do Sistema	Implementa os principais requisitos	Não implementa RM	Não implementa RM	Não implementa RM
Funcionalidades	Implementa as principais funcionalidades	Não implementa permissões	Não implementa classificação	Não implementa permissões

**Tabela 22: Resumo de Principais Avaliações de Software de Descrição Arquivística**

Avaliação	AtoM	Collective Access
Maturidade	Muito Maduro	Maduro
Arquitetura	Completa	Ainda deficitária
Funcionalidades	Todos os Serviços necessários	Não arquiva conteúdos
Standards	Implementa bem a ISAD(G)	Implementa a ISAD(G) e o CIDOC-CRM

**Tabela 23: Resumo de Principais Avaliações de Software de Preservação Digital**

Avaliação	Archivematica	Roda	Xena
Maturidade	Muito Maduro	Utilização limitada	Sem evolução

Avaliação	Archivematica	Roda	Xena
Arquitetura	Completa	Completa	Desatualizado
Funcionalidades	Todas	Todas	Limitadas
Standards	Modelo OAIS completo	Modelo OAIS completo	Modelo OAIS completo

Destas tabelas é possível para já, concluir que existem produtos com os quais se consegue instanciar uma implementação do modelo conceptual, ou seja: implementar uma solução para o problema da implementação de um Sistema de Gestão Documental usando software livre ou de código aberto.

## **4.4 Aproximação aos normativos**

A solução encontrada implementa funcionalmente as principais normas e boas praticas das três sub-áreas, Gestão Documental, Descrição Arquivística e Preservação Digital. Neste ponto elabora-se acerca da comparação entre cada um dos produtos seleccionados e as normas mais relevantes.

### **4.4.1 Alfresco vs MoReq2010**

O MoReq2010 vai mais além do que grande parte das normas, já que se apresenta como um Modelo de Requisitos, o que significa que define um conjunto de requisitos, funcionais, não funcionais e arquiteturais que permitem avaliar até que ponto um produto de Gestão Documental implementa o Modelo.

Para efeitos desta dissertação, foi usada a lista de requisitos publicada por Rafael António para avaliar qual o grau de cumprimento do Alfresco em relação ao MoReq2010 (António & Silva, 2012). Os resultados avaliação completa são apresentados no Anexo 2 .

Os resultados consolidados apresentados na avaliação demonstram que Alfresco, muito embora não implemente completamente todos os requisitos definidos pelo MoReq2010, implementa a maior parte das funcionalidades por ele definidas.

#### **4.4.2 Alfresco vs ISO 15489**

O Alfresco, complementado com o módulo de *Records Management* implementa grande parte da norma ISO 15489, nomeadamente:

- A criação de Registos;
- As políticas e responsabilidades;
- A avaliação usando as tabelas de seleção;
- Controlo do registo, concretamente: esquemas de meta-dados sobre os registos, plano de classificação, regras de acesso e permissões, autorização de destruição;
- Processos para criação, captura e gestão dos registos, concretamente: criação, captura, indexação, classificação, controlo de acessos, armazenamento, utilização, destruição.

A maior falha do Alfresco, no que se refere à ISO 15489, que pode no entanto facilmente ser colmatada, consiste na implementação de standards de migração e conversão de registos

#### **4.4.3 Alfresco vs MEF**

A MEF, Macro Estrutura Funcional, definida pela Direção Geral do Livro Arquivos e Bibliotecas consiste, em termos práticos, na definição concreta de um plano de classificação hierárquico com pelo menos 3 níveis (Lourenço, Henriques, Penteado, 2013).

O módulo RM do Alfresco permite, sem qualquer modificação, a implementação de planos de classificação com vários níveis, permitindo a cada nível a associação de um modelo de meta-informação próprio.

Desta forma, considera-se que o Alfresco permite implementar completamente a MEF.

#### **4.4.4 Alfresco vs MIP**

A MIP, Meta Informação para a Interoperabilidade consiste na definição de 17 elementos de meta-informação. O seu objetivo primário é permitir a interoperabilidade de documentos entre organismos.

No Alfresco após instalação não disponibiliza modelo de meta-dados que implemente a MIP.

No entanto, trata-se de um produto que permite facilmente a adição de modelos de meta-informação através da criação de Aspetos (classes de meta-dados que podem ser associados aos documentos), que consistem na definição em XML dos elementos que compõe um ou

mais conjuntos de meta-dados que descrevem um documento. As características da definição deste modelo, incluindo validação de tipos de dados, constrangimentos por lista ou por código e validações cruzadas permitem concluir que é trivial a configuração do Alfresco para implementar este normativo.

#### **4.4.5 AtoM versus ISAD(G)**

A ISAD(G) e as outras normas associadas – ISAAR(CPF), ISDIAH, entre outros - foram definidas pelo Conselho Internacional de Arquivos. Consistem em Princípios Orientadores, Regras Gerais e definição de modelos de meta-dados, concretamente dividido em sete zonas descritivas.

A empresa Archivemática desenvolveu o produto AtoM de tal forma em estreita ligação com o ICA, que o produto inicialmente se denominava de ICA-AtoM. Este fato poderia não ser relevante, no entanto a utilização do nome teve como objetivo associar o nome do ICA a uma implementação que, segundo a organização implementa os normativos por si definidos.

De facto, o produto foi construído tendo a ISAD(G) como base e construindo-se à sua volta, implementando completamente as sete zonas descritivas do modelo e cumprindo princípios orientadores e regras gerais.

Desta forma, o AtoM é um sistema informático que implementa completamente a norma ISAD(G).

#### **4.4.6 ArchiveMática vs Modelo de Referência OAIS**

O produto Archivemática foi construído para ser, desde o seu início, um Sistema conforme o modelo OAIS (ISO 14721).

Inclui a implementação completa do dicionário de dados PREMIS versão 3. Usa os formatos definidos pelo METS, nomeadamente para os AIP(s).

Assim, considera-se que o Archivemática é um Sistema de Preservação Digital que implementa o modelo OAIS de forma muito completa.

### **4.5 Arquitetura Concreta**

Após definir o Conceito Arquitetural a implementar para um sistema global de gestão de conteúdos de arquivo, complementando-o com um conjunto de ‘software’ selecionado,

construiu-se uma arquitetura em que as peças escolhidas são devidamente posicionadas, com adaptações derivadas das suas especificidades.

Desta forma, a arquitetura concreta, devidamente descrita com a Linguagem de descrição de Arquiteturas *Archimate*, apresenta-se na figura que se segue.

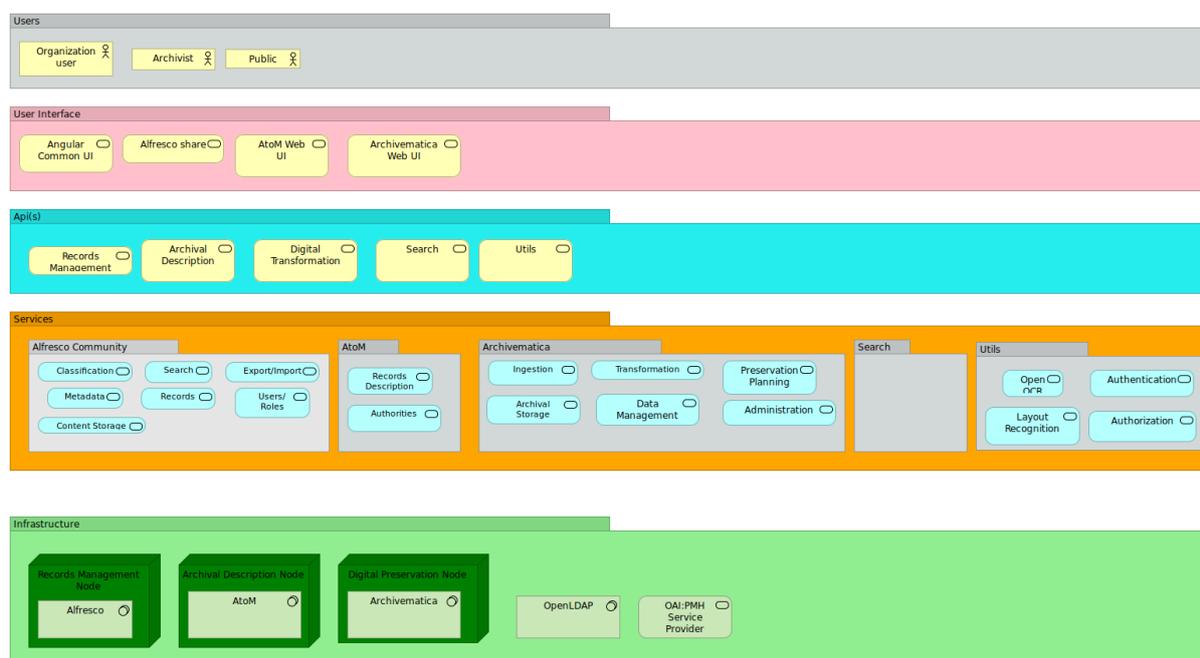


Figura 74: Arquitetura Concreta da Solução

## 4.5.1 Resumo de Produtos

A tabela que se segue apresenta os produtos, principais ou acessórios de gestão de conteúdos selecionados para implementação do protótipo desta dissertação.

**Tabela 24: Produtos de Software Selecionados para Gestão de Documentos de Arquivo**

Produto/Projeto	Classe	Função	Descrição
Alfresco Content Services	Core	Gestão Documental	Repositório de Conteúdos
Alfresco Records Management	Core	Records Management	Gestão de Registos
AtoM	Core	Descrição Arquivística	Sistema com interface web que implementa a ISAD(G)
Archivematica	Core	Preservação Digital	Sistema de PD em PHP que implementa o modelo de referência OAIS
esign-cert	Acessório GD	Assinatura Digital	Permite assinar digitalmente documentos no Alfresco. Suporta smart card como o cartão do cidadão.
Tesseract	Acessório GD	OCR	Reconhecimento Ótico de Caracteres em documentos digitalizados para imagem
OpenCV + OpenCV-python-web-services	Acessório GD	Layout Detection	Deteção do tipo de documento (ex: fatura de gasolina) com base num modelo previamente analisado
Elastic Search + elasticsearch-oai	Acessório GD	Enterprise Search	Coletar com o protocolo OAI:PMH, indexar e permitir uma pesquisa centralizada

Para suporte dos produtos atrás definidos, uma possível infraestrutura de referência, que foi implementada no protótipo faz uso dos seguintes produtos:

**Tabela 25: Produtos selecionados como Software de Infraestrutura**

<b>Produto/Projeto</b>	<b>Função</b>	<b>Descrição</b>
Linux Debian Server	Sistema Operativo	Sistema base no qual todos os produtos de infraestrutura ou de negócio irão ser executados.
KVM/QEMU	Virtualização	Criação e execução de máquinas virtuais num SO Linux
Docker Container	Contentor de Programas	Virtualização “leve” de programas em contentores
Kubernetes	Orquestração de Contentores	Criação, execução e gestão de contentores em ambiente distribuído.
OpenLDAP	Diretório de Utilizadores	Registo de utilizadores e suas passwords e grupos centralmente para integração nos restantes produtos
Postgresql	Sistema de gestão de bases de dados	Base de Dados relacional que suporta a persistência dos dados do Alfresco.
MariaDB	Sistema de gestão de bases de dados	Base de Dados relacional que suporta a persistência dos dados do AtoM.
ActiveMQ	Message Oriented Middleware	Servidor de Filas de Mensagens para comunicação entre as várias componentes
Spring Boot	Micro-serviços	Framework de criação rápida de micro-serviços
Elastic Search	Sistema de Indexação	Indexação e Pesquisa de dados indexados nas várias componentes
Nginx	Reverse Proxy	Disponibilização do acesso a todas as componentes a partir de um servidor de http central, mantendo-se o nome/ip do host e port

## 4.5.2 Gestão de Documentos de Arquivo

Tal como atrás descrito foram escolhidos um ou mais Pacotes de Software para implementar uma funcionalidade concreta do Sistema. Nos pontos que se seguem, descrevem-se os pacotes escolhidos para a implementação das Componentes Específicas da Gestão de Documentos de Arquivo.

### 4.5.2.1 Alfresco: Gestão Documental

O armazenamento e manipulação dos conteúdos (objetos digitais) fica a cargo de uma peça a que chamamos repositório.

O repositório de conteúdos é responsável por:

- Armazenar os objetos digitais nos dispositivos físicos

- Gerir a descrição dos objetos digitais com meta-dados
- Proteger os objetos digitais com permissões adequadas
- Automatizar com regras as ações sobre os objetos digitais
- Disponibilizar serviços para que as aplicações possam interagir com os objetos digitais e seus meta-dados, nomeadamente: criação, acesso, alteração, remoção, pesquisas.
- Indexar os objetos digitais para permitir as pesquisas.

A peça escolhida para armazenar os conteúdos é o Alfresco, nomeadamente a componente chamada de Content Services.

O Alfresco Content Services interage com o repositório de autenticação (implementado no protótipo utilizando o OpenLDAP), por forma a garantir que pesquisas, acessos e (ou) alterações são efetuadas de acordo com as permissões que são afetadas aos documentos.

Disponibiliza um conjunto de interfaces que permitem às outras peças do sistema ou aos utilizadores interagirem com a informação nele armazenado, nomeadamente:

**Tabela 26: Interfaces Disponibilizados pelo Alfresco**

Interface	Função
Alfresco Share	Interface Web que permite aos utilizadores interagir com o sistema, nomeadamente inserir documentos, atualizar meta-dados
CMIS	Protocolo Standard para interação com repositórios documentais («OASIS Content Management Interoperability Services (CMIS)», 2019)
REST	Serviços Web (sobre HTTP) usando o estilo arquitetural nomeado REST (Fielding & Taylor, 2000)
ADF	Componentes em Angular que implementam as principais interações dos utilizadores com um repositório e que poderão ser usadas na construção de aplicações de negócio que usam o Alfresco como repositório documental.

Foi instalado o módulo de Records Management, que é fundamental para que o Alfresco, para além de sistema de ECM possa implementar as funcionalidades adicionais fundamentais para um sistema de Gestão Documental.

#### 4.5.2.2 AtoM: Descrição Arquivística

O AtoM foi desenvolvido pela empresa Artefactual, com sede perto de Vancouver no Canadá em estreita colaboração com o ICA e com a Unesco. A sigla AtoM significa “Access to Memory”. Utiliza tecnologias de suporte que nos dão boas garantias de continuidade, nomeadamente: SGBDR MariaDB, Indexador Elastic Search, armazenamento dos objetos

digitais em sistema de ficheiros, PHP e javascript como linguagens de desenvolvimento, *memcached* para otimizar acessos aos dados, *Framework* Symfony para acelerar o desenvolvimento.

Implementa as normas ISAD(G), ISAAR(CPF), ISDIAH e ISDF.

É disponibilizado com a licença GNU Affero Public Licence (A-GPL 3.0).

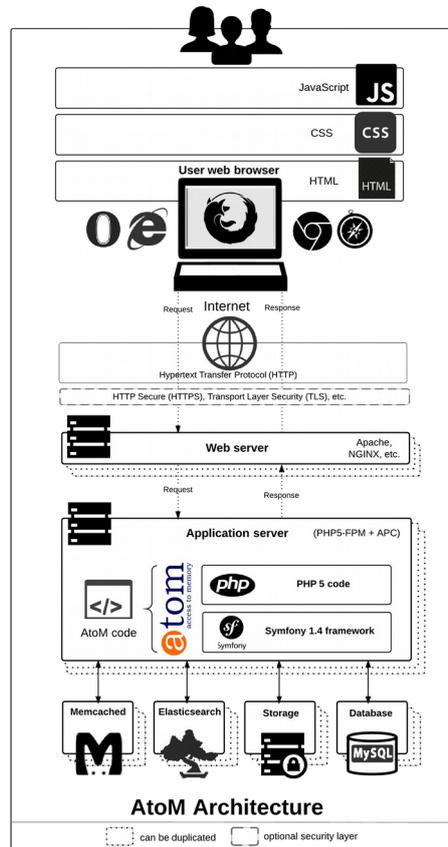


Figura 75: Arquitetura do AtoM

O interface com o utilizador funciona num browser. Esta solução facilmente consegue estar disponível de forma autónoma num contentor *docker*, com persistência da informação em volumes que nele são montados, garantindo assim uma grande independência deste sub-sistema face às versões do sistema operativo e das ferramentas existentes para outras componentes do sistema. Em termos de integração, a API REST que disponibiliza é infelizmente muito limitada, no entanto, as principais operações de importação e exportação de conteúdos podem ser efetuadas através de operações pela linha de comando, o que permite construir integrações.

