U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

Estudo de mapeamento entre o ISAD/ISAAR e o modelo CIDOC-CRM para a descrição de objetos culturais da Torre do Tombo

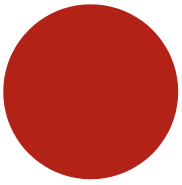
Inês Dias Koch

M

2019

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA
FACULDADE DE LETRAS



Inês Dias Koch

Estudo de mapeamento entre o ISAD/ISAAR e o modelo CIDOC-CRM para a descrição de objetos culturais da Torre do Tombo

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada pela Professora Doutora Maria Cristina Ribeiro

Faculdade de Engenharia e Faculdade de Letras

Universidade do Porto

julho de 2019

Estudo de mapeamento entre o ISAD/ISAAR e o modelo CIDOC-CRM para a descrição de objetos culturais da Torre do Tombo

Inês Dias Koch

Dissertação realizada no âmbito do Mestrado em Ciência da Informação, orientada pela Professora Doutora Maria Cristina Ribeiro

Membros do Júri

Presidente: Prof.ª Doutora Carla Alexandra Teixeira Lopes

Professora Auxiliar da Faculdade de Engenharia da Universidade do Porto

Arguente: Prof.ª Doutora Maria Manuel Lopes de Figueiredo Costa Marques Borges

Professora Associada da Faculdade de Letras da Universidade de Coimbra

Orientadora: Prof.ª Doutora Maria Cristina de Carvalho Alves Ribeiro

Professora Associada da Faculdade de Engenharia da Universidade do Porto

Agradecimentos

Ao longo deste percurso várias foram as pessoas que se foram cruzando comigo e que marcaram nesta caminhada. Desde já quero agradecer a todos os que me fizeram crescer e evoluir.

Agradecer à minha orientadora, Professora Doutora Cristina Ribeiro, que desde início se disponibilizou para me ajudar no que fosse necessário, pela orientação e todos os ensinamentos que me foi transmitindo.

A toda a equipa do projeto EPISA que sempre se mostrou prestável, em especial aos elementos da equipa do INESC TEC, por me ajudarem a desenvolver este projeto e me orientarem na direção certa, mesmo quando a relação ternária apareceu no nosso caminho.

À Maria José, da Torre do Tombo, por me ajudar nos momentos de sufoco em que pensei não conseguir realizar uma parte do projeto.

Aos investigadores do InfoLab por estarem sempre dispostos a ajudar e fazerem com que os dias passassem mais rápido.

Aos meus amigos por terem estado presentes ao longo do meu percurso académico e fazerem com que ele valesse a pena.

À minha família por estar sempre presente e me motivar a dar o melhor de mim, sem nunca desistir dos meus objetivos. Em especial à minha mãe por me fazer acreditar em mim e me apoiar em todas as decisões que tomo, deixando-me seguir os meus sonhos.

Um muito obrigada a todos.

Este trabalho é financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito do projeto DSAIPA/DS/0023/2018.

Resumo

O Arquivo Nacional da Torre do Tombo, que coordena o arquivo a nível nacional, tem na sua posse uma coleção única de objetos culturais históricos e contemporâneos que datam do século IX até a atualidade.

O arquivo tem identificado problemas no que toca à limitação da recuperação de informação relativa a entidades, visto que estas estão presentes nos campos textuais descritivos e necessitam de ser interpretadas e retiradas desses mesmos campos. As normas ISAD(G) e ISAAR(CPF), utilizadas atualmente para a representação dos objetos culturais, não respondem a todas as necessidades atuais, entre as quais se encontram a pesquisa por metadados que estão presentes nos campos textuais da norma. É, por isso, necessário encontrar um modelo de dados que permita recuperar toda a informação e ligar à rede internacional de dados. Tendo em vista a resolução deste problema, a Torre do Tombo considerou o CIDOC-CRM, modelo de dados criado e aplicado na área dos museus, como a ferramenta de metadados indicada a explorar num novo sistema de informação para os arquivos.

Este trabalho tem, por isso, como objetivo compreender no que é que consiste o CIDOC-CRM e como é que este se pode aplicar à área dos arquivos, utilizando o exemplo do Arquivo Nacional da Torre do Tombo para orientar a elaboração do modelo. Para isso serão analisadas as diversas entidades e propriedades presentes no CIDOC-CRM, de modo a ver como é que estas podem ser associadas e mapeadas no que diz respeito aos metadados utilizados atualmente nas normas de arquivo (ISAD(G) e ISAAR(CPF)).

Do presente trabalho resultou a ontologia *ArchOnto*, o novo modelo de dados para os arquivos. Esta tem como base o CIDOC-CRM e a ontologia *Data Object*, criada para se fazer a validação dos dados a inserir nos campos de descrição arquivística. Resultou ainda uma primeira caracterização dos objetos culturais da base de dados do *Digitarq*, base de dados onde se encontram os registos de todos os arquivos nacionais.

Palavras-Chave - ISAD(G), ISAAR(CPF), CIDOC-CRM, Arquivo Nacional da Torre do Tombo

Abstract

The Torre do Tombo National Archive, which coordinates the archive at national level, has in its possession a unique collection of historical and contemporary cultural objects that date from the ninth century to the present day.

The archive has identified problems with the limitation of the retrieval of entity information, since these are present in the descriptive textual fields and need to be interpreted and removed from those same fields. The ISAD(G) and ISAAR(CPF) standards, currently used for the representation of cultural objects, do not respond to all current needs, among which are the search for metadata that are present in the text fields of the standard. It is therefore necessary to find a data model that allows retrieving all the information and connecting to the international data network. In order to solve this problem, Torre do Tombo considered the CIDOC-CRM, a data model created and applied in the area of museums, as the metadata tool indicated to explore in a new information system for archives.

This work has, therefore, to understand what CIDOC-CRM is and how it can be applied to the archives area, using the example of the Torre do Tombo National Archive to guide the elaboration of the model. In order to do this, we will analyse the different entities and properties present in the CIDOC-CRM model, in order to see how these can be associated and mapped with respect to the metadata currently used in the ISAD(G) and ISAAR(CPF)).

The present work resulted in the *ArchOnto* ontology, the new data model for the archives. This is based on the CIDOC-CRM and the *Data Object* ontology, created to validate the data to be inserted in the archival description fields. It also resulted in a first characterization of the cultural objects of the *Digitarq* database, a database of the records of all national archives.

Key words - ISAD(G), ISAAR(CPF), CIDOC-CRM, Torre do Tombo National Archive

Lista de Figuras e Tabelas

Siglas e Abreviaturas

- ANTT - Arquivo Nacional da Torre do Tombo
- CIDOC - International Committee for Documentation
- CRM - Conceptual Reference Model
- DGLAB - Direção Geral do Livro, dos Arquivos e das Bibliotecas
- FRBR - Functional Requirements for Bibliographic Records
- ICA - International Council on Archives
- ICOM - International Council of Museums
- ISAAR(CPF) - International Standard Archival Authority Record for Corporate Bodies, Persons and Families
- ISAD(G) - General International Standard Archival Description
- ODA - Orientações para a Descrição Arquivística
- RiC-CM - Records in Context Conceptual Model

Sumário

1. Introdução

O Arquivo Nacional da Torre do Tombo, uma das instituições mais antigas de Portugal, foi, ao longo dos anos, coletando uma coleção única de objetos históricos e contemporâneos. Na sua posse tem documentos originais desde o século IX até à atualidade, preservando também os novos arquivos eletrónicos no âmbito de atuação dos organismos da Administração Pública. Tem ainda o mandato explícito para dar execução à lei que estabelece as bases da política e do regime de proteção e valorização do património cultural, na sua vertente de património arquivístico e património fotográfico.