### 4.5.2.3 Archivematica: Preservação Digital

O Archivematica foi implementado pela empresa canadiana Artefactual. É usado em instituições como o MoMA em New York, Harvard Business School ou a Unesco.

Implementa as principais normas usadas na Preservação Digital, como a ISO 14721 (Modelo OAIS), Premis, Mets e BagIT.

É disponibilizado com uma licença GNU Affero General Public Licence (A-GPL 3.0).

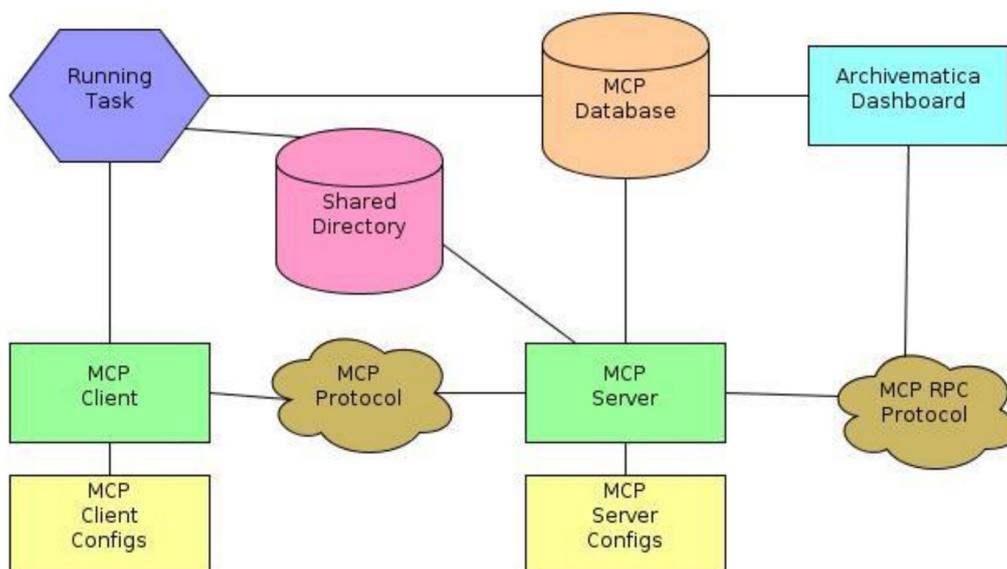


Figura 76: Arquitetura do Sistema Archivematica

No Archivematica, a Artefactual optou por implementar uma arquitetura baseada em micro-serviços executados por um sistema interno de workflow, tornando-o muito versátil.

A sub-componente MCP Server é responsável pela coordenação de todos os processos, configurações e execuções. Para comunicar com as outras componentes usa o MCP Protocol, armazenando a informação na MCP Database, monitorizando a Shared Directory e recebendo pedidos do MCP Cliente que é controlado por uma configuração própria. Disponibiliza Interface com o Utilizador que funciona em Browsers, inclusivamente Dashboards para auxiliar na gestão global e dos processos de preservação.

### 4.5.2.4 Esign-Cert: Assinatura Digital

O esign-cert é um módulo que pode ser implementado no Alfresco Share e que permite aos utilizadores assinarem digitalmente documentos existentes no repositório. Utiliza o projeto

@firma para interagir com os smart cards que contêm os certificados para assinatura digital qualificada como o Cartão do Cidadão de Portugal, disponibilizado pelo centro de Transferência de Tecnologia da Administração Pública Espanhola.

Ambos os projetos são disponibilizados com licença GNU General Public Licence 2.0.

#### 4.5.2.5 OpenOCR: Reconhecimento Ótico de Caracteres

O projeto OpenOCR, disponibilizado com licença GNU General Public Licence 2.0, utiliza um conjunto de tecnologias aplicando-as à implementação de um serviço escalável de OCR baseado em micro-serviços. Como motor de OCR usa o Tesseract. Para distribuição da carga usa o Kubernetes e uma Fila de Mensagens com um front-end único para as várias instâncias que podem crescer dinamicamente.

O Tesseract conta já com muitos anos de desenvolvimento. Foi inicialmente desenvolvido pela HP, e atualizado depois pela Google. Adicionalmente e porque se compreendeu ser complementar, selecionou-se o pacote open-ocr, que permite adicionar uma camada de serviços REST e capacidades de escalabilidade que permitem a criação de várias instâncias do Tesseract, cada uma funcionando sob a forma de um *container docker* usando ainda um *container* adicional que usa o RabbitMQ para sincronização e distribuição da carga entre instâncias.

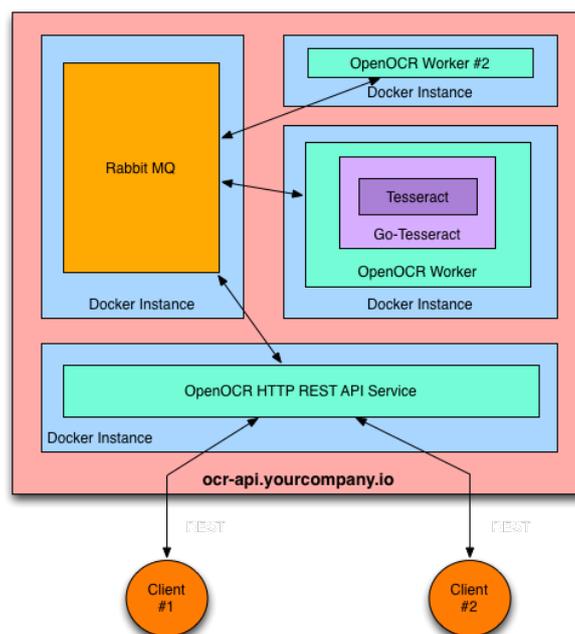


Figura 77: Arquitetura do OpenOCR

O open-ocr usa o Tesseract como “motor” de OCR a funcionar dentro de uma instância de um contentor docker. A esta componente chama-se OpenOCR worker. Para conseguir responder a necessidades de maiores cargas, podem ser criadas várias instâncias de *workers*. Estas são orquestradas pelo Kubernetes.

A disponibilização dos serviços é efetuada por um contentor de Front-End que comunica com o RabbitMQ para que este execute a extração no worker que estiver disponível.

#### **4.5.2.6 OpenCV: Detecção de Layouts**

O OpenCV é uma biblioteca de Visão Computacional originalmente desenvolvida pela Intel a partir do ano 1999. É disponibilizado com licença BSD e funciona em diversas plataformas, nomeadamente o Linux.

Implementa funcionalidades que permitem comparar duas imagens devolvendo um grau de semelhança entre ambas. No âmbito desta dissertação foram efetuados testes que mostraram que partindo de um documento base (como por exemplo uma fatura de gasolina), comparando documentos semelhantes, se consegue com um grande grau de certeza, compreender se os documentos comparados são do mesmo tipo do documento base.

O projeto “OpenCV python web services” disponibiliza a maior parte das funcionalidades do OpenCV sob a forma de uma API REST, devidamente encapsuladas num contentor.

Desta forma, com o DockerFile disponibilizado pelo projeto, é possível a criação de uma imagem que, ao ser executada, disponibiliza os serviços REST do OpenCV.

#### **4.5.2.7 Elasticsearch: Indexação e Pesquisa**

O Elasticsearch é um servidor de pesquisas baseado no Apache Lucene. É disponibilizado com Licença Apache 2.0.

A sua implementação permite a escalabilidade horizontal do problema de indexação e pesquisa.

A ingestão da informação que deve ser indexada pelo Elasticsearch é normalmente efetuada pelo produto logstash, no entanto, no caso concreto do Sistema de Gestão de Documentos de Arquivo investigado, é complementado pelo produto “OAI Harvester for Elasticsearch”. Este, substitui o logstash no caso dos servidores de OAI (OAI Data Providers) implementados pelo AtoM e pelo módulo de OAI:PMH para o Alfresco.

Desta forma, torna-se possível disponibilizar serviços de pesquisa aos utilizadores, remetendo os resultados para o sistema onde estes existem.

### **4.5.3 Componentes de Infraestrutura**

Tal como atrás descrito foram escolhidos um ou mais Pacotes de Software para implementar uma funcionalidade concreta do Sistema. Nos pontos que se seguem, descrevem-se os pacotes escolhidos para a implementação das Componentes de Infraestrutura com base nas quais as componentes relacionadas com a Gestão de Documentos de Arquivo forma implementadas.

#### **4.5.3.1 Linux: Sistema Operativo**

O Sistema Operativo Debian é disponibilizado sob licença Debian Free Software Guidelines DSFG. Duas das principais características do Debian são a estabilidade e a disponibilização apenas de pacotes de Software Livre.

A sua utilização no Sistema de Gestão de Documentos de Arquivo é bastante extensa já que foi usado como Sistema Operativo Hospedeiro nos principais servidores, como Sistema Operativo convidado e como base dos Contentores.

#### **4.5.3.2 KVM & Docker: Virtualização**

O KVM é disponibilizado com Licença GNU General Public Licence e o Docker com Licença Apache 2.0.

KVM significa Kernel-based Virtual Machine. É um módulo de virtualização existente sob a forma de módulo instalado no Kernel do Linux. Utiliza o emulador de CPU QEMU para executar as Máquinas Virtuais. Está disponível na distribuição standard do Debian, o que facilita a sua instalação e configuração no âmbito do Sistema de Gestão de Documentos de Arquivo.

É usado como sistema para virtualizar componentes cuja distribuição necessita de uma máquina para ser executado (que não estão preparados para funcionar num contentor).

O Docker começou por ser uma camada de gestão que usava o sistema de contentores em Linux, LXC. Ao implementar o seu próprio sistema de contentores com a linguagem Go, tornou-se autónomo e pode também estender a sua utilização aos sistemas operativos Windows e macOS.

As componentes do Sistema de Gestão de Documentos de Arquivo, sempre que possível, são implementadas sob a forma de contentores. Parte deles até já estão disponíveis no Docker HUB para download e execução.

### **4.5.3.3 Kubernetes: Orquestração de Contentores**

O Kubernetes automatiza as tarefas de criação, destruição e sincronização de contentores, permitindo desta forma que estes sejam criados quando a carga do sistema a isso obriga, e removidos quando não foram necessários. Desta forma, cada componente do sistema pode escalar ou ser reduzida horizontalmente com rapidez e facilidade.

### **4.5.3.4 OpenLDAP: Autenticação e Diretório de Utilizadores**

A autenticação nos principais sistemas é efetuada através do protocolo LDAP sobre o servidor OpenLDAP onde são registados os utilizadores e respetivas passwords encriptadas, contendo também um interface para administração dos utilizadores e do repositório, nomeadamente: criação, alteração, remoção de utilizadores, unidades orgânicas e outras configurações.

No caso da existência de uma camada aplicacional desenvolvida para o sistema global, o OpenDJ disponibiliza uma API REST que permite efetuar as operações sobre os utilizadores. Também disponibiliza um conjunto de comandos, permitindo assim a criação de scripts que efetuem manuseamentos que se julguem necessários.

Na arquitetura de referência, pode ser usado como uma componente instalada num dos nós da rede, ou ser usada sob a forma de um *container docker*, facilitando a administração de compatibilidade de versões.

### **4.5.3.5 PostgreSQL & MariaDB: Sistema de Gestão de Bases de Dados Relacionais**

O Alfresco utiliza um SGBDR PostgreSQL para persistir a sua informação. O AtoM e o Archivemática suportam-se no MariaDB.

O PostgreSQL é disponibilizado sob licença BSD. Trata-se de um SGDB lançado em 1997 como evolução do Ingres.

O MariaDB é um fork do MySQL efetuado pelo seu principal programador após a compra da Sun Microsystems pela Oracle. É disponibilizado com um licenciamento GNU General Public Licence.

As bases de dados usadas funcionam em contentores oficiais disponíveis no docker hub. Os dados são montados nos containers sob a forma de volumes.

#### **4.5.3.6 ActiveMQ: Filas de Mensagens Assíncronas**

Para integração, nomeadamente: para garantir que dados atualizados num sistema são replicados ou transferidos, consoante o padrão de integração que em cada caso faz sentido, define-se a utilização de um mecanismo standard de Filas de Mensagens, neste caso implementado pelo software ActiveMQ.

A integração com o ActiveMQ é efetuada através de Produtores e Consumidores que respetivamente Publicam ou Leem, a informação.

O ActiveMQ é disponibilizado com a Licença Apache Licence 2.0.

Foi implementado com um contentor de ActiveMQ ao qual os principais micro-serviços se irão ligar.

#### **4.5.3.7 Spring Boot &: Micro-serviços**

Para efeitos da evolução do sistema, foram criados vários exemplos de micro-serviços em Java com a Framework Spring Boot.

O Spring Boot é parte integrante da Framework Spring, e por isso disponível com uma Licença Apache 2.0.

Os exemplos implementados consistem em: consumidores de mensagens no ActiveMQ através de JMS; API(s) REST de fachada (que executa outro serviço REST, efetua o envio de uma mensagem ou grava informação numa base de dados).

### **4.6 Instalação e Configuração do Sistema**

O sistema foi montado sob a forma de uma máquina virtual com o sistema operativo Linux, distribuição Ubuntu, executada pelo Software de Virtualização KVM. Na máquina

virtual, para suportar a utilização de contentores especializados foi instalado o software Docker.

Em cada contentor foi instalado um pacote de software específico, garantindo assim um grande grau de independência entre cada um.

A utilização de contentores pré-existentes tornou o processo de configuração e instalação bem mais simples do que a primeira abordagem que consistia em instalar cada componente na mesma máquina. Dado os contentores se comportarem como se de uma máquina se tratasse, o nível de isolamento resolveu problemas como a utilização de bibliotecas com versões incompatíveis e(ou) utilização dos mesmos ports TCP-IP.

A tabela que se segue discrimina os conjuntos de contentores instalados, tanto para servir de componentes de infraestrutura como específicos da Gestão de Documentos de Arquivo. De notar que alguns projetos, como no caso do Archivematica, disponibiliza scripts de Docker Compose que permitem a instalação de um conjunto de contentores coerentes que assim executam a funcionalidade.

**Tabela 27: Módulos instalados em Contentores**

Módulo	Descrição
Alfresco	Alfresco Oficial Docker Compose
AtoM	AtoM Oficial Docker Compose
Archivematica	Archivematica Oficial Docker Compose
OpenOCR	OpenOCR Oficial Docker Compose
OpenCV	Dockerfile criado com o OpenCV e o OpenCVPythonWebServices
ElasticSearch	Elasticsearch Oficial Docker container
ElasticOAI	Dockerfile criado com o OAI:PMH harvester
OpenLDAP	OpenFrontier OpenLDAP Container
Postgresql	Postgresql Oficial Container
Maria DB	MariaDB Oficial Container
ActiveMQ	Dockerfile created container

As restantes componentes foram instaladas e configuradas de acordo com o respetivo contexto tal como se especifica na tabela que se segue:

**Tabela 28: Módulos não instalados em Contentores**

Módulo	Descrição da instalação
Debian	Sistema Operativo Instalado a partir da imagem ISO disponível no site da Debian e no caso dos contentores a

Módulo	Descrição da instalação
	partir da imagem oficial disponível no GitHub
KVM	Virtualizador instalado a partir do repositório Debian com “apt install kvm”
Docker	Sistema de contentores instalado a partir do repositório Debian com “apt install docker”
Kubernetes	Orquestração de contentores instalada com “apt install kubectl”
Esign-cert	Instalado sobre o(s) container(s) do Alfresco

# 5 Conclusões e trabalhos futuros

## 5.1 Principais Conclusões

A Gestão de Documentos de Arquivo é uma área extremamente relevante nas organizações na medida em que são responsáveis pelo arquivo, descrição, preservação e acesso produtivo à massa documental. Esta importância é tal, que se prevê atingir um volume de negócios da ordem dos 67 biliões de dólares em 2022.

Historicamente, a Gestão de Documentos de Arquivo derivou em três grandes áreas: Gestão Documental ou de Registos, Descrição Arquivística e Preservação Digital para as quais foram definidas normas, standards e conjuntos de boas práticas, sendo estas acompanhadas pela criação de Sistemas de Informação correspondentes.

A investigação efetuada, cuja questão de investigação consistiu em: *Poder-se-ia desenvolver um sistema que suporte o ciclo de vida dos Documentos de Arquivo baseado em tecnologias licenciadas como Software Livre ou Open Source ?*; permitiu identificar não só o contexto histórico, os standards e os tipos de sistemas criados, mas também compreender o estado da arte no que se refere à seleção de pacotes de Software Livre ou de Código Aberto e adicionalmente como usar uma Linguagem de Descrição de Arquiteturas para a definição de um Sistema Conceptual derivando num Sistema Concreto.

A abordagem metodológica utilizada, que permitiu identificar o estado da arte da Gestão de Documentos de Arquivo e a concretização de um artefacto, passando por uma especificação de conceito, permitiu concluir que:

É possível implementar um Sistema para Gestão de Documentos de Arquivo, tal como este é definido por organizações em normas e standards, com Software cuja disponibilização é efetuada com Licença de Software Livre ou de Código Aberto.

De facto, o artefacto criado, com base na proposta de solução final demonstra que é viável a construção de um sistema completo para Gestão de Documentos de Arquivo. Tal sistema será no entanto construído com recurso a vários sistemas que são coordenados entre si através da utilização de uma arquitetura que se poderá considerar de micro-serviços.

## 5.2 Trabalho Futuro

Ao longo do trabalho desenvolvido para esta dissertação, foram surgindo vários aspetos que podem ser estudados e analisados no futuro.

As questões a seguir apresentadas, que foram surgindo durante a investigação e implementação, são diretivas que podem estudadas expandindo o trabalho apresentado:

- Apesar de se ter estudado protocolos (por exemplo o OAI-PMH) que permitem efetuar pesquisas em vários sistemas, concluiu-se que a área da indexação e pesquisa de conteúdos, federada ou não, é por si uma área suficientemente abrangente e importante para ser estudada de forma autónoma.
- Embora para a Gestão de Documentos de Arquivo seja fundamental manter a informação de contexto e esta estar intimamente ligada aos processos que a cada registo deu origem, a gestão de *workflows* é uma área transversal nas tecnologias de informação. É importante em qualquer tipo de sistema informático (desde os ERP(s) aos CRM(s), passando pela Gestão de Documentos), mas esta aplicação abrangente versus a sua ligação aos documentos deve ser estudada em trabalhos futuros.
- A Classificação Automática de Documentos para além da utilização das técnicas de Visão Computacional podem ser melhoradas através da utilização de Machine Learning e Inteligência Artificial
- O Regulamento Geral de Proteção de Dados tem grandes implicações sobre os conteúdos armazenados sob a forma de registos ou documentos, sendo importante a investigação do seu impacto.
- O Controlo de Configuração com configuração automática de sistemas complexos que cumprem os normativos, usando ferramentas atualmente existentes para aprovisionamento e configuração automática, como por exemplo o Ansible e(ou) o Puppet é um aspeto importante de ser estudado para que estes sistemas possam ser criados de forma mais aproximada das necessidades das organizações concretas.
- A Possibilidade de criar e visualizar anotações públicas, privadas ou partilhadas, a apor em documentos, sem, no entanto efetuar quaisquer modificações nas partes do documento que tenham valor probatório são acréscimos que os utilizadores valorizam na medida em que permitem aumentar o valor do trabalho efetuado sobre os documentos.

- A possibilidade do registo da inserção e acesso aos documentos e aos registos, nomeadamente o acesso e indexação dos logs serem armazenados e tratados em Sistemas do tipo SIEM – Security Information and Event Management é uma área importante de estudo adicional.
- Poderia ser efetuada uma comparação de formatos de ficheiros de importação / exportação e dicionários de meta-dados mais extensiva e completa.

## 6 Referências Bibliográficas

- 2016 *Future of Open Source Survey Results*. (2016, Abril). Software. Obtido de <https://pt.slideshare.net/blackducksoftware/2016-future-of-open-source-survey-results>
- Abate, P., Di Cosmo, R., Treinen, R., & Zacchiroli, S. (2012). Dependency solving: A separate concern in component evolution management. *Journal of Systems and Software*, 85(10), 2228–2240. <https://doi.org/10.1016/j.jss.2012.02.018>
- ActiveMQ. (2019). Obtido 19 de Maio de 2019, de <https://activemq.apache.org/>
- Adewumi, A., Misra, S., & Omoregbe, N. (2013). A Review of Models for Evaluating Quality in Open Source Software. *IERI Procedia*, 4, 88–92. <https://doi.org/10.1016/j.ieri.2013.11.014>
- Aganette, E., Alvarenga, L., & Souza, R. R. (2010). Elementos constitutivos do conceito de Taxonomia. *Informação & Sociedade: Estudos*, 20(3). Obtido de <http://www.periodicos.ufpb.br/index.php/ies/article/view/3994>
- Agerwala, T., & Chatterjee, S. (2005). Computer architecture: Challenges and opportunities for the next decade. *IEEE Micro*, 25(3), 58–69. <https://doi.org/10.1109/MM.2005.45>
- Alberts, I., Schellinck, J., Eby, C., & Marleau, Y. (2010). Bridging Functions and Processes for Records Management / Combler le fossé entre les fonctions et les processus dans la gestion documentaire. *Canadian Journal of Information and Library Science*, 34(4), 365–390. <https://doi.org/10.1353/ils.2010.0010>
- Alföldi, I., & Réthy, I. (2017, Julho 2). Recommended Practices and Final Public Report on Pilots - E-Ark Project. Obtido 10 de Abril de 2018, de <http://www.eark-project.com/resources/project-deliverables/97-d25-1>
- Alfresco Community. (2018, Julho 23). Obtido 23 de Julho de 2018, de <https://community.alfresco.com/>

- Andrew Laurent. (2004). *Understanding Open Source and Free Software Licensing* [Book].  
Obtido 21 de Abril de 2019, de <https://www.oreilly.com/library/view/understanding-open-source/0596005814/>
- António, R. (2008). *O Sistema de gestão documental: Oportunidade do software livre nos Municípios Portugueses* (PhD Thesis).
- António, R., & Silva, A. (2012). *A gestão documental na perspectiva do MoReq2010* (1. ed).  
Lisboa: Edição de Autor.
- Apache Kafka. (2019). Obtido 19 de Maio de 2019, de Apache Kafka website:  
<https://kafka.apache.org/>
- Apache Solr. (2019). Obtido 19 de Maio de 2019, de <https://lucene.apache.org/solr/>
- ApacheDS — Apache Directory. (2019). Obtido 19 de Maio de 2019, de  
<https://directory.apache.org/apacheds/>
- Archivematica. (2019). Obtido 27 de Julho de 2018, de  
<https://wiki.archivematica.org/Overview>
- Arellano, M. Á. M. (2004). Preservação de documentos digitais. *Ciência da Informação*, 33(2).
- Atherton, J. (1985). From Life Cycle to Continuum: Some Thoughts on the Records Management–Archives Relationship. *Archivaria*, 21(0), 43–42.
- AtoM: Open Source Archival Description Software. (2019). Obtido 19 de Maio de 2019, de  
<https://www.accesstomemory.org/en/>
- BAD. (2019). Obtido 23 de Julho de 2018, de <http://www.apbad.pt/>
- Barbedo, F., & Corujo, L. (2012). MIP: Metainformação Para Interoperabilidade. *Lisboa: DGLAB. Acedido a*, 30.
- Barbedo, F., Corujo, L., Castro, R., Faria, L., Ramalho, J. C., & Ferreira, M. (2007). RODA: Repositório de Objectos Digitais Autênticos. *Actas do Congresso Nacional de*

- Bibliotecários, Arquivistas e Documentalistas*, 0(9). Obtido de <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/535>
- Bartlett, N. (1992). Respect des Fonds: *Primary Sources & Original Works*, 1(1–2), 107–115. [https://doi.org/10.1300/J269V01N01\\_07](https://doi.org/10.1300/J269V01N01_07)
- .be eID Sign (RedTree). (2019). Obtido 19 de Maio de 2019, de <https://community.alfresco.com/docs/DOC-7735-be-eid-sign-redtree>
- Björnander, S. (2019). *Architecture Description Languages*.
- Bluhm, M., Getting, B., Hayft, M., & Walz, S. (2011). *United States Patent N. US7941431B2*. Obtido de <https://patents.google.com/patent/US7941431B2/en>
- Blumberg, R., & Atre, S. (2003). The problem with unstructured data. *Dm Review*, 13(42–49), 62.
- BPMN Specification. (2011). Obtido 18 de Junho de 2019, de <http://www.bpmn.org/>
- Braconnier, S. (2019). *JODConverter* [Java]. Obtido de <https://github.com/sbraconnier/jodconverter> (Original work published 2016)
- Bredenberg, K., Skog, B., Nielsen, A. B., Johansen, K. H. E., Johansen, K. H. E., Wilson, A., & Luis Faria. (2016, Dezembro 16). E-ARK Common Specification. Obtido 10 de Abril de 2018, de <http://www.eark-project.com/resources/specificationdocs/67-e-ark-draft-common-specification-ver-017>
- Bruseker, G., Carboni, N., & Guillem, A. (2017). Cultural Heritage Data Management: The Role of Formal Ontology and CIDOC CRM. Em M. L. Vincent, V. M. López-Menchero Bendicho, M. Ioannides, & T. E. Levy (Eds.), *Heritage and Archaeology in the Digital Age: Acquisition, Curation, and Dissemination of Spatial Cultural Heritage Data* (pp. 93–131). [https://doi.org/10.1007/978-3-319-65370-9\\_6](https://doi.org/10.1007/978-3-319-65370-9_6)
- Caplan, P. (2009). *Understanding premis*. Library of Congress Washington DC, USA.
- CCSDS. (2012). *Reference Model for an Open Archival Information System (OAIS)*.
- CCSDS. (2019). Obtido 13 de Março de 2019, de <https://public.ccsds.org>

- Chen, P. P.-S. (1976). The Entity-relationship Model—Toward a Unified View of Data. *ACM Trans. Database Syst.*, 1(1), 9–36. <https://doi.org/10.1145/320434.320440>
- CIDOC CRM. (2019). Obtido 19 de Junho de 2019, de <http://www.cidoc-crm.org/>
- Collective Access. (2019). Obtido 19 de Maio de 2019, de <https://www.collectiveaccess.org/>
- Comitê de Normas de Descrição, ICA. (2004). ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition | International Council on Archives. Obtido 11 de Abril de 2018, de <https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd>
- Congress, L. of, & Committee, P. E. (2015). PREMIS Data Dictionary for Preservation Metadata, Version 3.0 (Library of Congress) [Webpage]. Obtido 27 de Março de 2018, de <http://www.loc.gov/standards/premis/v3/index.html>
- Cook, M. (2011). Desenvolvimento na descrição arquivística: Algumas sugestões para o futuro. *Acervo*, 20(1/2 janDez), 125–132.
- Crowston, K., Wei, K., Howison, J., & Wiggins, A. (2008). Free/Libre Open-source Software Development: What We Know and What We Do Not Know. *ACM Comput. Surv.*, 44(2), 7:1–7:35. <https://doi.org/10.1145/2089125.2089127>
- Cunningham, A. (2011). O Poder da Proveniência na Descrição Arquivística: Uma perspectiva sobre o desenvolvimento da segunda edição da ISAAR(CPF). *Acervo*, 20(1/2 janDez), 77–92.
- Cusumano, M. A. (2008). The changing software business: Moving from products to services. *Computer*, 41(1), 20–27.
- da Rosa Righi, R., Rodrigues, V. F., Rostirolla, G., André da Costa, C., Roloff, E., & Navaux, P. O. A. (2018). A lightweight plug-and-play elasticity service for self-organizing resource provisioning on parallel applications. *Future Generation Computer Systems*, 78, 176–190. <https://doi.org/10.1016/j.future.2017.02.023>

Dappert, A., & Enders, M. (2008). Using METS, PREMIS and MODS for Archiving eJournals. *D-Lib Magazine*, 14(9/10). <https://doi.org/10.1045/september2008-dappert>

Data Publisher for Earth & Environmental Science. (2019). Obtido 18 de Junho de 2019, de <https://pangaea.de/>

Davenport, T. H. (2005). The coming commoditization of processes. *Harvard business review*, 83(6), 100–108.

DCMI. (2019). Obtido 19 de Junho de 2019, de <http://dublincore.org/>

de Lima Baldam, R. (2002). *GED-Gerenciamento eletrônico de documentos*. Érica.

Debian -- The Universal Operating System. (2019). Obtido 19 de Maio de 2019, de <https://www.debian.org/>

DeNardis, L. (2010). E-Governance Policies for Interoperability and Open Standards. *Policy & Internet*, 2(3), 129–164. <https://doi.org/10.2202/1944-2866.1060>

Deprez, J.-C., & Alexandre, S. (2008). Comparing Assessment Methodologies for Free/Open Source Software: OpenBRR and QSOS. *Product-Focused Software Process Improvement*, 189–203. [https://doi.org/10.1007/978-3-540-69566-0\\_17](https://doi.org/10.1007/978-3-540-69566-0_17)

DGLAB, A. (2013, Junho 28). DGLAB. Obtido 13 de Março de 2019, de <http://dglab.gov.pt/>

DiBona, C., Stone, M., & Cooper, D. (2005). *Open sources 2.0: The continuing evolution*. O'Reilly Media, Inc.