Com a grande quantidade de documentos que a esta instituição compete preservar, há a utilização de metadados descritivos que são a sua descrição. No entanto, os metadados não possibilitam, por exemplo, a pesquisa dos diversos documentos, uma vez que não recupera todos os documentos existentes. Isto torna-se essencial quando, ao executar uma pesquisa, os metadados são a única maneira de alcançar estes mesmos objetos, porque é através da descrição dos documentos que a pesquisa é elaborada. Com os metadados consegue-se encontrar os diversos tipos de documentos que a esta instituição dizem respeito, como o caso dos documentos digitais com descrição e dos documentos físicos com e sem descrição.

Para além da realização de pesquisas, as descrições dos documentos garantem a autenticidade[1] e o contexto dos mesmos, dando-lhes um valor probatório. Assim será possível obter dados que não estão no documento, mas que são essenciais para o interpretar.

Esta instituição tem vindo, ao longo dos anos, a fazer a descrição de todos os documentos que tem em sua posse, sendo que esta é uma das tarefas que tem mais recursos humanos alocados. Tendo em conta que esta instituição presta mais serviços à comunidade, desde a pesquisa científica ao acesso legal a documentos, a renovação da infraestrutura de dados existentes para a descrição arquivística representa um dos maiores desafios. Esta renovação é um dos maiores projetos do Arquivo Nacional da Torre do Tombo, sendo que estes projetos foram o que originou este trabalho.

[1] Projeto InterPARES, o qual visa o desenvolvimento do conhecimento essencial para a preservação, a longo prazo, de registos autênticos criados ou mantidos em formato digital e fornecer a base para padrões, políticas, estratégias e planos de ação capazes de assegurar a longevidade de tal material e a capacidade de seus utilizadores confiarem na sua autenticidade. *Website* InterPARES – consultado pela última vez a 30/06/2019 – Disponível em: - http://www.interpares.org/

A Torre do Tombo tem utilizado as normas ISAD(G) e ISAAR(CPF) para a descrição dos seus objetos culturais, no entanto estas têm vindo a mostrar algumas limitações que necessitam de ser ultrapassadas. Este trabalho identifica limitações nos modelos de descrição existentes e contribui para a definição e avaliação de um novo modelo, tendo em vista a representação dos registos de descrição dos arquivos, a sua exploração por profissionais e por leigos e a sua ligação a outras coleções e informação relacionada.

1.1 Enquadramento da dissertação

O projeto EPISA (*Entity and Property Inference for Semantic Archives*) tem como parceiros o INESC TEC, a DGLAB (Direção Geral do Livro, dos Arquivos e das Bibliotecas) e a Universidade de Évora. Este projeto visa, entre outros objetivos, o desenvolvimento um protótipo para uma plataforma de grafo de conhecimento de código aberto, adotando o modelo de dados para descrição arquivística utilizado pela DGLAB. Este projeto também se dedica ao desenvolvimento de algoritmos para garantir a migração efetiva de conteúdos armazenados de acordo com os padrões ICA para um modelo baseado em ontologias, exigindo o uso de cross-walks e a inferência de novas relações com métodos semi-automatizados (Almeida and Runa 2018).

O projeto EPISA encontra-se na sua fase inicial, visto que começou no início de 2019 e termina em 2021. É no âmbito deste projeto que surge o presente trabalho, o qual corresponde ao primeiro semestre do EPISA. Ao longo deste trabalho vai ser desenvolvido um novo modelo de descrição arquivística.

Daqui em diante, o presente trabalho será denominado de projeto, sendo que o projeto EPISA será denominado apenas de EPISA.

1.1.1 Problemas

No início do desenvolvimento deste projeto foi necessário compreender quais os principais problemas com que se deparam os arquivos, refletindo seguidamente nas possíveis soluções.

Primeiramente, foi necessário compreender no que consiste o atual modelo utilizado na Torre do Tombo, de modo a perceber o porquê de este não corresponder a todas os requisitos atuais. Com isto pôde-se concluir que um dos principais problemas são as limitações do modelo atual em relação aos campos existentes.

Seguidamente, outro problema verificado foi a dificuldade em recuperar informação relevante, pois é limitada a capacidade para captar as ligações entre os diversos elementos descritos (objetos, produtores, autores). Isto acontece, porque os metadados existentes não têm uma relação entre si, como no caso de Alexandre Herculano autor, com Alexandre Herculano indivíduo singular.

Existe ainda a necessidade de acompanhar o crescimento sistemático de registos digitais que deve de ser acompanhado por metadados que permitam satisfazer os requisitos de interoperabilidade em redes semânticas.

Para além destes problemas, o problema geral dos arquivos passa pela abertura dos dados de descrição arquivística existentes no Arquivo Nacional da Torre do Tombo, o qual tem em sua posse cerca de 3.5 milhões de registos, à abertura ao acesso como dados abertos.

Com todos estes problemas identificados pela Torre do Tombo, a solução neste projeto passa pela exploração e experimentação do modelo de dados CIDOC-CRM como ferramenta de metadados. Esta ferramenta, criada no âmbito dos museus, possibilita uma descrição mais rica, uma vez que as descrições arquivísticas terão por base uma descrição mais flexível, completa e refinada.Com isto pretende-se dizer que haverá o atomizar dos registos, representando a informação dos registos arquivísticos de modo a que as entidades e relações estejam mais percetíveis. Haverá ainda um grau mais fino na representação, onde as representações textuais terão, por exemplo, as entidades temporais e os indivíduos identificados.

1.1.2 Objetivos

Para a elaboração do novo modelo de descrição arquivística será tido em conta um objetivo principal, o qual consiste na elaboração de um modelo de dados baseado no CIDOC-CRM para a descrição de objetos culturais do Arquivo Nacional da Torre do Tombo. Para garantir que este objetivo é alcançado, será necessário, primeiramente, ter em conta os seguintes objetivos específicos:

1. **Análise** do CIDOC-CRM, considerando os conteúdos já desenvolvidos pela DGLAB;
2. **Construção** de um modelo CIDOC-CRM atendendo as normas ISAD/ISAAR, de modo a saber quais os conteúdos do atual modelo podem ser incorporados no futuro modelo;

3. **Experimentação** do modelo elaborado - Identificação dos metadados de épocas e arquivistas diferentes e elaborar uma primeira experiência; Escolha de um pequeno conjunto de registos relevantes (fundos, épocas e técnicas) e fazer a descrição em CIDOC-CRM, verificando como se pode tornar o processo sistemático para migrar de ISAD/ISAAR para o CIDOC-CRM; Observar se o modelo é capaz de acomodar a informação que já se encontra no modelo ISAD/ISAAR. Tudo isto leva à afinação do modelo e à identificação do conjunto de dificuldades na migração;
4. **Avaliação** do modelo, analisando as perguntas frequentes utilizadas na pesquisa pelos utilizadores, com a finalidade de melhorar o modelo;
5. **Validação** do modelo de dados proposto no domínio dos conteúdos culturais e da administração pública com a equipa da DGLAB. Para isto será necessário recorrer a experiências com alguns arquivistas da DGLAB para estimar as necessidades de revisão da migração.