Digital Library Federation. (2019). Obtido 19 de Junho de 2019, de DLF website: <https://www.diglib.org/>

Dikopoulou, A., & Mihiotis, A. (2012). The contribution of records management to good governance. *The TQM Journal*, 24(2), 123–141. <https://doi.org/10.1108/17542731211215071>

DLF. (2019). Obtido 13 de Março de 2019, de DLF website: <https://www.diglib.org/>

DLM Forum. (2010). *MoReq2010: Modular requirements for records systems, Volume 1. Core services & plug-in modules. Version 1.1.*

- DLM Forum. (2018). Obtido 27 de Julho de 2018, de <http://www.dlmforum.eu/>
- Docker Swarm. (2019, Maio 14). Obtido 19 de Maio de 2019, de Docker Documentation website: <https://docs.docker.com/engine/swarm/>
- DocuSign Connector for Alfresco. (2019). Obtido 19 de Maio de 2019, de <https://community.alfresco.com/docs/DOC-7763-docusign-connector-for-alfresco>
- DoD Standard 5015.2. (2007). Obtido 18 de Junho de 2019, de National Archives website: <https://www.archives.gov/records-mgmt/initiatives/dod-standard-5015-2.html>
- Dragoni, N., Lanese, I., Larsen, S. T., Mazzara, M., Mustafin, R., & Safina, L. (2018). Microservices: How To Make Your Application Scale. Em A. K. Petrenko & A. Voronkov (Eds.), *Perspectives of System Informatics* (pp. 95–104). Springer International Publishing.
- DSpace. (2019). Obtido 18 de Junho de 2019, de Duraspace.org website: <https://duraspace.org/dspace/>
- Duarte, F. S. G. (2017). *O acesso à informação no Departamento de Sócios do Sporting Clube de Portugal : 1906-1982*. Obtido de <http://repositorio.ul.pt/handle/10451/30410>
- Duchein, M. (1983). Theoretical Principles and Practical Problems of Respect des fonds in Archival Science. *Archivaria*, 16(0), 64–82.
- Duchein, M. (1992). The history of European archives and the development of the archival profession in Europe. *The American Archivist*, 55(1), 14–25.
- EAC-CPF. (2019). Obtido 19 de Junho de 2019, de <https://eac.staatsbibliothek-berlin.de/>
- EAD. (2019). Obtido 19 de Junho de 2019, de <https://www.loc.gov/ead/>
- E-Ark Project. (2014). Obtido 21 de Março de 2019, de <http://www.eark-project.com/>
- esign-cert - Alfresco client certificate signature (including Chrome)* [JavaScript]. (2019). Obtido de <https://github.com/keensoft/alfresco-esign-cert> (Original work published 2016)
- EXIF.org. (2019). Obtido 19 de Junho de 2019, de <https://www.exif.org/>

- Fedora. (2019). Obtido 19 de Maio de 2019, de <http://getfedora.org>
- Ferreira, M. (2006). *Introdução à preservação digital : Conceitos, estratégias e actuais consensos*. Obtido de <http://eprints.rclis.org/8524/>
- FFmpeg. (2019). Obtido 19 de Junho de 2019, de <https://ffmpeg.org/>
- Fielding, R. T., & Taylor, R. N. (2000). *Architectural styles and the design of network-based software architectures* (Vol. 7). University of California, Irvine Irvine, USA.
- Firebird: The true open source database for Windows, Linux, Mac OS X and more. (2019). Obtido 19 de Maio de 2019, de Home website: <https://firebirdsql.org/>
- Fitzgerald, B. (2006). The Transformation of Open Source Software. *MIS Quarterly*, 30(3), 587–598. <https://doi.org/10.2307/25148740>
- Flores, A. C. (2004). *A importância dos sistemas de arquivos na qualidade das informações empresarias*. Obtido de <http://repositorio.ufsm.br/handle/1/548>
- Flores, D., & Hedlund, D. C. (2014). Análise e aplicação do ICA-AtoM como ferramenta para descrição e acesso ao Patrimônio Documental e Histórico do município de Santa Maria – RS. *Informação & Informação*, 19(3), 86–106. <https://doi.org/10.5433/1981-8920.2014v19n3p86>
- Fowler. (2014). Microservices. Obtido 31 de Março de 2019, de [martinfowler.com](http://martinfowler.com) website: <https://martinfowler.com/articles/microservices.html>
- Fowler, M. (2002). *Patterns of Enterprise Application Architecture*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Francke, H. (2005). What's in a Name? Contextualizing the Document Concept. *Literary and Linguistic Computing*, 20(1), 61–69. <https://doi.org/10.1093/llc/fqh043>
- Frank Upward. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond Ö a personal reflection | Records Management Journal | Vol 10, No 3. Obtido 23 de Julho de 2018, de <https://www.emeraldinsight.com/doi/pdfplus/10.1108/EUM0000000007259>

- Free Software Foundation. (1989). GNU GENERAL PUBLIC LICENSE. Obtido 21 de Abril de 2019, de <https://www.gnu.org/licenses/gpl.html>
- Free Software Foundation. (2019). Obtido 18 de Junho de 2019, de <https://www.fsf.org/>
- Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future*, 2007(2012), 1–16.
- Garriga, M. (2018). Towards a Taxonomy of Microservices Architectures. Em A. Cerone & M. Roveri (Eds.), *Software Engineering and Formal Methods* (pp. 203–218). Springer International Publishing.
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, 30(4), 378–391.  
<https://doi.org/10.1177/0165551504045856>
- GitHub Octoverse 2018 | Highlights from the last twelve months. (2018). Obtido 27 de Julho de 2018, de <https://octoverse.github.com/>
- Global enterprise content management market size 2017-2022. (2018). Obtido 27 de Julho de 2018, de <https://www.statista.com/statistics/506914/enterprise-content-management-market-size/>
- gnu.org. (2019). Obtido 17 de Maio de 2019, de <https://www.gnu.org/>
- GOCR. (2019). Obtido 19 de Maio de 2019, de <http://jocr.sourceforge.net/>
- Goldman, E. (2008). Search Engine Bias and the Demise of Search Engine Utopianism. Em A. Spink & M. Zimmer (Eds.), *Web Search: Multidisciplinary Perspectives* (pp. 121–133). [https://doi.org/10.1007/978-3-540-75829-7\\_8](https://doi.org/10.1007/978-3-540-75829-7_8)
- Google. (2019). Obtido 18 de Junho de 2019, de <https://www.google.com/>
- Grimes, S. (2008). Unstructured data and the 80 percent rule. *Carabridge Bridgepoints*, 10.
- Healy, S. (2010). ISO 15489 records management: Its development and significance. *Records Management Journal*, 20(1), 96–103.

- Hevner. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2), 4.
- Hevner, March, S., Park, J., & Ram, S. (2008). Design Science in Information Systems Research. *Management Information Systems Quarterly*, 28(1). Obtido de <http://aisel.aisnet.org/misq/vol28/iss1/6>
- Hill, M. D. (1990). What is Scalability? *SIGARCH Comput. Archit. News*, 18(4), 18–21. <https://doi.org/10.1145/121973.121975>
- Hinkelmann, K., Gerber, A., Karagiannis, D., Thoenssen, B., van der Merwe, A., & Woitsch, R. (2016). A new paradigm for the continuous alignment of business and IT: Combining enterprise architecture modelling and enterprise ontology. *Computers in Industry*, 79, 77–86. <https://doi.org/10.1016/j.compind.2015.07.009>
- Hohpe, G., & Woolf, B. (2003). *Enterprise integration patterns*.
- Hohpe, G., & Woolf, B. (2004). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley Professional.
- Huang, T. (1996). *Computer Vision: Evolution and Promise*. Obtido de <http://cds.cern.ch/record/400313/files/p21.pdf>
- Hussain, S. (2013). *Investigating Architecture Description Languages (ADLs) A Systematic Literature Review*. Obtido de <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-104856>
- ICA. (2011). ISAD(G): General International Standard Archival Description - Second edition | International Council on Archives. Obtido 20 de Março de 2019, de <https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition>
- ICA. (2019). Obtido 13 de Março de 2019, de <https://www.ica.org/en>
- ICOM. (2019). Obtido 13 de Março de 2019, de <https://icom.museum/en/>
- ImageMagick. (2019). Obtido 19 de Junho de 2019, de ImageMagick website: <https://imagemagick.org/>

Inkscape. (2019). Obtido 19 de Junho de 2019, de <https://inkscape.org/pt/>

IPQ. (2018). Obtido 13 de Março de 2019, de <http://www.ipq.pt>

ISDIAH: International Standard for Describing Institutions with Archival Holdings |  
International Council on Archives. (2008, Março 11). Obtido 27 de Julho de 2018, de <https://www.ica.org/en/isdiah-international-standard-describing-institutions-archival-holdings>

ISO. (2016). ISO 15489-1:2016. Obtido 27 de Abril de 2018, de <https://www.iso.org/standard/62542.html>

ISO. (2018). Obtido 13 de Março de 2019, de ISO website: <http://www.iso.org>

ISO/IEC/IEEE 42010. (2011). Obtido 22 de Maio de 2018, de <https://www.iso.org/standard/50508.html>

Jarvinen, P. (2000). Research Questions Guiding Selection of an Appropriate Research Method. *ECIS 2000 Proceedings*. Obtido de <https://aisel.aisnet.org/ecis2000/26>

Joy, A. M. (2015). Performance comparison between Linux containers and virtual machines. *2015 International Conference on Advances in Computer Engineering and Applications*, 342–346. <https://doi.org/10.1109/ICACEA.2015.7164727>

Kaplan, S. (2002, Janeiro 15). Cool Tool: Comparing Content Management and Document Management. Obtido 3 de Abril de 2018, de CIO website: <https://www.cio.com/article/2441032/collaboration/cool-tool--comparing-content-management-and-document-management.html>

Katuu, S. (2012). Enterprise content management (ECM) implementation in South Africa. *Records Management Journal*, 22(1), 37–56. <https://doi.org/10.1108/09565691211222081>

Kubernetes : Production-Grade Container Orchestration. (2019). Obtido 19 de Maio de 2019, de <https://kubernetes.io/>

- Kulathuramaiyer, N., & Balke, W.-T. (2006). Restricting the View and Connecting the Dots-Dangers of a Web Search Engine Monopoly. *J. UCS*, 12(12), 1731–1740.
- KVM. (2019). Obtido 19 de Maio de 2019, de [https://www.linux-kvm.org/page/Main\\_Page](https://www.linux-kvm.org/page/Main_Page)
- Lage de Medeiros, N., & Gomes do Amaral, C. M. (2010). A Representação do ciclo vital dos documentos: Uma discussão sob a ótica da gestão de documentos. *Em Questão*, 16(2). Obtido de <http://www.redalyc.org/resumen.oa?id=465645963019>
- Lankhorst, M. M., Proper, H. A., & Jonkers, H. (2009). The Architecture of the ArchiMate Language. Em *Lecture Notes in Business Information Processing. Enterprise, Business-Process and Information Systems Modeling* (pp. 367–380). [https://doi.org/10.1007/978-3-642-01862-6\\_30](https://doi.org/10.1007/978-3-642-01862-6_30)
- Laumer, S., Beimborn, D., Maier, C., & Weinert, C. (2013). Enterprise Content Management. *Business & Information Systems Engineering*, 5(6), 449–452. <https://doi.org/10.1007/s12599-013-0291-3>
- Lavoie, B. F. (2008). The Open Archival Information System Reference Model: Introductory Guide. *Microform & Imaging Review*, 33(2), 68–81. <https://doi.org/10.1515/MFIR.2004.68>
- Lee, Slattery, O., Lu, R., Tang, X., & McCrary, V. (2002). The State of the Art and Practice in Digital Preservation. *Journal of Research of the National Institute of Standards and Technology*, 107(1), 93–106. <https://doi.org/10.6028/jres.107.010>
- Lee, T., & Iio, J. (2015). Document Management System Based on ISAD(G). *2015 18th International Conference on Network-Based Information Systems*, 685–689. <https://doi.org/10.1109/NBiS.2015.103>
- Leyden, T. (2019). *open-ocr - Run your own OCR-as-a-Service using Tesseract and Docker* [Go]. Obtido de <https://github.com/tleyden/open-ocr> (Original work published 2014)
- Library of Congress. (2015). Metadata Encoding and Transmission Standard (METS). Obtido 18 de Junho de 2019, de <http://www.loc.gov/standards/mets/>

- Library of Congress [Web page]. (2019). Obtido 13 de Março de 2019, de Library of Congress, Washington, D.C. 20540 USA website: <https://www.loc.gov/>
- Liu, C., Shie, M., Lee, Y., Lin, Y., & Lai, K. (2014). Vertical/Horizontal Resource Scaling Mechanism for Federated Clouds. *2014 International Conference on Information Science Applications (ICISA)*, 1–4. <https://doi.org/10.1109/ICISA.2014.6847479>
- LogicalDoc. (2019). Obtido 19 de Maio de 2019, de LogicalDOC website: <https://www.logicaldoc.com/pt/>
- Lopes, V. (2008). Preservação digital. *Portugal: Universidade do Minho, Guimarães*.
- Lourenço, Henriques, Penteado. (2013, Março 28). Macroestrutura funcional – MEF. Obtido 27 de Julho de 2018, de <http://arquivos.dglab.gov.pt/programas-e-projectos/modernizacao-administrativa/macroestrutura-funcional-mef/macroestrutura-funcional-mef/>
- Love, R. (2010). *Linux Kernel Development: Linux Kernel Development \_p3*. Pearson Education.
- Lusch, R. F., Vargo, S. L., & O'Brien, M. (2007). Competing through service: Insights from service-dominant logic. *Journal of Retailing*, 83(1), 5–18. <https://doi.org/10.1016/j.jretai.2006.10.002>
- Madalli, D. P., Barve, S., & Amin, S. (2012). Digital Preservation in Open-Source Digital Library Software. *The Journal of Academic Librarianship*, 38(3), 161–164. <https://doi.org/10.1016/j.acalib.2012.02.004>
- MariaDB.org. (2019). Obtido 19 de Maio de 2019, de MariaDB.org website: <https://mariadb.org/>
- Messaging that just works — RabbitMQ. (2019). Obtido 19 de Maio de 2019, de <https://www.rabbitmq.com/>
- Millar, L. A. (2017). *Archives: Principles and practices*. Facet Publishing.
- Minix3. (2019). Obtido 19 de Maio de 2019, de <https://www.minix3.org/>

MIT CSAIL. (2019). Obtido 18 de Junho de 2019, de <https://www.csail.mit.edu/>

Moreau. (2013). *Qualification And Selection of Open Source Software*. Obtido de <https://www.qsos.org/method>

Morgan, L., & Finnegan, P. (2014). Beyond free software: An exploration of the business value of strategic open source. *The Journal of Strategic Information Systems*, 23(3), 226–238. <https://doi.org/10.1016/j.jsis.2014.07.001>

Mori, S., Suen, C. Y., & Yamamoto, K. (1992). Historical review of OCR research and development. *Proceedings of the IEEE*, 80(7), 1029–1058. <https://doi.org/10.1109/5.156468>

MyBatis. (2019). Obtido 19 de Maio de 2019, de <https://blog.mybatis.org/>

Netflix Technology. (2013, Janeiro 28). Announcing Ribbon: Tying the Netflix Mid-Tier Services Together. Obtido 31 de Março de 2019, de Medium website: <https://medium.com/netflix-techblog/announcing-ribbon-tying-the-netflix-mid-tier-services-together-a89346910a62>

Norma NP 4438-1. (2005). Obtido 27 de Abril de 2018, de <http://www1.ipq.pt/PT/site/clientes/pages/Norma.aspx?docId=IPQDOC-185-149830>

Nuxeo. (2019). Obtido 19 de Maio de 2019, de Nuxeo website: <https://www.nuxeo.com/>

O Cartão de Cidadão. (2019). Obtido 18 de Junho de 2019, de <https://www.autenticacao.gov.pt/o-cartao-de-cidadao>

OASIS Content Management Interoperability Services (CMIS). (2019). Obtido 20 de Março de 2019, de [https://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=cmis](https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=cmis)

ODLIS. (2018, Julho 23). Obtido 23 de Julho de 2018, de [https://www.abc-clio.com/ODLIS/odlis\\_i.aspx](https://www.abc-clio.com/ODLIS/odlis_i.aspx)

Online Computer Library Center. (2019, Março 19). Obtido 20 de Março de 2019, de OCLC website: <https://www.oclc.org/en/home.html>

- Open Archives Initiative. (2019). Obtido 20 de Março de 2019, de <https://www.openarchives.org/>
- Open Journal Systems. (2019). Obtido 18 de Junho de 2019, de <https://pkp.sfu.ca/ojs/>
- Open Source ArchiMate Modelling. (2019). Obtido 18 de Maio de 2019, de <https://www.archimatetool.com/>
- Open Source Search & Analytics · Elasticsearch. (2019). Obtido 19 de Maio de 2019, de <https://www.elastic.co/>
- OpenCV. (2019). Obtido 19 de Maio de 2019, de <https://opencv.org/>
- OpenKM. (2019). OpenKM. Obtido 19 de Maio de 2019, de OpenKM website: <https://www.openkm.com/>
- OpenLDAP. (2019). Obtido 19 de Maio de 2019, de <http://www.openldap.org/>
- OpenOCR. (2019). Obtido 19 de Maio de 2019, de SourceForge website: <https://sourceforge.net/projects/openocr/>
- Oracle VM VirtualBox. (2019). Obtido 19 de Maio de 2019, de <https://www.virtualbox.org/>
- OSSPAL. (2019). Obtido 19 de Maio de 2019, de <http://osspal.org/>
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Penteado, P., Henriques, C., & Lourenço, A. (2012). O desafio da interoperabilidade na gestão dos arquivos da Administração: Propostas do órgão de coordenação nacional de arquivos. *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, 0(11). Obtido de <https://www.bad.pt/publicacoes/index.php/congressosbad/article/view/452>
- Pérotin, Y. (1966). Administration and the "three ages" of archives. *The American Archivist*, 29(3), 363–369.