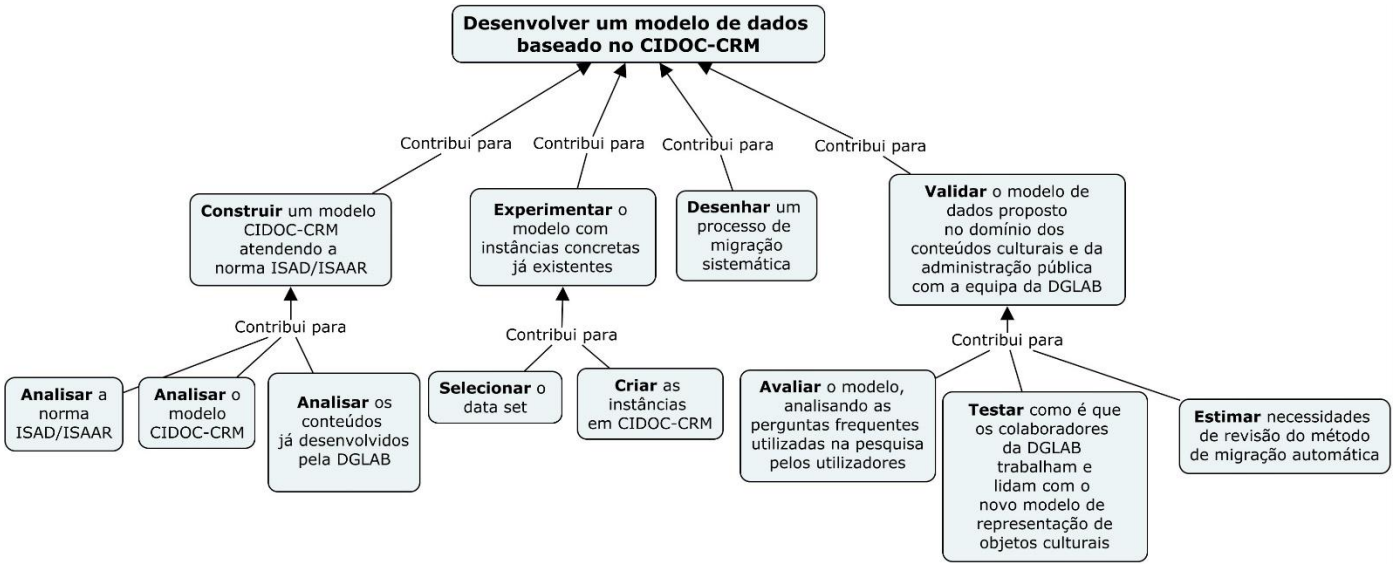


Figura 1 – Árvore de Objetivos

Fonte: Autoria Própria

1.2 Abordagem metodológica

Tendo em conta que este trabalho se foca no mapeamento entre as normas atuais de descrição arquivística da Torre do Tombo, ISAD(G) e ISAAR(CPF), para o CIDOC-CRM, existem duas metodologias aplicadas neste projeto, uma ligada à parte teórica, estado da arte, e outra ligada à parte prática, mapeamento do modelo em si.

Na parte teórica, a metodologia utilizada baseia-se na pesquisa exploratória. Esta metodologia, que tem como finalidade inteirar o autor acerca de um determinado tema, permite ter um maior conhecimento acerca dos temas abordados na dissertação, entre os quais se encontram as normas ISAD(G) e ISAAR(CPF) e o CIDOC-CRM. Para além de adquirir conhecimento em relação às normas utilizadas nesta dissertação, ainda permite a consciencialização acerca do modo como atualmente são elaboradas as descrições arquivísticas pelo Arquivo Nacional e como é que estas são apresentadas aos utilizadores, através do portal de pesquisa disponibilizado pela Torre do Tombo, *Digitarq*.

Já na parte prática, a metodologia utilizada baseia-se no método iterativo incremental (Figura 2). Este método, muito utilizado no desenvolvimento de *software*, caracteriza-se pela formalização do projeto de forma gradual (Cordeiro 2003). Com isto pretende-se dizer que, à medida que o projeto vai sendo desenvolvido, este vai estando sujeito a avaliação e validação, havendo assim, sempre que necessário, o refazer de um passo para que se consiga chegar ao modelo desejado. Esta metodologia permite que se vá refinando o modelo que se está a desenvolver, de modo a que o produto final do projeto seja conforme os requisitos estabelecidos.

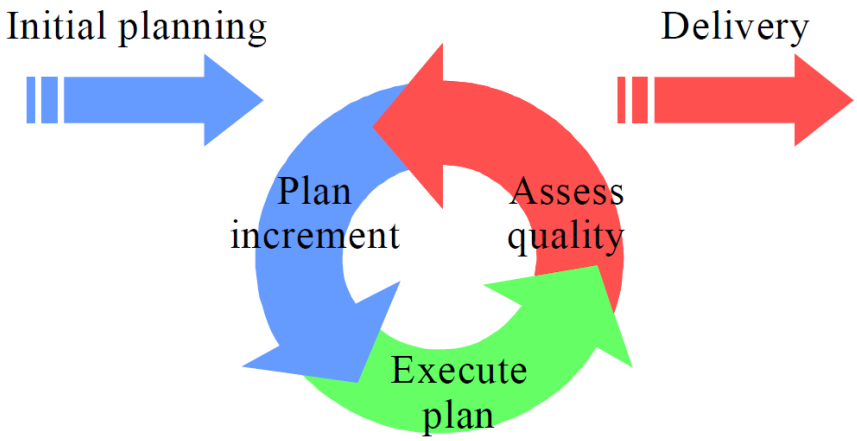


Figura 2 – Método Iterativo-Incremental

Fonte: (Börstler 2001)

Esta metodologia, como se pode observar na figura acima, é cíclica e ágil, sendo que as conclusões de um ciclo levam ao desenvolvimento de um outro ciclo, o que permite a existência de mudanças no que foi elaborado até então, havendo, assim, uma melhoria sistemática do resultado.

Para se trabalhar segundo esta metodologia vão sendo tidos em conta diversos exemplos de documentos que se encontram, atualmente, descritos de acordo com as normas de descrição arquivística ISAD(G) e ISAAR(CPF).

Com isto, podemos dizer que a metodologia utilizada neste projeto passa por um modelo sistemático com avaliação e utilização de casos reais.

Para além das metodologias referidas, para a elaboração do presente projeto foi fundamental a utilização de tecnologias que permitissem a realização do modelo de dados para o Arquivo Nacional da Torre do Tombo. Para isso foi utilizado o Protégé, tecnologia de desenvolvimento de ontologias, para a interpretação da ontologia do CIDOC-CRM e criação da nova ontologia para o ANTT.

1.3 Estrutura da dissertação

A dissertação encontra-se dividida em 5 capítulos, os quais serão apresentados seguidamente. Após estes capítulos existem ainda duas secções, a primeira destinada às referências bibliográficas e a segunda aos anexos.

O primeiro capítulo, destinado à *Introdução*, tem como propósito elucidar o leitor em relação ao contexto da presente dissertação, mostrando no que é que esta consiste, fazendo o seu enquadramento. No enquadramento da dissertação é possível constatar quais as problemáticas que levaram ao desenvolvimento deste trabalho, os objetivos e resultados esperados e a metodologia utilizada ao longo de todo o projeto.

O segundo capítulo, denominado *Normas de descrição para arquivo*, destina-se ao estado da arte. Este encontra-se dividido em quatro tópicos, sendo que o primeiro é relativo à norma ISAD(G). Neste são apresentados os conceitos da norma, no que é que consiste a descrição multinível, característica que a distingue de outras normas, e a estrutura desta norma. Por sua vez, o segundo tópico, pretende explicitar a norma ISAAR(CPF), mostrando no que é que esta consiste e como é que se encontra estruturada. Seguidamente, no terceiro tópico é abordada a norma CIDOC-CRM, explicitando no que é que consiste o modelo, a hierarquia das classes e propriedades e ainda os diversos modelos

de CRM existentes. Seguidamente, no quarto tópico, são estudadas diversas aplicações do CIDOC-CRM, sendo aqui apresentadas a adoção do CIDOC-CRM em diversas áreas, incluindo nos arquivos. Por fim, no quinto tópico é apresentada a A*nálise das ferramentas online do ANTT*, onde são analisados o *website* e o portal de pesquisa da Torre do Tombo.

Por sua vez, o terceiro capítulo, denominado *Coleção Torre do Tombo*, tem como objetivo conhecer a coleção de registos existentes na base de dados da *Digitarq*, base de dados onde estão inseridos os registos de todos os arquivos a nível nacional. Para se conhecer a coleção do Arquivo Nacional da Torre do Tombo foram tidos em conta, primeiramente, as características para a análise da base de dados do *Digitarq* e, seguidamente a caracterização da sua coleção onde foram elaboradas algumas interrogações para se analisar a base de dados.

No quarto capítulo dá-se a conhecer o desenvolvimento do modelo de dados para a Torre do Tombo, mais especificamente o processo que se desenvolveu para se chegar a este mesmo modelo. Aqui será apresentado o estudo de mapeamento dos campos ISAD(G) para as classes e propriedades do CIDOC-CRM, no tópico *ISAD(G) para CIDOC-CRM*. É também apresentado o desenvolvimento das ontologias criadas para representar o novo modelo de dados, no tópico *Ontologia*. Este capítulo conta, ainda, com a *Extração de metadados de descrições ISAD(G)*, com o objetivo de saber quais os metadados presentes nas descrições do ANTT que podem vir a ser explorados como *Linked Open Data*.

No quinto capítulo são tidas em conta as *Conclusões* do trabalho desenvolvido e os conhecimentos que se obtiveram com este. Para além disso, são apresentados os diversos desafios que se foram encontrando ao longo do percurso da dissertação e como é que estes foram sendo contornados e solucionados. Por fim, são apresentadas tarefas a ser realizadas em trabalho futuro.

Para terminar, no final da dissertação são apresentadas as referências bibliográficas tidas como suporte para elaboração deste trabalho e os anexos, os quais complementam a dissertação.

A par da minha dissertação, foi desenvolvida uma outra, dentro do EPISA, intitulada de "*ArchGraph: Design of a vertical prototype infrastructure for semantic archives*". Esta dissertação, desenvolvida pelo Nuno Miguel Cardoso Lopes de Freitas, aluno do Mestrado Integrado em Engenharia Informática, visa na elaboração do grafo de conhecimento, um dos objetivos do EPISA.

2. Normas de descrição para arquivo

Atualmente, o Arquivo Nacional da Torre do Tombo encontra-se regido pelas normas arquivísticas ISAD(G) e ISAAR(CPF), na segunda versão, estando estas conjugadas com as normas nacionais, nomeadamente pelas Orientações para a Descrição Arquivística (ODA). Para que seja possível a elaboração da dissertação, e tendo em conta as normas que nesta vão ser analisadas e utilizadas, (ISAD(G), ISAAR(CPF) e CIDOC-CRM), demonstra-se, desde logo, fundamental a compreensão de cada uma destas.

A ISAD(G) e a ISAAR(CPF) são normas que se complementam, devendo, por isso, ser utilizadas de forma conjugada, com o objetivo de potenciar o trabalho de descrição e a posterior recuperação da informação (Direcção Geral De Arquivo 2007).

Por sua vez, o CIDOC-CRM é uma ontologia formal de referência que representa, através de Entidades e Propriedades, o domínio dos museus. Esta procura responder a uma crescente demanda por pesquisa orientada, estudos comparativos, transferência e migração de dados entre fontes heterogéneas de conteúdos culturais (Santos 2016).

2.1 ISAD(G)

A ISAD(G), norma internacional de descrição arquivística, estabelece diretrizes gerais para a preparação de descrições arquivísticas que podem ser aplicadas em qualquer tipo de documento, independentemente da sua dimensão, forma ou suporte. Esta norma, foi elaborada pelo *International Council on Archives* (ICA) e está em vigor a sua segunda edição, que data de 1999.

Segundo a norma, "o objetivo da descrição arquivística é identificar e explicar o contexto e o conteúdo de documentos de arquivo a fim de promover o acesso aos mesmos". Com este objetivo, esta norma permite acompanhar o processo dos documentos de arquivo ao longo de toda a sua vida, podendo este começar aquando da produção dos documentos ou mesmo antes desta, influenciando os sistemas de informação que vão produzir os documentos.

Tendo em conta a natureza dinâmica dos registos dos documentos, esta pode ser submetida a alterações, uma vez que com um maior conhecimento do seu conteúdo e contexto de criação, a descrição arquivística pode ser enriquecida.

2.1.1 Conceitos da norma

Os conceitos essenciais que são identificados para a descrição arquivística serão apresentados em seguida, tendo em conta a norma ISAD(G) (International Council on Archives 2002). Considerando que estes não são os únicos conceitos da norma, os restantes encontram-se disponíveis no Anexo 1.

- Descrição arquivística (*archival description*) - A elaboração de uma representação exata de uma unidade de descrição e das partes que a compõem, caso existam, através da recolha, análise, organização e registo de informação que sirva para identificar, gerir, localizar e explicar a documentação de arquivo, assim como o contexto e o sistema de arquivo que a produziu. Este termo também se aplica ao resultado desse processo.

- Documento (*document*) - Informação registada num suporte, independentemente das características deste. (Ver tb. Documento de arquivo).

- Documento de arquivo (*record*) - Informação de qualquer tipo, registada em qualquer suporte, produzida ou recebida e conservada por uma instituição ou pessoa no exercício das suas competências, ou atividades.

- Nível de descrição (level of description) - Posição de uma unidade de descrição na hierarquia de um fundo.

- Unidade de descrição (*unit of description*) - Documento ou conjunto de documentos, sob qualquer forma física, tratado como um todo e que, como tal, serve de base a uma descrição singular.