- PostgreSQL: The world's most advanced open source database. (2019). Obtido 19 de Maio de 2019, de <https://www.postgresql.org/>
- Prante, J. (2019). *OAI client for Elasticsearch* [Java]. Obtido de <https://github.com/jprante/elasticsearch-oai> (Original work published 2012)
- Premis Working Group. (2016, Agosto 5). Obtido 20 de Março de 2019, de OCLC website: <https://www.oclc.org/research/activities/pmwg.html>
- QSOS. (2018). Obtido 18 de Maio de 2019, de <https://www.qsos.org/>
- Ramalho, J. C., Ferreira, M., Castro, R., Faria, L., Barbedo, F., & Corujo, L. (2007, Fevereiro 15). *XML e preservação digital*. Obtido de <http://repositorium.sdum.uminho.pt/handle/1822/6224>
- Raymond, E. S. (1998). The cathedral and the bazaar. *First Monday*, 3(2).  
<https://doi.org/10.5210/fm.v3i2.578>
- Ribeiro, C. J. S. (2018). Soluções em ECM (Enterprise Content Management) e softwares para repositórios: Convergência de requisitos nas atividades para gestão documental. *Informação & Tecnologia*, 3(2), 4–20.
- RLG. (2019). Obtido 20 de Março de 2019, de <http://www.rlg.org/>
- RODA - Repositório de Objetos Digitais Autênticos. (2019). Obtido 19 de Maio de 2019, de <https://roda.arquivos.pt/#welcome>
- Ross, S. (2012). Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries. *New Review of Information Networking*, 17(1), 43–68.  
<https://doi.org/10.1080/13614576.2012.679446>
- Rousseau, J.-Y., Couture, C., & Arès, F. (1998). *Os fundamentos da disciplina arquivística*. Publicações Dom Quixote.
- Rowley, J., & Hartley, R. (2017). *Organizing Knowledge : An Introduction to Managing Access to Information*. <https://doi.org/10.4324/9781315247519>

- Runa, L., Barbedo, F., & Almeida, M. J. (2018). Projeto ICON - Integração de CONteúdos: Para um novo modelo de descrição. *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, 0(13). Obtido de <https://bad.pt/publicacoes/index.php/congressosbad/article/view/1738>
- SAA. (2019). Obtido 13 de Março de 2019, de <https://www2.archivists.org/>
- Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 44(1.2), 206–226.
- Saraiva, N. de L., & Nogueira, R. de F. (2015). Software livre: Um diferencial competitivo para a Arquivística. *Informação Arquivística*, 3(2). <https://doi.org/10.18377/2316-7300/informacaoarquivistica.v3n2p%>
- Sayão, L. F., & Marcondes, C. H. (2008). O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. *Transinformação*, 20(2), 133–148.
- Schellenberg, T. R. (1956). *Modern archives. Principles and techniques*.
- Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science*, 2(1–2), 1–19. <https://doi.org/10.1007/BF02435628>
- Seymour, J. (2017). The Modern Records Management Program: An Overview of Electronic Records Management Standards. *Bulletin of the Association for Information Science and Technology*, 43(2), 35–39. <https://doi.org/10.1002/bul2.2017.1720430212>
- Shadija, D., Rezai, M., & Hill, R. (2017). Towards an understanding of microservices. *2017 23rd International Conference on Automation and Computing (ICAC)*, 1–6. <https://doi.org/10.23919/IConAC.2017.8082018>
- Shallcross, M. (2016). Appraising digital archives with Archivematica. *2016 IEEE International Conference on Big Data (Big Data)*, 3272–3276. <https://doi.org/10.1109/BigData.2016.7840985>
- Shokouhi, M., & Si, L. (2011). Federated Search. *Foundations and Trends® in Information Retrieval*, 5(1), 1–102. <https://doi.org/10.1561/15000000010>

- Silva, J. T. e. (2016). NORMAS ISO PARA GESTÃO DE DOCUMENTOS: Uma introdução. *Archeion Online*, 4(1), 04–21.
- Singh, B., Robin, Narendra, C., Singh, T., Geetam, & Shailendra, S. (2017). *Exploring Enterprise Service Bus in the Service-Oriented Architecture Paradigm*. IGI Global.
- Smith, H. A., & McKeen, J. D. (2003). Developments in practice VIII: Enterprise content management. *The Communications of the Association for Information Systems*, 11(1), 41.
- Söderberg, J. (2015). *Hacking Capitalism : The Free and Open Source Software Movement*.  
<https://doi.org/10.4324/9780203937853>
- Sousa, R. T. B. de. (2006). *Classificação de documentos arquivísticos : Trajetória de um conceito*. Obtido de <http://repositorio.unb.br/handle/10482/948>
- Spring Boot. (2019). Obtido 19 de Maio de 2019, de <https://spring.io/projects/spring-boot>
- Stallman, R. (1998). *The GNU project*.
- Szary, R. (2006). Encoded Archival Context (EAC) and Archival Description: Rationale and Background. *Journal of Archival Organization*, 3(2–3), 217–227.  
[https://doi.org/10.1300/J201v03n02\\_16](https://doi.org/10.1300/J201v03n02_16)
- Taylor, A. G., & Joudrey, D. N. (2017). *The Organization of Information, 4th Edition*. ABC-CLIO.
- tesseract-ocr. (2019). Obtido 19 de Maio de 2019, de GitHub website:  
<https://github.com/tesseract-ocr>
- The Open Group. (2017). ArchiMate® 3.0.1 Specification. Obtido 23 de Maio de 2018, de <http://pubs.opengroup.org/architecture/archimate3-doc/>
- Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. *The state of digital preservation: an international perspective*, 4–31.

- Thomaz, K. P., & Soares, A. J. (2009). A preservação digital e o modelo de referência Open Archival Information System (OAIS). *DataGramZero, Rio de Janeiro*, 5(1), A01–1001.
- Tom. (2019). *OCROpus - Python-based tools for document analysis and OCR*. [Jupyter Notebook]. Obtido de <https://github.com/tmbdev/ocropy> (Original work published 2014)
- Torres, S. (2018). GESTÃO DA INFORMAÇÃO NA CÂMARA MUNICIPAL DO PORTO. *Páginas a&b: arquivos e bibliotecas*, 0(0), 99–117.
- Tyrväinen, P., Päivärinta, T., Salminen, A., & Iivari, J. (2006). Characterizing the evolving research on enterprise content management. *European Journal of Information Systems*, 15(6), 627–634. <https://doi.org/10.1057/palgrave.ejis.3000648>
- Ubuntu - The leading operating system for PCs, IoT devices, servers and the cloud. (2019). Obtido 19 de Maio de 2019, de <https://www.ubuntu.com/>
- Umar, A., & Zordan, A. (2009). Reengineering for service oriented architectures: A strategic decision model for integration versus migration. *Journal of Systems and Software*, 82(3), 448–462. <https://doi.org/10.1016/j.jss.2008.07.047>
- Umm-e-Laila, Zahoor, A., Mehboob, K., & Natha, S. (2017). Comparison of open source maturity models. *Procedia Computer Science*, 111, 348–354. <https://doi.org/10.1016/j.procs.2017.06.033>
- Unified Modeling Language Specification. (2017). Obtido 18 de Junho de 2019, de <https://www.omg.org/spec/UML/About-UML/>
- Upward, F. (1996). Structuring the records continuum (Series of two parts) Part 1: Post custodial principles and properties. *Archives and Manuscripts*, 24(2), 268.
- Van de Sompel, H., Nelson, M. L., Lagoze, C., & Warner, S. (2004). Resource harvesting within the OAI-PMH framework. *D-lib magazine*, 10(12).

- Vieira, R., & Borbinha, J. (2011). MoReq2010 – Uma Apresentação. *Actas do Encontro Nacional de Arquivos Municipais*, 0(10). Obtido de <http://bad.pt/publicacoes/index.php/arquivosmunicipais/article/view/19>
- Walfield, N. H., & Brinkmann, M. (2007). A Critique of the GNU Hurd Multi-server Operating System. *SIGOPS Oper. Syst. Rev.*, 41(4), 30–39. <https://doi.org/10.1145/1278901.1278907>
- Warasart, M., & Kuacharoen, P. (2012). based document authentication using digital signature and QR code. *4th International Conference on Computer Engineering and Technology. International Proceedings of Computer Science and Information Technology*, 94–98.
- Williams, C. (2006). *Managing archives: Foundations, principles and practice*. Elsevier.
- Xena - Digital Preservation Software. (2019). Obtido 19 de Maio de 2019, de [https://sourceforge.net/p/xena/wiki/Main\\_Page/](https://sourceforge.net/p/xena/wiki/Main_Page/)
- Xiao, Z., Wijegunaratne, I., & Qiang, X. (2016). Reflections on SOA and Microservices. *2016 4th International Conference on Enterprise Systems (ES)*, 60–67. <https://doi.org/10.1109/ES.2016.14>

## 7 Apendix A – Requisitos Alfresco vs

### MoReq2010

A avaliação do Alfresco que se segue foi efetuada de acordo com os requisitos funcionais definidos pelo MoReq2010, com base nos quais foi realizada uma proposta de tradução na obra “A Gestão Documental na perspetiva do MoReq2010” (António, 2008).

A avaliação dos requisitos funcionais original foi ligeiramente modificada sendo esta efetuada de acordo com as seguintes possibilidades:

- Sim – O software implementa totalmente a funcionalidade.
- Não - O sistema não implementa de todo a funcionalidade.
- Parcial – A funcionalidade está implementada de forma parcial

#### 7.1.1.1 Serviços do Sistema

**Tabela 29: Avaliação de Requisitos de Serviços de Sistema em relação ao MoReq**

ID	Funcionalidade	Sim/Não/ Parcial	Observações
1	Serviços do Sistema		
1.1	Implementação do serviço de utilizadores e grupos; serviço de perfis; serviço de classificação; serviço de registo de documentos; serviço de meta-dados; serviço de seleção e eliminação; serviço de retenção; serviço de pesquisa e serviço de exportação. Cada serviço pode ser implementado individualmente ou alguns serviços podem ser agrupados	Sim	O Alfresco, com o módulo de Records Management devidamente instalado implementa todos os serviços descritos sendo alguns agrupados (ex: utilizadores, grupos e perfis estão implementados num serviço; classificação, eliminação, retenção está implementado nos serviços do módulo de Records Management)
1.2	Configuração do um conjunto de meta-dados para cada serviço: executar a identificação do sistema; executar a identificação do módulo; identificação da certificação do sistema e informação do fornecedor; identificação do idioma padrão, título, descrição, informações do proprietário; tipos de entidade; histórico de eventos e meta-dados contextuais	Parcial	Quase todas as entidades no Alfresco são consideradas nós, os quais podem ter associados conjuntos de meta-dados (aspetos ou tipos). Alguns serviços contêm meta-dados outros, por serem tão <i>core</i> da implementação não.
1.3	Permitir a um utilizador autorizado navegar através dos serviços, ou conjunto de serviços, e inspecionar os meta-dados.	Parcial	Nem todos os serviços podem ser “navegados”.
1.4	Permitir a um utilizador autorizado modificar os meta-dados para cada serviço, ou conjunto de serviços, incluindo: título; descrição;	Parcial	Um utilizador autorizado pode modificar os meta-dados nos serviços em que estes existirem.

	informações do proprietário e meta-dados contextuais.		
1.5	Permitir a um utilizador autorizado criar um relatório de conformidade com o MoReq2010, listando os serviços activos, ou conjuntos de serviços, e para cada um a lista com os seus meta-dados.	Parcial	A verificação dos serviços activos bem como os seus meta-dados não é efetuada num único local. Para algumas verificações tem de ser o administrador a efetuar tal verificação existindo complexidade técnica associada.
1.6	Assegurar que cada serviço, ou conjunto de serviços, tem um interface que implemente um dos módulos do MoReq2010®, 100. Interface Series.	Não	
1.7	Sempre que o sistema não complete uma função solicitada pelo próprio ou por um utilizador autorizado, o sistema deve registar pelo menos as informações de erro seguintes: data/hora da falha; identificador do sistema da função que foi tentada; identificador do sistema do utilizador autorizado que iniciou a função; identificador do sistema de qualquer entidade participante; informações de erro com a descrição da falha.	Parcial	O Alfresco regista as falhas em log(s) do sistema, mas nem sempre é possível compreender qual o utilizador que iniciou a função.
1.8	Em caso de erro de uma função solicitada por um utilizador, o sistema deve disponibilizar uma forma de recuperar informações de erro sobre a função que falhou, sem aceder ao diário externo.	Não	
1.9	Permitir a um utilizador autorizado ver os tipos de entidades associados a cada documento e inspecionar os seus meta-dados	Sim	No Alfresco as associações entre nós são possíveis ser visualizadas. Sempre de acordo com as respetivas permissões.
1.10	Cada entidade deve ter os meta-dados seguintes: identificador do sistema, título; descrição; definições de elementos de meta-dados do sistema para esse tipo de entidade; definições de funções para esse tipo de entidade; histórico de eventos; lista de permissões de acesso.	Sim	As principais entidades podem conter os meta-dados, contêm histórico quando configurado e a elas podem ser associadas listas de permissões de acessos.
1.11	Para cada tipo de entidade permitir a um utilizador autorizado navegar na definição de funções de entidades associadas a esse tipo de entidade e inspecionar os seus meta-dados.	Não	Roles e permissões não são nodes e por isso apesar de serem entidades e extensíveis (permitindo adicionar novos) não é possível nelas navegar nem associar meta-dados.
1.12	Para cada definição de função deve permitir a um utilizador autorizado especificar se um evento deve ou não ser criado pelo sistema, quando a função é executada.	Não	Pela mesma razão do ponto anterior
1.13	Deve ser criado um evento sempre que uma função for executada.	Não	Não está previsto no Alfresco
1.14	Sempre que uma função for realizada por uma entidade do sistema, sujeita à opção de criação de evento, o sistema deve criar automaticamente o novo evento que descreve a função e incluí-lo no histórico de eventos das entidades participantes.	Não	Pelas mesmas razões dos anteriores