2.1.2 Descrição Multinível

A norma ISAD(G) estabelece uma descrição multinível e uniforme. Com isto, os descritores que são utilizados para cada unidade de descrição são os mesmos quer se trate de um fundo, uma série ou um documento. No entanto cada um dos descritores tem, ou pode ter, uma descrição independente, porque é uma unidade diferente. Este tipo de descrição baseia-se na representação do geral para o particular, representando o "contexto e a estrutura hierárquica do fundo e suas partes componentes".

Com uma descrição multinível, é desde logo fundamental saber quais é que são os níveis de descrição recomendados. Estes níveis são os seguintes (International Council on Archives 2002):

- **Fundo** (*Fonds*) - Conjunto de documentos de arquivo, independentemente da sua forma ou suporte, organicamente produzido ou acumulado e utilizado por uma pessoa singular, família ou pessoa coletiva, no decurso das suas atividades e funções.

- **Subfundo** (*Subfonds*) - Subdivisão de um fundo compreendendo um conjunto de documentos relacionados que corresponde a subdivisões administrativas da agência ou instituição produtora ou, quando tal não é possível, correspondendo a uma divisão geográfica, cronológica, funcional ou agrupamento de documentos similares. Quando o organismo produtor tem uma estrutura hierárquica complexa, cada secção tem tantas subdivisões subordinadas quantas forem necessárias, de modo a refletir os níveis da estrutura hierárquica da unidade administrativa subordinada primária.

- **Série** (*Series*) - Conjunto de documentos organizados de acordo com um sistema de arquivo e conservados como uma unidade, por resultarem de um mesmo processo de acumulação, do exercício de uma mesma atividade, por terem uma tipologia particular, ou devido a qualquer outro tipo de relação resultante do processo de produção, receção ou utilização. É também designada como série documental (*records series*).

- **Subsérie** (*Sub-series*) - Subdivisão de uma Série determinada pela sua ordem original ou por exigências de preservação [2].

- **Processo** (*File*) - Unidade organizada de documentos agrupados, quer para utilização corrente pelo seu produtor, quer no decurso da organização arquivística, por se referirem a um mesmo assunto, atividade ou transação. Um processo é geralmente a unidade básica de uma série.

- **Peça** (*Item*) - A mais pequena unidade arquivística intelectualmente identificável, por exemplo: carta, memorando, relatório, fotografia, registo sonoro.

[2] Multilingual Archival Terminology – consultado pela última vez a 30/12/2018 - Disponível em: http://www.ciscra.org/mat/mat/term/4069

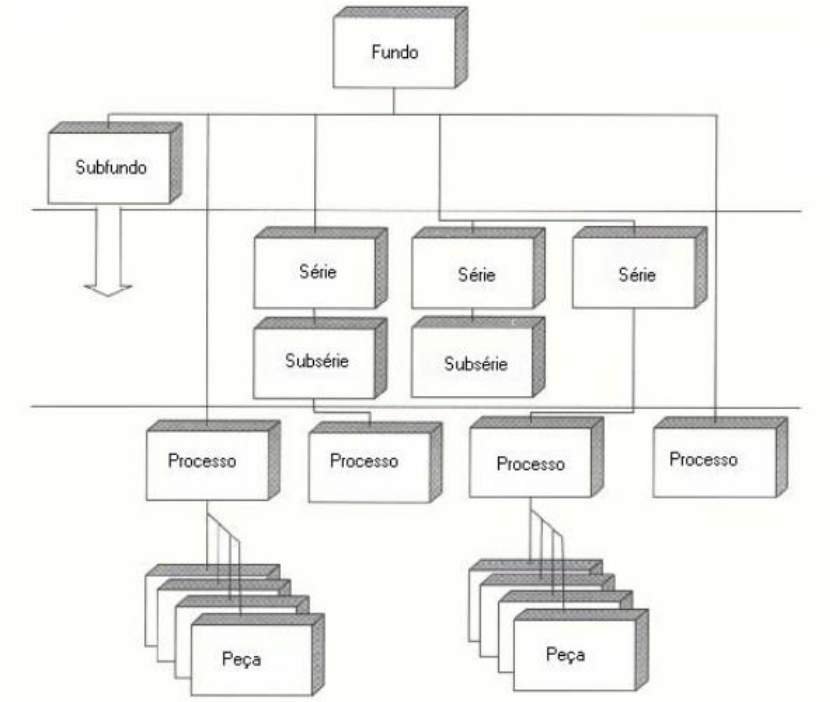


Figura 3 - Modelo dos níveis de arranjo de um fundo

Fonte: (International Council on Archives 2002)

2.1.3 Estrutura da Norma

A norma ISAD(G) encontra-se estruturada em sete zonas distintas de descrição, estando estas compostas por diversos elementos, havendo um total de 26 elementos distintos. Todos os elementos de descrição apresentam uma mesma estrutura, sendo que esta inclui o nome do elemento, o seu objetivo e as regras para a sua aplicação, assim como exemplos ilustrativos dessas mesmas regras.

As zonas de descrição existentes são as seguintes:

- Zona da identificação - Zona onde se encontram os metadados essenciais para a identificação da unidade de descrição. Aqui estão presentes o(s) código(s) de referência, o título, a(s) data(s), nível de descrição e dimensão e suporte.

Exemplo:

- **Código de referência**: PT-TT-JC
- **Título**: Junta do Comércio
- **Datas de produção**: 1755-1834
- **Nível de descrição**: Fundo

- **Dimensão e suporte** (quantidade, volume ou extensão): 829 u.i.[3] (449 liv., 380 mç.); papel

- Zona do contexto - É nesta zona que são registados os metadados sobre a origem e a custódia da unidade de descrição. Estão, assim, presentes o nome do(s) produtor(es), história administrativa/biográfica, história custodial e arquivística e fonte imediata de aquisição ou transferência.

Exemplo:

- **Nome do Produtor**: Real Junta do Comércio, Agricultura, Fábricas e Navegação destes Reinos e seus Domínios. 1755-1834
- **História administrativa**: ----------
- **História custodial e arquivística**: A documentação foi preparada pela Comissão encarregue de dar cumprimento ao Decreto de extinção da Junta do Comércio. Parte da documentação foi, então, entregue a diversas instituições, nomeadamente, Ministério do Reino, Ministério dos Negócios Estrangeiros, Tesouro Público, Tribunal do Comércio, Companhia de Seguros Bonança, Alfândega de Lisboa; a relativa à Aula do Comércio foi confiada ao Comissário dos Estudos. A restante deu entrada na Torre do Tombo entre Janeiro e Abril de 1835. Num relatório de José Feliciano de Castilho sobre o Arquivo da Torre do Tombo elaborado em 21 de Janeiro de 1843, cujo anexo n.º 2 apresenta a "relação das repartições e mosteiros extintos de que vieram papéis e livros para o arquivo", é referido que da Junta do Comércio entraram 714 maços e 401 livros. A documentação em causa ficou depositada no edifício do antigo Mosteiro de S. Bento da Saúde, onde funcionava, à época, o Arquivo Nacional da Torre do Tombo, tendo aí permanecido até 1990, altura em que, no âmbito da reinstalação do Arquivo Nacional, esta documentação foi transferida para as actuais instalações.