1.15	Para cada evento associado a uma função que está a ser realizada, o sistema deve incluir meta-dados: identificador do sistema; registo de criação; registo do evento; identificador da função do evento; identificador do utilizador, comentário do evento, entre outros.	Não	
1.16	Sempre que os meta-dados de uma entidade são modificados como resultado da execução de uma função e é criado um evento relacionado com a mesma, o sistema deve incluir uma entrada de evento relativo à alteração dos meta-dados, com os valores seguintes: identificador da definição do elemento de meta-dados; valor anterior e novo valor.	Sim	O Alfresco, com o módulo de RM e com o audit ligado esta informação é armazenada
1.17	Sempre que o sistema realize uma função solicitada pelo utilizador, que altere os meta-dados, o sistema deve permitir ao utilizador a introdução de um comentário sobre a razão da execução da função.	Não	Apesar de se poder adicionar comentários às versões, alterações aos meta-dados não são considerados novas versões
1.18	Permitir a um utilizador autorizado navegar na história dos eventos de uma entidade, segundo uma sequência lógica e/ou cronológica, e inspecionar os seus meta-dados.	Parcial	A informação de audit está disponível para administradores. A informação de histórico dos registos (do módulo RM) pode ser navegada.
1.19	Permitir a um utilizador autorizado especificar se um evento é criado ou não pelo sistema quando a função for realizada e se deve ser mantido caso a entidade a que pertence seja eliminada.	Parcial	No Alfresco eventos/regras podem ser associados a todas as entidades consideradas nodes.
1.20	Permitir a um utilizador autorizado excluir um acontecimento de um histórico de eventos de uma entidade residual, desde que o utilizador justifique a eliminação e se crie um novo evento relativo a essa ação.	Sim	Pode ser efetuado por um administrador do sistema
1.21	Quando um utilizador estiver a navegar por um conjunto de entidades, o sistema deve por defeito limitar a procura às entidades ativas, exceto se o utilizador especificar a navegação por entidades ativas e residuais.	Parcial	O Alfresco funciona dessa forma para registos, não para todas as entidades
1.22	Usar os identificadores do sistema disponibilizados pelo MoReq2010.		O Alfresco permite a criação de identificadores simples ou hierárquicos.
1.23	Criar identificadores do sistema para novas entidades segundo o UUID e permitir que esses identificadores possam ser modificados.	Parcial	Podem ser usados uuid(s) mas não podem ser modificados
1.24	Definir automaticamente a marca de dia de criação e, caso exista, a data/hora original da produção para todas as entidades.	Sim	No Alfresco estas datas são automáticas.
1.25	Permitir a um utilizador autorizado modificar a data/hora original de produção de uma entidade ativa para uma data anterior à indicada na marca de dia de criação, desde que o utilizador o justifique.	Sim	Utilizadores autorizados podem fazer alterações às entidades
1.26	Criar marcas do dia compatíveis com os formatos W3C XML e incluir informações	Sim	A informação do Fuso horário é armazenada nas datas

	sobre o fuso horário.		
1.27	Armazenar os elementos de meta-dados textuais em Unicode acompanhados pelo identificador de idioma compatível com o RFC5646 e IANA Language Subtag Registry.	Não	Os meta-dados textuais não são armazenados com identificador de idioma

ID	Funcionalidade		
2	Serviço de Utilizadores e Grupos		
2.1	Poder ser acessido por utilizadores autenticados e ter a entidade de utilizador ativa com, pelo menos, os meta-dados seguintes: identificador do sistema; data/hora de criação; registo do primeiro acesso; identificador de grupo; título; descrição; registo de eliminação; histórico de eventos; controlo de acesso; entre outros.	Sim	Os utilizadores são entidades com grande parte dos meta-dados adicionados e existem extensões para os meta-dados em falta.  O histórico é armazenado em audit, mas pode ser configurado para ficar armazenado sob a forma de versões
2.2	Permitir a criação de entidades de novos utilizadores no sistema, com os meta-dados indicados em 2.1.	Sim	
2.3	Permitir a atualização do título e da descrição da uma entidade de utilizador ativo, bem como os seus meta-dados contextuais de forma a espelhar as mudanças ocorridas com esse utilizador.	Sim	
2.4	Permitir a adição/eliminação de utilizadores activos nos grupos activos, e facilitar o registo desse evento no histórico.	Sim	
2.5	Permitir a eliminação de um utilizador que nunca tenha acessido ao Sistema para realizar qualquer função.	Sim	
2.6	Permitir a eliminação de um utilizador que tenha acessido ao Sistema para realizar uma função.	Sim	
2.7	Permitir a criação de um relatório por um utilizador autorizado, com a lista de grupos activos ao qual ele pertenceu num determinado momento.	Parcial	A informação acerca dos grupos a que o utilizador pertenceu pode ser obtida no audit
2.8	Ter capacidade de manter grupos com, pelo menos, os meta-dados seguintes: identificador	Sim	Os grupos são nodes. Parte dos meta-dados existem out of the box. Os restantes

	do sistema; data/hora de criação; registo de criação; registo de quando foi usado pela primeira vez, título; descrição; registo de eliminação; entre outros.		podem ser adicionados por extensão
2.9	Permitir a criação de novos grupos no Sistema com os meta-dados indicados em 2.8.	Sim	
2.10	Permitir a atualização do título e da descrição de um grupo ativo, bem como os seus meta-dados contextuais de forma a espelhar as mudanças ocorridas nesse grupo.	Sim	
2.11	Permitir a eliminação de um grupo que nunca teve utilizadores associados.	Sim	
2.12	Permitir a eliminação de um grupo que teve utilizadores associados.	Sim	
2.13	Permitir a criação de um relatório por um utilizador autorizado, com a lista dos utilizadores activos pertencentes a um determinado grupo num determinado momento.	Parcial	Relatório terá de ser construído a partir do audit
2.14	Permitir a um utilizador autorizado navegar e inspecionar utilizadores e grupos, para: procurar entre os utilizadores do serviço de utilizadores e grupos e inspecionar os seus meta-dados; procurar entre os grupos do serviço de utilizadores e grupos e inspecionar os seus meta-dados; procurar a partir de um utilizador os grupos aos quais está associado e inspecionar os seus meta-dados; e, procurar a partir de um grupo os utilizadores que lhe estão associados e inspecionar os seus meta-dados.	Sim	

**Tabela 30: Avaliação de Requisitos de Serviços de Pesquisa em relação ao MoReq**

ID	Funcionalidade		
3	Serviço de Perfis		
3.1	Permitir a um utilizador autorizado criar perfis com os meta-dados seguintes: identificador do sistema; registo de criação; data/hora de origem; registo da primeira utilização; título; descrição; âmbito; registo de eliminação; configurações sobre heranças; entre outros.	Parcial	Os perfis no Alfresco não são nodes (entities), mas podem ser criados por utilizadores autorizados.

3.2	Permitir a atualização do título e da descrição de um perfil ativo, bem como os seus meta-dados contextuais.	Não	
3.3	Permitir a um utilizador autorizado a criação de um perfil de administração ou geral, mas somente se esse perfil nunca foi incluído numa das entradas de permissões de acesso.	Não	
3.4	Permitir que uma função seja adicionada ou removida de um perfil ativo, assegurando que cada função está sempre associada a pelo menos um perfil ativo.	Não	
3.5	Permitir a um utilizador autorizado eliminar um perfil que nunca tenha estado associado a uma entrada de permissões de acesso, assegurando que cada função está sempre associada a pelo menos um perfil ativo.	Sim	
3.6	Permitir a um utilizador autorizado eliminar um perfil que tenha previamente estado associada a uma entrada de permissões de acesso, desde que cada função esteja associada a pelo menos um perfil.	Sim	
3.7	Permitir a um utilizador autorizado navegar e inspecionar perfis e funções, para: procurar entre os perfis do serviço de perfil e inspecionar os seus meta-dados; procurar entre as funções do serviço de perfil e inspecionar os seus meta-dados; procurar a partir de um perfil as funções que lhe estão associadas e inspecionar os seus meta-dados; e, procurar a partir das funções os perfis que lhe estão associados e inspecionar os seus meta-dados.	Não	O Alfresco permite a navegação meta-dados e pesquisas nas entidades que estão implementadas sob a forma de nodes. Os perfis não estão implementados desta forma.
3.8	Criar automaticamente uma lista com as permissões de acesso para cada serviço e para cada entidade, com o meta-dado relativo a 'Herança de perfil' e o controle de acesso de entradas para essa entidade.	Sim	As permissões são em lista e podem ser herdadas de nodes relacionados (ex: documentos de pastas).
3.9	Permitir a um utilizador autorizado ver a lista de permissões de acesso de uma entidade e verificar as entradas que contém.	Sim	
3.10	Permitir a um utilizador autorizado modificar a lista de permissões de acesso de forma a alterar o meta-dado 'Include Inherited Roles Flag' e adicionar, modificar ou eliminar as entradas de acesso com os meta-dados seguintes: identificador do utilizador ou grupo e identificador do perfil.	Sim	As permissões no Alfresco podem ser herdadas ou específicas. Existe no interface possibilidade de um utilizador autorizado modifica-las.
3.11	Permitir a qualquer utilizador ativo executar qualquer função numa entidade, desde que a função a ser realizada esteja incluída num perfil ativo e tenha sido concedida ao utilizador ou grupo ativo do qual é membro, incluindo os perfis herdados do serviço, entidade superior ou classe.	Sim	
3.12	Permitir a um utilizador autorizado verificar quais as funções que pode executar em relação	Sim	

	a qualquer entidade.		
3.13	Permitir a um utilizador autorizado criar um relatório descritivo das funções que determinado utilizador pode exercer em relação a qualquer entidade, e como tal foi estabelecido.	Sim	É possível, no entanto terá de se navegar entidade a entidade e para esta aceder às suas permissões.
3.14	Permitir a um utilizador autorizado criar um relatório descritivo das funções pertencentes a perfil específico em um determinado momento.	Parcial	Não existe um relatório que o permita, no entanto esta informação pode ser inferida a partir do audit log.
3.15	Permitir a um utilizador autorizado procurar e/ou encontrar: entidades para as quais uma entrada de acesso contém um perfil específico e entidades para as quais uma entrada de acesso contém um utilizador ou grupo específico.	Sim	Navegando nas permissões

**Tabela 31: Avaliação de Requisitos de Serviços de Classificação em relação ao MoReq**

ID	Funcionalidade		
4	Serviço de Classificação		
4.1	Incorporar a funcionalidade que gere classes de um esquema de classificação compatível com um dos módulos do MoReq2010® 200. Classification Series.	Sim	Através de scripting
4.2	Permitir a um utilizador autorizado criar novas classes com, pelo menos, os dados seguintes: identificador do sistema; registo da criação; data/hora original; registo da primeira utilização; título; descrição; notas; identificador do prazo de eliminação; data de eliminação; entre outros.	Sim	No interface do utilizador do módulo de Records Management
4.3	Permitir a um utilizador autorizado modificar o título, descrição e notas de âmbito de uma classe ativa e os seus meta-dados contextuais.	Sim	
4.4	Permitir a um utilizador autorizado alterar a data de eliminação de uma classe ativa, desde que a Tabela de Seleção nova esteja ativa e sujeita aos requisitos funcionais do serviço de eliminação, relativos ao meta-dado 'Retention Trigger Code from date of last review'.	Sim	
4.5	Permitir a um utilizador autorizado eliminar uma classe que nunca tenha sido usada para a classificação de documentos ou agregações.	Sim	
4.6	Permitir a um utilizador autorizado eliminar uma classe ativa, desde que não esteja associada a uma agregação e/ou documento ativo.	Sim	
4.7	Permitir a um utilizador autorizado navegar entre classes e pelas suas entidades associadas, para: procurar entre as classes do Serviço de Classificação e inspecionar os seus meta-dados; procurar a partir de uma classe a Tabela de Seleção e inspecionar os seus meta-dados;	Sim	

	e, procurar a partir de uma classe qualquer associação e inspecionar os seus meta-dados.		
4.8	Permitir a um utilizador autorizado substituir uma classe por outra ,em todas as agregações e documentos classificados nessa classe.	Parcial	Apenas se não gerar nenhuma incongruência.

**Tabela 32: Avaliação de Requisitos de Serviços de Registo de Documentos em relação ao MoReq**

ID	Funcionalidade		
5	Serviço de Registo de Documentos		
5.1	Permitir a um utilizador autorizado a criação de agregações com os seguintes meta-dados: identificador do sistema; registo do momento de criação; data/hora original; registo de quando foi usado pela primeira; registo de quando foi modificado pela última vez; identificador da classe; título; descrição; notas; registo da data de eliminação; meta-dados específicos de agregações de topo; meta-dados específicos de agregações secundárias; entre outros.	Sim	Alguns dos meta-dados fazem parte da base do RM. Os restantes são adicionados sob a forma de aspeto.
5.2	Possibilitar a criação de agregações de topo, desde que lhe seja atribuída uma classe no momento de criação; ou agregações secundárias, sobre a agregação de topo, desde que colocadas numa agregação ativa que ainda não contenha registos, activos ou residuais, sendo-lhe atribuída uma classe ativa ou a herde da sua agregação de topo, e não ultrapasse os níveis máximos de agregações secundárias estipulados para a agregação de topo.	Parcial	
5.3	Permitir a um utilizador autorizado a modificação do título, âmbito e descrição de uma agregação e de qualquer um dos seus meta-dados.	Sim	
5.4	Permitir a reclassificação de uma agregação através: da remoção da classe diretamente atribuída a uma agregação secundária; ou através da atribuição e/ou substituição de uma nova classificação para uma agregação.	Sim	
5.5	Permitir a um utilizador autorizado adicionar, modificar e eliminar um valor máximo de níveis de agregação numa agregação de topo, desde que o valor estabelecido não seja menor que o número de níveis de agregação existentes na agregação de topo.	Não	
5.6	Permitir a um utilizador autorizado encerrar ou abrir agregações ativas, garantindo que uma agregação não poderá ser encerrada caso todas as agregações secundárias não estejam encerradas, e facilite a eliminação das agregações encerradas se cumprirem as condições seguintes: a agregação ter sido	Sim	

	utilizada para integrar agregações e documentos; ou todas as entidades secundárias da agregação (ou agregações inferiores ou documentos) tenham sido eliminadas.		
5.7	Permitir a um utilizador autorizado eliminar qualquer agregação que nunca tenha sido usada para integrar outras agregações.	Sim	
5.8	Permitir a um utilizador autorizado mover qualquer agregação, incluindo as agregações de topo, para: uma nova agregação de topo de forma a manter a sua classificação original, desde que a agregação de topo esteja ativa e não contenha documentos activos ou residuais, e que a inclusão da agregação não exceda os níveis de agregações estabelecido para a agregação de topo onde será colocada; para uma nova agregação de topo para que adote a classificação dessa agregação, desde que a agregação de topo esteja ativa, não contenha documentos, activos ou residuais, e que a inclusão da agregação não exceda os níveis de agregações estabelecido para a agregação de topo onde será colocada; ou, para que se torne uma nova agregação de topo, mantendo a sua classificação original.	Parcial	
5.9	Permitir a um utilizador autorizado navegar e inspecionar agregações, pelo menos, das seguintes formas: procurar em todas as agregações de topo do Serviço de Registo de Documentos e inspecionar os seus meta-dados; procurar a partir de uma agregação de topo as suas agregações secundárias e inspecionar os seus meta-dados; procurar a partir das agregações secundárias a sua agregação de topo e inspecionar os seus meta-dados; procurar a partir de uma agregação a sua classe no Plano de Classificação e inspecionar os seus meta-dados; e, procurar a partir de uma agregação qualquer prazo de conservação associado e inspecionar os seus meta-dados.	Sim	
5.10	Permitir a um utilizador autorizado criar documentos numa agregação ativa que não contenha qualquer agregação, ativa ou residual, com os meta-dados seguintes: identificador do sistema; registo de criação; data/hora original; assunto; descrição; identificador de duplicado; identificador da agregação de topo; registo de agregação; identificador da classe; data de eliminação; data de início de retenção; código de ação. de eliminação; entre outros.	Sim	
5.11	Permitir a um utilizador autorizado modificar o assunto e a descrição de um documento e qualquer um dos seus meta-dados.	Sim	
5.12	Assegurar que cada documento criado numa agregação ativa e aberta herda a classe da sua agregação de topo e permitir a um utilizador autorizado reclassificar um documento no ato	Sim	

	da captura ou em qualquer outro momento, através: da atribuição direta de uma classe ativa, substituindo a sua classificação anterior e a herança da agregação; ou, retirar diretamente a classe a um documento para que este herde a classe da sua agregação.		
5.13	Permitir a um utilizador autorizado mover um documento da sua agregação de topo para qualquer outra agregação ativa e aberta, que não contenha qualquer agregação ativa ou aberta, e: permitir guardar a classificação anterior do documento através da sua aplicação direta; e, substituir a classificação anterior do documento pela classificação da nova agregação de topo, eliminando qualquer tipo de classe aplicada diretamente ao documento.	Sim	
5.14	Assegurar que o documento herda os prazos de conservação e eliminação estipulados para a sua classe sempre que um documento ativo é criado, classificado ou reclassificado, a menos que os prazos de seleção e eliminação sejam substituídos.	Sim	
5.15	Permitir a um utilizador autorizado alterar os prazos de seleção e eliminação, quer pela: aplicação direta ao documento de uma Tabela de Seleção, substituindo os prazos herdados por defeito através da agregação de topo; ou retirar prazos de seleção e eliminação aplicado diretamente a um documento, fazendo-o herdar dos prazos estabelecidos por defeito pela sua classe.	Sim	
5.16	Permitir a um utilizador autorizado fazer um duplicado de um documento, incluindo os seus meta-dados - do sistema e contextuais; lista de controlo de acessos, histórico de eventos, componentes e respetivos conteúdos.	Parcial	O histórico de eventos não é duplicado
5.17	Permitir a um utilizador autorizado navegar e inspecionar documentos, pelo menos, das seguintes formas: procurar um documento na sua agregação através da data/hora original; procurar a partir de um documento a sua agregação de topo e inspecionar os seus meta-dados; procurar a partir de um documento a sua classe, seja herdada ou por aplicação direta, e inspecionar os seus meta-dados; procurar a partir de um documento os seus prazos de conservação e eliminação, herdados ou aplicados diretamente, e inspecionar os seus meta-dados; procurar a partir de um documento as associações que possam existir e inspecionar os seus meta-dados; procurar a partir de um documento o(s) seu(s) componente(s) e inspecionar os seus meta-dados; e, procurar através de um componente o documento a que pertence e inspecionar os seus meta-dados.	Parcial	As pesquisas sobre meta-dados das classes encontram as classes permitindo navegar e ver os documentos das classes, não ficando os documentos como resultado direto da pesquisa.