- Zona do conteúdo e estrutura - Zona onde se encontram descritos os metadados relativos à organização e ao conteúdo da unidade de descrição. Aqui estão expostos o âmbito e conteúdo, avaliação, seleção e eliminação, ingresso(s) adicional(ais) e sistema de organização.

[3] Unidade de instalação (u.i) - é o conjunto de documentos agrupados ou conservados numa mesma unidade física de cotação, instalação e inventariação. Não corresponde a uma unidade intelectual. São unidades de instalação: caixas, maços, livros, rolos, cadernos, pastas, disquetes, bobinas, cassetes, capa ou *dossier*, disco óptico, volume, etc. (Direcção Geral De Arquivo 2007)

Exemplo:

- **Âmbito e conteúdo**: Documentação com informações para a história económica do período pombalino, fim do século XVIII e início do XIX, nos mais diversos aspectos: comércio interno e externo, de retalho, fiscalização alfandegária, tráfego marítimo, indústria, obras públicas, sendo de realçar a sua importância para o estudo do comércio ultramarino, nomeadamente no que respeita ao Brasil, embora também exista bastante documentação referente à Ásia e à África.
- **Avaliação, selecção e eliminação**: ----------
- **Ingressos adicionais**: ----------
- **Sistema de organização**: O plano de classificação adoptado corresponde a uma estrutura orgânico-funcional, baseado nos relatórios da "Comissão encarregada de dar cumprimento ao Decreto de extinção da Junta do Comércio".

- Zona das condições de acesso e de utilização - Zona destinada ao registo de metadados sobre a acessibilidade e disponibilidade da unidade de descrição. Os elementos que aqui estão presentes são as condições de acesso, condições de reprodução, idioma/escrita, características físicas e requisitos técnicos e instrumentos de descrição.

Exemplo:

- **Condições de acesso**: Comunicável sem restrições legais.
- **Condições de reprodução**: Constantes no regulamento interno que prevê algumas restrições tendo em conta o tipo dos documentos, o seu estado de conservação ou o fim a que se destina a reprodução de documentos, analisado, caso a caso, pelo Serviço Núcleo de Transferência de Suportes, de acordo com as normas que regulam os direitos de propriedade do IA /TT e a legislação sobre direitos de autor e direitos conexos.
- **Idioma/Escrita**: ----------
- **Características físicas e requisitos técnicos**: Decorrente da necessidade de preservar originais, alguns documentos só são comunicáveis em microfilme.
- **Instrumentos de descrição**: AZEVEDO, Pedro A. de; BAIÃO, António - "Junta do Comercio". in O Arquivo da Torre do Tombo: sua história, corpos que o compõem e organização. Lisboa: ANTT; Livros Horizonte, 1989. (Fac-Símile). p. 167-171. Reprodução fac-similada da edição de 1905. PORTUGAL. Instituto dos Arquivos Nacionais / Torre do Tombo. Direcção de Serviços de Arquivística - "Junta do Comércio". In Guia Geral dos Fundos da Torre do Tombo: *Instituições do Antigo Regime, Administração Central* (2). Lisboa: IAN/TT, 1999. vol. 3. (Instrumentos de Descrição Documental). ISBN 972-8107-60-9. p. 1-34. Acessível no IAN/TT, IDD (L. 602). SERRÃO, Joel; LEAL, Maria José da Silva; PEREIRA, Miriam Halpern - "Junta do Comércio". In *Roteiro de Fontes da História*

Portuguesa Contemporânea: Arquivo Nacional da Torre do Tombo. Col. de Ana Maria Cardoso de Matos; Maria de Lurdes Henriques. Lisboa: Instituto Nacional de Investigação Científica, 1984. vol. 1. p. 256-276.

- Zona da documentação associada - Tal como o próprio nome indica, aqui são apresentados os metadados acerca dos documentos ou outras fontes com uma relação importante com a unidade de descrição. Aqui encontra-se a existência e localização de originais, existência e localização de cópias, unidades de descrição relacionadas e, ainda, nota de publicação.

Exemplo:

- **Existência e localização de originais**: ----------
- **Existência e localização de cópias**: ----------
- **Unidades de descrição relacionadas**: Relação completiva: Portugal, Torre do Tombo, Ministério dos Negócios Estrangeiros (PT-TT-MNE) Relação complementar: Portugal, Arquivo Histórico do Ministério das Obras Públicas, Junta do Comércio.
- **Nota de publicação**: FRUTUOSO, Eduardo; GUINOTE, Paulo; LOPES, António - O movimento do porto de Lisboa e o comércio luso-brasileiro: 1769-1836. Lisboa: Comissão Nacional para as Comemorações dos Descobrimentos Portugueses, 2001. 795p. ISBN 972-787-048-1. MADUREIRA, Nuno Luís - Mercado e privilégios: a indústria portuguesa entre 1750 e 1834. Lisboa: Estampa, 1997. 514 p. ISBN 972-33-1330-8. PEDREIRA, Jorge Miguel - Estrutura industrial e mercado colonial: Portugal e Brasil (1780-1830). Linda-a-Velha: [s.n.], 1994. 582 p. ISBN 972-29-0305-5. RATTON, Diogo - Reflexões sobre a Junta do Comércio, sobre as alfândegas, sobre os depósitos, e sobre as pautas. Lisboa: Imprensa Nacional, 1821. 8 p. Acessível na Biblioteca Nacional, S.C. 5604/16A. SANTANA, Francisco - Documentos do cartório da Junta do Comércio respeitantes a Lisboa: 1755-1834. Lisboa. Câmara Municipal de Lisboa.

- Zona das notas - Nesta zona é possível acrescentar metadados (especializados ou não) que não poderiam ser acrescentados em mais nenhuma das zonas de descrição referidas. Nesta zona apenas existe um elemento, denominado de notas.

- Zona do controlo da descrição - zona onde são registados metadados relativos à informação sobre quando, como e por quem foi elaborada a descrição arquivística. Aqui estão presentes a nota do(s) arquivista(s), regras ou convenções e data(s) da(s) descrição(ões).

Exemplo:

- **Nota do arquivista**: Descrição elaborada por Joana Braga.

Fontes para a História Custodial: PORTUGAL. Arquivo Nacional da Torre do Tombo - Minuta de ofício. 1834-10-3. "Minuta de ofício da Torre do Tombo referindo que ainda não tinha dado entrada um único papel da extinta Junta do Comércio no arquivo". Acessível no Instituto dos Arquivos Nacionais / Torre do Tombo, Lisboa, Portugal. Arquivo do Instituto dos Arquivos Nacionais / Torre do Tombo, cx. 33.

PORTUGAL. Arquivo Nacional da Torre do Tombo - Minuta de ofício. 1835-1-16. "Minuta de ofício da Torre do Tombo sobre a entrada de duas carradas de papéis e livros da extinta Junta do Comércio". Acessível no Instituto dos Arquivos Nacionais / Torre do Tombo, Lisboa, Portugal. Arquivo do Instituto dos Arquivos Nacionais / Torre do Tombo, cx. 33.

PORTUGAL. Arquivo Nacional da Torre do Tombo - Minuta de ofício. 1835-4-6. "Minuta de ofício da Torre do Tombo sobre a entrada do resto dos papéis e livros da extinta Junta do Comércio". Acessível no Instituto dos Arquivos Nacionais / Torre do Tombo, Lisboa, Portugal. Arquivo do Instituto dos Arquivos Nacionais / Torre do Tombo, cx. 33.

Ministério do Reino - Maço 3532. [Manuscrito]. 1843. Acessível no Instituto dos Arquivos Nacionais / Torre do Tombo, Lisboa, Portugal.

- **Regras ou convenções**: DIRECÇÃO GERAL DE ARQUIVOS; PROGRAMA DE NORMALIZAÇÃO DA DESCRIÇÃO EM ARQUIVO; GRUPO DE TRABALHO DE NORMALIZAÇÃO DA DESCRIÇÃO EM ARQUIVO – *Orientações para a descrição arquivística*. 2.ª v. Lisboa: DGARQ, 2007. ISBN 978-972-8107-91-8.
- **Data da descrição**: Elaboração: 2004, Fevereiro; Revisões: 2006, Setembro, 8; 2006, Dezembro, 14.

2.2 ISAAR(CPF)

A ISAAR(CPF), norma internacional de registo de autoridade arquivística para Entidades Coletivas, Pessoas e Famílias, tal como a ISAD, foi elaborada pelo *International Council on Archives*. A segunda versão da norma foi publicada em 2004, após uma revisão que foi feita à primeira edição da norma, que consistiu numa reestruturação e ampliação da primeira edição.

O objetivo da norma ISAAR(CPF) é, segundo a mesma, a partilha de descrições dos produtores de documentos, promovendo a preparação de descrições consistentes, apropriadas e auto-explicativas das pessoas coletivas, das pessoas singulares e das famílias que os produziram. Com este objetivo, será possível uma melhor compreensão do contexto que se encontra na base dos arquivos, garantindo que as várias ocorrências de uma mesma entidade estão ligadas por uma identificação comum. Será, também, permitido a

identificação precisa dos produtores de documentos de arquivo, incorporando aqui a descrição das relações entre as diferentes entidades.

Tendo em conta a natureza da norma ISAAR(CPF), esta deve de ser utilizada em parceria com a norma ISAD(G) e com normas de descrição arquivística nacionais (no caso de Portugal as ODA).

2.2.1 Conceitos da norma

Assim como na ISAD(G), a ISAAR(CPF) também tem na sua génese conceitos que serão apresentados em seguida. Existem, no entanto, conceitos que são transversais a estas duas normas, como acontece com os termos Descrição Arquivística, Documento de arquivo, Pessoa Coletiva, Ponto de acesso, Produtor e Proveniência. Apesar disso, existem dois termos que não existem na ISAD(G) e que necessitam de ser explicitados. Esses dois termos são os seguintes:

- **Qualificativo (Qualifier) -** A informação adicionada a um elemento de descrição para ajudar a identificação, compreensão ou utilização do registo de autoridade. Os qualificativos para as pessoas singulares podem ser: Pré-título, Datas, Título (Nobiliárquico, Honorífico, Académico), Epíteto ou outros.

  Exemplo:

  - **Ponto de acesso normalizado:** Rocha, Adolfo Correia. 1907-1995, médico e escritor
  - **Outras formas do nome**: Miguel Torga (pseudónimo) (Direcção Geral De Arquivo 2007)

- **Registo de autoridade** (*Authority record*) **-** A forma autorizada do nome de uma entidade combinada com outros elementos de informação que identificam e descrevem essa entidade, podendo remeter para outros registos de autoridade relacionados.

Apesar de todos os outros termos estarem presentes no glossário da ISAD(G), existem pequenas diferenças em alguns destes, uma vez que na ISAAR(CPF) são tidos em conta os registos de autoridade. Os restantes conceitos desta norma encontram-se expostos no Anexo 2.

2.2.2 Estrutura da Norma

A ISAAR(CPF), tal como a norma ISAD(G), encontra-se dividida por zonas, sendo que esta apenas tem quatro, com um total de 27 elementos. Na norma os elementos de descrição são apresentados com uma mesma estrutura, semelhante à usada na ISAD(G) (nome do elemento, objetivos, regras e exemplos, quando aplicáveis).

As zonas de descrição existentes são as seguintes:

- **Zona de Identificação** - Zona onde se encontram os metadados que visam a identificação específica da entidade que será descrita, definindo aqui também os pontos de acesso normalizados para o registo. Aqui estão presentes o tipo de entidade, forma(s) autorizada(s) do nome, formas paralelas do nome, formas normalizadas do nome de acordo com outras regras, outras formas do nome e identificadores para pessoas coletivas.

Exemplo:

- **Tipo de entidade**: Pessoa colectiva
- **Formas autorizadas do nome**: Real Junta do Comércio, Agricultura, Fábricas e Navegação destes Reinos e seus Domínios. 1755-1834
- **Formas paralelas do nome**: ----------
- **Formas autorizadas do nome de acordo com outras regras**: ----------
- **Outras formas do nome**: Junta do Comércio destes Reinos e seus Domínios; Junta do Comércio
- **Identificadores para pessoas colectivas**: ----------

- **Zona de descrição** - Nesta zona são registados metadados que se demonstrem pertinentes sobre a natureza, funções, contexto e atividades da entidade que está a ser descrita. Aqui estão presentes as datas de existência, história, lugares, estatuto legal, funções, ocupações e atividades, mandatos/ fontes de autoridade, estruturas internas/ genealogia e contexto geral.

Exemplo:

- **Datas de existência**: 1755-1834
- **História**: Criada pelo Decreto de 30 de Setembro de 1755 a Junta do Comércio destes Reinos e seus Domínios obteve a confirmação dos seus estatutos por Decreto de 16 de Dezembro de 1756. Pela Lei de 5 de Junho de 1788 foi elevada a tribunal supremo passando a designar-se por Real Junta do Comércio, Agricultura, Fábricas e Navegação. A Direcção da Junta era constituída por um provedor, um secretário, um procurador, seis deputados, um juiz conservador (por lhe ter sido concedida jurisdição privativa) e um procurador fiscal. Os deputados eram, obrigatoriamente, homens de negócio acreditados nas praças de Lisboa ou

do Porto. A Junta do Comércio tinha vastas atribuições: fiscalização do comércio de retalho na cidade de Lisboa, definição da política mercantil, tomada de medidas de prevenção, repressão e fiscalização de contrabandos, fiscalização da indústria a nível nacional, supervisão da Mesa do Bem Comum dos Mercadores, poder judicial nas causas de comércio, naturalização de estrangeiros, supervisão da Real Fábrica das Sedas, administração e inspecção dos faróis, tudo o que diz respeito à navegação e à Aula do Comércio. Tinha ainda uma função de carácter consultivo relativamente à agricultura e minas. A Junta do Comércio foi extinta pelo Decreto de 18 de Setembro de 1834.

- **Lugares**: ----------
- **Estatuto legal**: ----------
- **Funções, ocupações e actividades**: ----------
- **Mandatos/Fontes de autoridade**: ----------
- **Estruturas internas/Genealogia**: ----------
- **Contexto geral**: ----------

- **Zona das Relações** - Zona onde se registam metadados relativos à descrição de relações com outras pessoas coletivas, pessoas singulares ou famílias que podem ser descritas noutros registos de autoridade. Estão presentes nesta zona os nomes/ identificadores da pessoa coletiva, pessoa singular ou família relacionadas, categoria do relacionamento, descrição do relacionamento e datas do relacionamento.