5.18q	Permitir a um utilizador autorizado procurar e encontrar: agregações e/ou documentos classificados através da nomeação de classe; ou, documentos com prazos de seleção e eliminação indicados.	Sim	
5.19	Garantir que todos os documentos são criados com um ou mais componentes, implementando a funcionalidade de um dos módulos do MoReq2010® 330. Component Series, e são criados com, pelo menos, os seguintes meta-dados: identificador do sistema; registo de criação; data/hora original; título; descrição; identificador de duplicação; data de eliminação; entre outros.	Parcial	O Alfresco não implementa todas as funcionalidades do módulo 330. Os meta-dados existem ou são acrescentados sob a forma de aspeto.
5.20	Permitir a um utilizador autorizado modificar o assunto e a descrição de um componente ativo e qualquer um dos seus meta-dados contextuais.	Sim	
5.21	Sempre que o Sistema gere um evento para o componente deve incluir um identificador de registo nos meta-dados do evento, para que o evento possa aparecer no histórico de eventos do documento, bem como no histórico de eventos do componente.	Não	

**Tabela 33: Avaliação de Requisitos de Serviços de Meta-dados em relação ao MoReq**

ID	Funcionalidade		
6	Serviço de Metadados		
6.1	Cada Serviço de Metadados deve gerir definições de elementos de meta-dados, com pelo menos: identificador do sistema; título; descrição; âmbito; ordem de apresentação; ocorrências mínimas ou máximas; tipo de dados; identificador do idioma; entre outros.	Sim	Parte dos meta-dados existem de base. Os restantes são adicionados sob a forma de aspetos.
6.2	Permitir a um utilizador autorizado a criação de definições de elementos de meta-dados contextuais, com os seguintes meta-dados: registo de criação; data/hora de criação; registo da primeira utilização; registo de eliminação.	Parcial	É possível criar definições de elementos de meta-dados, no entanto quaisquer meta-dados sobre estas definições são meros comentários no ficheiro XML que os define.
6.3	Assegurar que em cada nova definição de elemento de meta-dados contextuais, o utilizador autorizado pode indicar qual o objetivo do elemento de meta-dado: mantendo uma referência a uma entidade armazenando o seu identificador de sistema ou mantenha um valor de dados válido especificado por um tipo de dados XML W3C.	Parcial	É possível identificar o objetivo, no entanto não se mantém referência à entidade.
6.4	Garantir que em cada nova definição de elemento de meta-dados contextuais, o utilizador autorizado pode especificar o número mínimo e/ou máximo de ocorrências do elemento de meta-dados.	Sim	
6.5	Garantir que cada definição de elementos de	Sim	A ordem de apresentação é parametrizada

	meta-dados é feita numa ordem de apresentação única.		e só pode diferir na alteração dos meta-dados e eventualmente na pesquisa.
6.6	Para cada definição de elementos de meta-dados activos, exceto para os elementos do sistema de meta-dados que armazenam identificadores do sistema, deve ser permitido a um utilizador autorizado especificar os elementos de meta-dados que devem ser mantidos quando a entidade a que pertencem for eliminada.	Não	Quando uma entidade é eliminada, os meta-dados também o são.
6.7	Permitir a um utilizador autorizado eliminar qualquer elemento de meta-dados de uma entidade residual, exceto se for um elemento de meta-dados do sistema que contenha um identificador de sistema, desde que seja justificada a eliminação e seja criado um registo da mesma.	Não	Não é efetuado qualquer registo da eliminação de um elemento de meta-dados.
6.8	Permitir a um utilizador autorizado modificar em qualquer definição de elementos de meta-dados definitivos, incluindo definições do sistema de elementos de meta-dados, os seguintes meta-dados: título, descrição, âmbito, ordem de apresentação, valor padrão, identificador do idioma.	Sim	
6.9	Permitir a um utilizador autorizado modificar qualquer definição de elementos de meta-dados contextuais activos, que nunca foi aplicada a uma entidade, os seguintes meta-dados: ocorrência mínima; ocorrência máxima; etiqueta modificável; etiqueta de referência de entidade; identificador de entidade; tipo de dados; etiqueta textual.	Sim	
6.10	Permitir a um utilizador autorizado excluir uma definição de um elemento de meta-dados que nunca foi aplicado a uma entidade.	Sim	
6.11	Permitir a um utilizador autorizado eliminar uma definição de um elemento de meta-dados contextual, mesmo que tenha sido aplicado a uma entidade. Deve-se assegurar que quando uma definição é eliminada: novos elementos de meta-dados com essa definição não podem ser criados ou aplicados a entidades ou elementos de meta-dados com essa definição contextual permanecem associados a entidades, mas os seus valores não podem ser modificados.	Parcial	Os elementos de meta-dados removidos deixam de estar presentes nas entidades que os utilizavam.
6.12	Permitir a um utilizador autorizado navegar e inspecionar definições de elementos de meta-dados e formulários no Serviço de Metadados e tipos de entidades, para: procurar entre as definições de elementos de meta-dados do Serviço de Metadados por ordem de apresentação e inspecionar os seus meta-dados; procurar em todos os formulários do sistema de meta-dados e inspecionar os seus meta-dados; procurar a partir de uma definição de elementos do sistema de meta-dados tipos de	Parcial	

	entidades e inspecionar os seus meta-dados; procurar a partir do tipo de entidade o sistema de definições de elementos de meta-dados associado; procurar a partir de uma definição de elementos de meta-dados contextual o formulário de meta-dados; procurar a partir de um formulário uma definição de elementos de meta-dados contextuais associada; procurar a partir de um formulário um tipo de entidade associada; e, navegar a partir de um tipo de entidade para os formulários associados.		
6.13	Sempre que sejam adicionados elementos de meta-dados a entidades, mediante a criação da entidade ou quando um modelo é aplicado a uma entidade já existente, e sempre que o valor de um elemento de meta-dados é modificado, deve garantir-se as regras seguintes: o elemento de meta-dados deve ser inicializado com o valor por defeito da sua definição; ao elemento de meta-dados nunca deve ser dado um valor incompatível com o tipo de dados da sua definição; se o elemento de meta-dados contém uma referência a uma entidade o seu valor deve representar um identificador de sistema válido pertencente a uma entidade do tipo indicado pelo identificador de tipo de entidade de referência; se o elemento de meta-dados é textual então deve ter sempre um identificador de idioma, que por defeito deriva do identificador de idioma para a definição de elementos de meta-dados; o elemento de meta-dados nunca deve ter menos valores fornecidos do que os permitidos pelo valor mínimo de ocorrências na sua definição de elemento de meta-dados; o elemento de meta-dados nunca deve ter mais valores do que os permitidos pelo valor máximo de ocorrências na sua definição de elemento de meta-dados, deve ser impedido que um valor de elemento de meta-dados seja alterado por um utilizador autorizado se o elemento de meta-dados estiver configurado de forma a não permitir alterações; deve ser impedido que o valor de elemento de meta-dados seja alterado por um utilizador autorizado caso a definição do elemento de meta-dados tenha sido eliminada.	Parcial	
6.14	Permitir a um utilizador autorizado criar formulários activos, com os seguintes meta-dados: identificador do sistema, registo de criação; data/hora original; registo da primeira utilização; título; descrição; formulário de identificador do tipo de entidade; formulário de identificador de tipo de entidade; formulário de identificador de classe; formulário de identificador de serviço; identificador de definição do elemento de meta-dados contextual; registo de eliminação; entre outros.	Sim	
6.15	Permitir a um utilizador autorizado modificar	Sim	

	os meta-dados de qualquer formulário ativo: título; descrição, identificador do tipo de entidade; identificador do serviço modelo; identificador do modelo de classe, identificador da definição dos elementos de meta-dados contextuais; entre outros.		
6.16	Permitir a um utilizador autorizado eliminar qualquer formulário que nunca foi aplicado a uma entidade.	Sim	
6.17	Permitir a um utilizador autorizado eliminar um formulário ativo, impedindo que seja aplicado a outras entidades, desde que o formulário já tenha sido aplicado a uma entidade.	Sim	
6.18	Aplicar automaticamente um formulário ativo aquando da criação de uma nova entidade, quando: a entidade é criada num Serviço, ou conjunto de Serviços, para os quais existem um ou mais formulários do tipo de entidade, indicados através da combinação do identificador do formulário de serviço e do identificador do formulário de tipo de entidade; a entidade é uma agregação criada e classificada com uma classe para a qual existem um ou mais formulário de agregação, indicados através da combinação do identificador do formulário de classe e o identificador do formulário do tipo de entidade; a entidade é um registo e é criada e classificada com uma classe para a qual existem uma ou mais formulário, indicados através da combinação do identificador do formulário de classe e o identificador do formulário de tipo de entidade.	Parcial	
6.19	Permitir a um utilizador autorizado aplicar um formulário ativo a uma entidade ativa em qualquer momento, desde que o tipo de entidade coincida com o identificador do formulário do tipo de entidade.	Não	
6.20	Garantir que os elementos de meta-dados adicionados a uma entidade nunca são apagados enquanto a entidade continuar ativa.	Não	

ID	Funcionalidade		
7	Serviço de Seleção e Eliminação		
7.1	Permitir a um utilizador autorizado criar novos prazos de conservação, com os meta-dados seguintes: identificador do sistema; registo de criação; data/hora original; registo da primeira utilização; título; descrição; mandato; âmbito; código de ação. de eliminação; identificador do	Não	Os prazos de conservação não são entidades logo a eles não podem ser adicionados meta-dados.

	código de retenção de eliminação; intervalo do código de retenção; confirmação do código de intervalo de retenção; entre outros.		
7.2	Permitir que o Código de Ação de Eliminação seja definido para um dos valores seguintes: conservação definitiva; revisão; transferência ou destruição. Sempre que o Código de Ação de Eliminação estiver definido como conservação permanente, o sistema deve assegurar que os meta-dados relativos aos períodos de conservação não são associados com a entidade.	Parcial	
7.3	Permitir que o Código de Ação de Eliminação esteja definido num dos valores seguintes: a partir de agora; a partir da data da última revisão; data original do registo; a partir da data original da agregação; a partir da data de adição na agregação; a partir da data da última adição na agregação; a partir da data de fecho da agregação; a partir da data dos meta-dados do documento ou a partir da data dos meta-dados da agregação.	Sim	
7.4	Não permitir que uma Tabela de Seleção com o Código de Ação de Eliminação a partir da última revisão seja associado por defeito a uma classe. Também não deve ser permitido que essa Tabela de Seleção seja aplicada diretamente a um documento que não tenha sido analisado previamente, exceto durante o processo de confirmação de uma decisão de revisão.	Não	
7.5	Permitir que o Código do Intervalo do prazo de conservação seja definido para um dos valores seguintes: sem prazo de conservação; dias; semanas; meses ou anos. Sempre que esse código esteja definido como sem prazo de conservação o sistema deve assegurar que os meta-dados sobre o período de duração do prazo de conservação, o código de	Sim	

	compensação do prazo de conservação e o código de compensação em meses do prazo de conservação não estão incluídos.		
7.6	Permitir que o Código de Compensação do prazo de conservação seja definido para um dos valores seguintes: sem compensação; no início do mês seguinte; no início do próximo trimestre ou no início de um determinado mês. Quando o código de compensação do prazo de conservação estiver definido como sem compensação ou começar no início do mês seguinte o sistema deve assegurar que o código de compensação em meses, do prazo de conservação, não está incluído no conjunto de meta-dados de eliminação. Contrariamente, se o código de compensação do prazo de conservação estiver definido para o início do trimestre seguinte ou de um determinado mês o código de compensação, em meses, do prazo de conservação deve estar incluído.	Parcial	
7.7	Permitir que o Código de Confirmação do intervalo do prazo de conservação esteja definido em dias ou semanas.	Sim	
7.8	Permitir a um utilizador autorizado modificar em qualquer Tabela de Seleção os meta-dados seguintes: título, descrição, mandato, âmbito e elementos de meta-dados contextuais.	Parcial	Alguns meta-dados não existem para a tabela de seleção.
7.9	Permitir a um utilizador autorizado modificar numa qualquer Tabela de Seleção, que nunca tenha sido aplicada a um documento, os meta-dados seguintes: código de ação. de eliminação; código do início do prazo de conservação; identificador do início do prazo de conservação; código do intervalo do prazo de conservação; número de duração do prazo de conservação; código de compensação do prazo de conservação; código de compensação mensal do prazo de conservação; código de confirmação do intervalo do prazo de	Parcial	

	conservação e numero de confirmação do prazo de conservação.		
7.10	Permitir a um utilizador autorizado excluir uma Tabela de Seleção que nunca foi aplicada a um documento, desde que não esteja associada a uma classe ativa.	Sim	
7.11	Permitir a um utilizador autorizado destruir uma Tabela de Seleção ativa, desde que não esteja aplicada a um documento ativo ou a uma classe ativa.	Sim	
7.12	Permitir a um utilizador autorizado navegar por todas as Tabelas de Seleção existentes e inspecionar os seus meta-dados.	Sim	
7.13	Permitir a um utilizador autorizado substituir uma Tabela de Seleção ativa por outra Tabela de Seleção ativa, aplicando tal substituição a todos os documentos activos e classes ativas que lhe estão associadas.	Parcial	
7.14	Atualizar o estado do prazo de conservação de um documento sempre que solicitado por um utilizador autorizado, bem como efetuar a atualização diária de todos os documentos activos.	Sim	
7.15	Permitir a um utilizador autorizado visualizar um alerta de confirmação de eliminação (por e-mail, SMS, serviço de assinatura, entre outros.) de documentos ou agregações sempre que uma ação. de eliminação esteja prevista.	Sim	
7.16	Permitir a um utilizador autorizado navegar e inspecionar todos os documentos activos prontos para eliminação e agrupá-los por: classe; agregação; Tabela de Seleção; código de ação. de eliminação; data da ação. da eliminação e data da confirmação da ação. da eliminação.	Parcial	É possível obter este comportamento mas o produto tem de ser modificado.
7.17	Permitir a um utilizador autorizado completar uma revisão em qualquer documento com um Código de Ação de Eliminação de Revisão que	Sim	

	<p>tenham atingido a data de eliminação, aplicando-se uma nova Tabela de Seleção e a decisão de revisão para: a revisão de um documento individual; a revisão de um conjunto de documentos; a revisão de todos os documentos associados a uma determinada Tabela de Seleção; a revisão de todos os documentos associados a uma determinada agregação (seja uma agregação de topo ou uma agregação secundária); a revisão de todos os documentos associados a uma determinada classe.</p>		
7.18	<p>Permitir a um utilizador autorizado cancelar a transferência de qualquer documento portador do Código de Ação de Eliminação de Transferência que tenha atingido a Data de Ação de eliminação, ou cancelar a eliminação de um documento com um Código de Ação de Eliminação que requeira confirmação, ou: um documento devido à transferência ou eliminação; um conjunto determinado de documentos nomeados para transferência; um conjunto determinado de documentos nomeados para a eliminação; um conjunto de documentos nomeados para transferência sob um determinado prazo de conservação; um conjunto de documentos nomeados para a eliminação sob um determinado prazo de conservação; um conjunto de documentos nomeados para a transferência numa determinada agregação; um conjunto de documentos nomeados para eliminação numa determinada agregação; um conjunto de documentos nomeados para a transferência numa determinada classe; um conjunto de documentos nomeados para a eliminação numa determinada agregação. Deve ser permitido o cancelamento de uma transferência ou eliminação antes da existência do alerta de confirmação da ação.</p>	Parcial	

7.19	<p>Permitir a um utilizador autorizado confirmar se a transferência foi concluída para qualquer documento que tenha atingido o prazo de conservação que determine a transferência, ou: a transferência de um documento individualmente, um conjunto de documentos nomeados para a transferência; um conjunto de documentos nomeados para a transferência sob uma determinada Tabela de Seleção; um conjunto de documentos nomeados para a transferência sob uma determinada agregação; um conjunto de documentos nomeados para a transferência sob uma determinada classe.</p>	Sim	
7.20	<p>Sempre que a Data da Ação de Eliminação afeta um documento, deve ser verificado se todos os seus componentes podem ser eliminados automaticamente. Se tal não for possível, o sistema deve definir um período de confirmação e permitir a um utilizador autorizado confirmar a conclusão da eliminação. O sistema deve permitir a um utilizador autorizado confirmar a exclusão do componente tanto para: documentos individuais; um conjunto de documentos nomeados que atingiram a data de eliminação; um conjunto de documentos nomeados para a eliminação sob uma determinada Tabela de Seleção; um conjunto de documentos nomeados para a eliminação sob uma agregação ou um conjunto de documentos nomeados para a eliminação sob uma determinada classe.</p>	Parcial	
7.21	<p>Permitir que um prazo de conservação possa ser aplicado a qualquer documento ativo com um Código de Ação de Eliminação que ainda não tenha atingido a data para eliminação. Nestes casos, deve ser possível alterar o Código de Ação de Eliminação e substituir a sua data de ação. de eliminação.</p>	Sim	

7.22	Após a eliminação de documentos, o sistema deve permitir eliminar automaticamente agregações sem documento activos, desde que estejam encerradas.	Sim	
7.23	Sempre que um utilizador autorizado confirma a eliminação de todos os documentos activos de uma agregação, deve ser permitido o encerramento da mesma, bem como das agregações secundárias associadas.	Sim	
7.24	Impedir qualquer alteração da Tabela de Seleção aplicada a um documento residual.	Não	

**Tabela 34: Avaliação de Requisitos de Serviços de Retenção em relação ao MoReq**

ID	Funcionalidade		
8	Serviço de Retenção		
8.1	Permitir a um utilizador autorizado criar prazos de retenção activos, com os meta-dados seguintes: identificador do sistema; registo de criação; data/hora original; registo da primeira utilização; identificador de retenção do registo; identificador de retenção da agregação; identificador de retenção da classe; título; descrição; mandato; âmbito; registo de destruição; entre outros.	Parcial	Nem todos os meta-dados existem e como os prazos não são nodes não podem ser adicionados.
8.2	Permitir a um utilizador autorizado alterar os meta-dados de um prazo de retenção, incluindo o título, descrição, mandato; âmbito e quaisquer meta-dados contextuais.	Não	
8.3	Permitir a um utilizador autorizado associar e/ou desassociar um prazo de retenção ativo a um documento, agregação ou classe.	Sim	
8.4	Impedir a destruição de um documento que: esteja diretamente associado com um prazo de retenção; pertença a uma agregação associada a um prazo de retenção; e, tenha sido classificado com uma classe associada a um prazo de conservação.	Sim	
8.5	Permitir a um utilizador autorizado eliminar um prazo de retenção que nunca tenha sido associado a documentos, agregações ou classes.	Sim	
8.6	Permitir a um utilizador autorizado levantar um prazo de retenção através da eliminação, possibilitando a eliminação de qualquer documento a ele associado.	Sim	
8.7	Permitir a um utilizador autorizado navegar e inspecionar o Serviço de Retenção, e as	Parcial	

	entidades associadas em outros Serviços, para: procurar todos os prazos de retenção do Serviço de Retenção e inspecionar os seus meta-dados ou procurar a partir de um prazo de retenção qualquer documento, agregação ou classe associada e inspecionar os seus dados.		
--	---	--	--

**Tabela 35: Avaliação de Requisitos de Serviços de Pesquisa em relação ao MoReq**

ID	Funcionalidade		
9	Serviço de Pesquisa		
9.1	Permitir ao utilizador encontrar quaisquer entidades a que tenha recebido autorização para navegar ou inspecionar.	Sim	
9.2	Permitir ao utilizador restringir o resultado da pesquisa a um ou mais tipos de entidades particulares.	Sim	
9.3	Permitir ao utilizador especificar uma consulta que compreenda a pesquisa de texto percorrendo todos os elementos de meta-dados textuais.	Sim	
9.4	Quando se realizar uma pesquisa de texto, deve estar disponível um índice de relevância para cada entidade encontrada.	Parcial	
9.5	Permitir ao utilizador especificar a pesquisa através de uma combinação de critérios, onde cada critério de pesquisa compara um elemento de meta-dados contextuais particular com o valor fornecido pelo utilizador.	Sim	
9.6	Permitir ao utilizador especificar um critério de pesquisa que devolva uma correspondência para qualquer valor específico de elemento de meta-dados.	Sim	
9.7	Permitir ao utilizador especificar um critério de pesquisa que devolva uma correspondência para meta-dados textuais com base em pesquisa de texto.	Sim	
9.8	Permitir ao utilizador especificar um critério de pesquisa que devolva uma correspondência para a data, data/hora e registo de meta-dados com base nas opções seguintes: valor que ocorre antes de uma determinada data; valor que ocorre após uma determinada data; valor que ocorre numa determinada data; valores diários/semanais/mensais/trimestrais/anuais; valores ocorridos no dia/semana/mês/trimestre/ano anterior; valores ocorridos no dia/semana/mês/trimestre/ano seguinte.	Sim	Obriga à configuração das pesquisas
9.9	Permitir a um utilizador autorizado configurar o Serviço de Pesquisa o primeiro dia da semana e o primeiro mês do ano de uma organização de forma a facilitar a pesquisa por data.	Não	

9.10	Ao pesquisar, usando critérios baseados na data, deve ser possível administrar o fuso horário local para encontrar de forma precisa valores de meta-dados correspondentes a um período ou data específica.	Não	
9.11	Permitir ao utilizador especificar um critério de pesquisa para meta-dados numéricos que correspondam ao valor fornecido, com base em qualquer uma das opções seguintes: igual, maior que ou menor que.	Sim	
9.12	Permitir ao utilizador especificar um critério de pesquisa baseado em meta-dados booleanos que verifique se o valor do elemento é verdadeiro ou falso.	Sim	
9.13	Permitir ao utilizador especificar um critério de pesquisa através de elementos de meta-dados que contenham identificadores do sistema, baseados na correspondência com a entidade fornecida pelo utilizador.	Sim	
9.14	Permitir ao utilizador especificar um critério de pesquisa para documentos, agregações e componentes dependentes de qualquer agregação.	Sim	
9.15	Permitir ao utilizador combinar diversos critérios de pesquisa baseados nas combinações dos operadores booleanos (E, OU e Não), ou mudar a ordem de precedência através do uso de parêntesis ou outro método equivalente.	Sim	
9.16	De forma a responder a expressões complexas de pesquisa, deve ser permitido ao utilizador combinar ou juntar os resultados de diversas pesquisas.	Sim	
9.17	Por definição, somente devem ser devolvidas as entidades existentes no resultado de pesquisa, a menos que o utilizador pretenda incluir entidades ativas e residuais nos resultados de pesquisa.	Não	
9.18	Quando o utilizador inicia uma pesquisa, deve poder especificar: quais os elementos de meta-dados a incluir nos resultados da pesquisa; se inclui o tipo de entidade no resultado de pesquisa de cada entidade; se inclui o índice de relevância; se inclui as entidades ativas e residuais; se a ordem do resultado de pesquisa é por relevância; que elementos de meta-dados serão usados para ordenar os resultados de pesquisa.	Não	
9.19	Para grandes conjuntos de resultados de pesquisa, deve ser implementado um método de paginação, ou outro, onde apenas um subconjunto do resultado total da pesquisa seja fornecido ao utilizador sendo os subconjuntos adicionais fornecidos quando solicitados.	Sim	
9.20	Fornecer o número total de entidades que	Sim	

	correspondam à busca, como parte dos resultados de pesquisa. Este total não deve incluir as entidades excluídas devido a permissões de acesso ou por serem entidades residuais.		
9.21	Não permitir a um utilizador não autorizado pesquisar, navegar ou aceder a entidades ou aos seus meta-dados, devendo estas serem excluídas do resultado de pesquisa.	Sim	
9.22	Sempre que um utilizador realiza uma pesquisa, deve ser criado um evento incluído no histórico de eventos da entidade do utilizador. O evento deve conter uma descrição da pesquisa realizada e o número total de entidades encontradas.	Sim	Se o audit estiver ativo
9.23	Permitir a um utilizador autorizado guardar, modificar, eliminar ou partilhar pesquisas, designadas por "pesquisas guardadas".	Sim	
9.24	Permitir a um utilizador autorizado criar relatórios detalhados com base numa pesquisa, com os itens seguintes: um cabeçalho fornecido pelo utilizador; a data e hora em que o relatório foi criado; a numeração de páginas; detalhes do Serviço de Pesquisa; detalhes do utilizador; a descrição da pesquisa usada para o relatório; o número total de resultados de pesquisa no relatório; colunas com base nos elementos de meta-dados selecionados para o relatório, entre outros.	Não	
9.25	Permitir a um utilizador autorizado criar relatórios sumários com base em consultas de pesquisa múltipla, num formato comum de relatório, com os itens seguintes: um cabeçalho fornecido pelo utilizador; a data e hora em que o relatório foi criado; a numeração de páginas; detalhes do Serviço de Pesquisa; detalhes do utilizador; a descrição da consulta de pesquisa usada para o relatório; o número total de resultados de pesquisa.	Não	
9.26	Sempre que um utilizador criar um relatório detalhado ou um relatório sumário, deve ser criado um evento no histórico de eventos da entidade do utilizador. O evento deve incluir a descrição da pesquisa realizada para criar o relatório e o número total de entidades encontradas.	Não	
9.27	Permitir a um utilizador autorizado guardar, modificar, apagar e partilhar definições de relatórios tanto para relatórios detalhados como para relatórios sumários.	Não	

**Tabela 36: Avaliação de Requisitos de Serviços de Exportação em relação ao MoReq**

ID	Funcionalidade		
----	----------------	--	--

10	Serviço de Exportação		
10.1	Permitir a um utilizador autorizado exportar entidades num ficheiro XML compatível com o esquema XML do MoReq2010®. E permitir a exportação dos dados seguintes: todos os utilizadores e grupos de utilizadores de um Serviço de Utilizadores e Grupos; qualquer grupo nomeado individualmente com acesso a entidades que pertencem a esse grupo; todos os utilizadores nomeados individualmente; todos os perfis existentes num Serviço de Perfis; qualquer perfil nomeado individualmente; todas as classes existentes num Serviço de Classificação; qualquer classe nomeada individualmente; todas as agregações e documentos, com os seus componentes, existentes num Serviço de Registo de Documentos; qualquer agregação nomeada individualmente bem como os documentos contidos nela, com os seus componentes; qualquer documento nomeado individualmente, com os seus componentes; todas as definições e modelos existentes num Serviço de Metadados; qualquer modelo indicado individualmente com as definições de elementos de meta-dados contextuais que o compõem; qualquer definição de elementos de meta-dados contextuais nomeados individualmente; todos os prazos de retenção existentes num Serviço de Retenção intermédia; qualquer prazo de retenção nomeado individualmente.	Não	Não existe mas o Alfresco pode ser estendido para o permitir.
10.2	Permitir a exportação de entidades residuais unicamente quando um utilizador autorizado optar especificamente por essa tarefa.	Não	
10.3	Determinar quais as entidades a exportar na totalidade e quais as entidades que devem exportar-se como "etiquetas". Entidades a exportar na íntegra: entidades nomeadas pelo utilizador autorizado; as entidades significativas de qualquer entidade a ser exportada na íntegra; eventos de qualquer entidade a ser exportada na íntegra. Entidades a exportar como "etiquetas": entidades descritas pelos elementos de meta-dados da entidade a exportar na íntegra, incluindo os eventos; as definições dos elementos de meta-dados contextuais para todos os elementos de meta-dados contextuais da entidade a ser exportada na íntegra; todas as entidades significativas para as entidades a exportar na íntegra; as entidades descritas nas Listas de Permissões de Acesso da entidade a serem exportadas na íntegra ou como "etiqueta".	Não	
10.4	Criar um Identificador de Exportação único para cada exportação	Não	
10.5	Permitir a um utilizador autorizado escrever	Não	

	um comentário a incluir nos dados, no âmbito da exportação e do evento de exportação.		
10.6	No processo de exportação deve ser considerada a seguinte informação: a Data/hora do início da Exportação; o Identificador de Exportação criado; os comentários incluídos na exportação; um cabeçalho de exportação com a identificação dos serviços suportados pelo sistema; os metadados e as listas de permissões de acesso para cada serviço; para cada serviço determinar as entidades a serem exportadas, agrupando as de exportação completa e as "etiquetas" e a Data/hora da conclusão com sucesso da Exportação.	Não	
10.7	Para cada entidade exportada integralmente, devem exportar-se os valores dos elementos de metadados do sistema e os elementos de metadados contextuais, bem como as suas listas de permissões de acesso.	Sim	Mas apenas no formato de exportação próprio do Alfresco
10.8	Para cada componente exportado integralmente, deve exportar-se o conteúdo do componente.	Sim	
10.9	Para cada entidade exportada como "etiqueta", deve exportar os valores dos elementos de metadados do sistema e a lista de permissões de acesso.	Não	
10.10	Para cada entidade exportada integralmente e para cada entidade exportada como "etiqueta", deve acrescentam-se no seu histórico de eventos um acontecimento que inclui: o Identificador de Exportação; a confirmação de Exportação Completa e um comentário sobre a exportação.	Não	