Exemplo:

- **Nome/Identificador da pessoa coletiva, da pessoa singular ou da família relacionadas:** Real Fábrica das Sedas e Obras de Águas Livres. 1734-1833 (identificador do registo de autoridade: PT-FNA126).
- **Tipo de relação**: Hierárquica
- **Descrição da relação**: ----------
- **Datas da relação**: ----------

- **Zona de Controlo** - Nesta zona os metadados identificam, de forma unívoca, o registo de autoridade, indicando como, quando e por que serviço o registo de autoridade foi criado e mantido. Estão nesta zona presentes o identificador do registo de autoridade, identificador(es) da instituição, regras ou convenções, estatuto, nível de detalhe, datas de criação, revisão ou eliminação, línguas e escritas, fontes e notas de manutenção.

Exemplo:

- **Identificador do registo de autoridade**: PT-FNA106
- **Identificadores da instituição**: PT-IAN/TT

- **Regras e/ou convenções**: DIRECÇÃO GERAL DE ARQUIVOS; PROGRAMA DE NORMALIZAÇÃO DA DESCRIÇÃO EM ARQUIVO; GRUPO DE TRABALHO DE NORMALIZAÇÃO DA DESCRIÇÃO EM ARQUIVO – Orientações para a descrição arquivística: partes 2 e 3. 2.ª v. Lisboa: DGARQ, 2007. ISBN 978-972-8107-91-8.
- **Estatuto**: ----------
- **Nível de detalhe**: Médio
- **Datas de criação, revisão ou eliminação**: Criado em 08/09/2006
- **Idiomas e escritas**: Português
- **Fontes**: ----------
- **Notas de manutenção**: Descrição elaborada por Joana Braga (IAN/TT).

Após a apresentação das quatro zonas acima referidas, a presente norma inclui, ainda, uma parte denominada "Relações das pessoas coletivas, pessoas singulares e famílias com documentação de arquivo e outros recursos". Nesta parte são apresentadas diretrizes para a ligação dos registos de autoridade arquivística às descrições dos documentos produzidos pela entidade ou a outros recursos de informação sobre ela ou para ela produzidos (International Council on Archives 2004). Esta diretrizes encontram-se estruturadas em quatro pontos, os quais são:

- **"Identificadores e títulos do recurso relacionado** - Identificar, de forma unívoca, o(s) recurso(s) relacionados e/ou estabelecer a ligação entre o registo de autoridade e a descrição dos recursos relacionados quando estes existam.

- **Tipo do recurso relacionado** - Identificar o tipo do(s) recurso(s) relacionado(s) referenciado(s).

- **Natureza da relação** - Identificar a natureza da(s) relação(ões) entre a pessoa colectiva, a pessoa singular ou a família e o(s) recurso(s) relacionado(s).

- **Datas do recurso relacionado** – Fornecer quaisquer datas relevantes para o(s) recurso(s) relacionado(s) e/ou as datas da relação entre a pessoa coletiva, a pessoa singular ou a família e o recurso relacionado, e indicar o significado dessas datas." (International Council on Archives 2004)

Exemplos:

- **1.ª relação**
    - **Identificadores e títulos dos recursos relacionados**: Junta do Comércio (código de referência: PT-TT-JC)
    - **Tipos de recursos relacionados**: Fundo documental
    - **Natureza das relações**: Produtor

- **Datas dos recursos relacionados e/ou das relações**: 1755-1834 (datas de produção da documentação)

- **2.ª relação**
  - **Identificadores e títulos dos recursos relacionados**: Junta do Comércio (PT- Arquivo Histórico do Ministério das Obras Públicas)
  - **Tipos de recursos relacionados**: Fundo documental
  - **Natureza das relações**: Produtor
  - **Datas dos recursos relacionados e/ou das relações**: 1755-1834 (datas de produção da documentação)

2.3 CIDOC-CRM

O CIDOC-CRM (CIDOC Conceptual Reference Model) é uma ontologia formal de referência, destinada a facilitar a integração, mediação e troca de informações no âmbito do património cultural. O CRM foi o culminar de um trabalho desenvolvido, ao longo de mais de uma década, por parte do Comité Internacional para a Documentação (CIDOC) do Conselho Internacional de Museus (ICOM). No ano de 2006 esta foi considerada uma norma ISO, a ISO 21127:2006.

Segundo a norma, o objetivo do CRM consiste no fornecimento das definições semânticas e os esclarecimentos necessários para transformar fontes de informações localizadas e díspares num recurso global coerente, seja onde for (internet, intranets, instituições de grande dimensão). Mais especificamente, o CRM pretende servir de referência em termos de conceitos e, portanto, de semântica a esquemas de bases de dados e estruturas de organização de documentos em sistemas de informação diversos.

2.3.1 O Modelo

O modelo CRM é uma norma internacional desenvolvido por um grupo de trabalho multidisciplinar, com o objetivo de fazer a ponte entre os profissionais da cultura e os informáticos e implementadores de sistemas de informação e de aplicações. Os segundos têm alguma dificuldade em compreender os conceitos culturais em toda a sua abrangência, enquanto os primeiros têm dificuldade em explicar estes mesmos conceitos aos cientistas da computação e os implementadores de sistemas (Santos 2016).

Uma das inovações do CRM é a estruturação dos conceitos em torno dos eventos temporais, em oposição à maioria dos modelos de metadados que têm o recurso como objeto central de interesse (Lima 2008).

A Figura 4 mostra as entidades e relacionamento centrais do modelo. O CIDOC-CRM é baseado em eventos, estando no seu núcleo as **Entidades Temporais** (*E2 Temporal Entity*), isto é, coisas que aconteceram num determinado período de tempo específico. Apenas as Entidades Temporais podem estar ligadas ao tempo e ter **Períodos de Tempo** (*E52 Time-Span*). Por sua vez, os **Objetos** (**Conceptuais** (*E28 Conceptual Object*) e **Físicos** (*E18 Physical Thing*), **Atores** (*E39 Actor*) e **Lugares** (*E53 Place*) devem de estar ligados a um evento - uma Entidade Temporal. Sabendo que um **Lugar** pode ser em qualquer parte, este pode ser definido geograficamente.

Já um **Ator** é uma entidade com responsabilidade legal e pode ser um indivíduo ou um grupo que interage com outras coisas, como sejam Objetos Físicos e Objetos Conceptuais. (Oldman and CRM Labs 2014)

Objetos Físicos são todos os itens físicos persistentes com uma forma relativamente estável, artificial ou natural que ocupa não só um espaço geométrico particular, mas, ao longo de sua existência, também forma uma trajetória através do espaço-tempo, que ocupa num volume real. Esse objeto novo torna-se parte do nosso domínio de interesse.

Por sua vez, os **Objetos Conceptuais** são produtos não materiais que se encontram nas nossas mentes ou outros dados produzidos pelo homem que se tornaram objetos de um discurso sobre a sua identidade, circunstâncias de criação ou implicação histórica. A produção de tais objetos pode ter sido apoiada pela utilização de dispositivos técnicos, tais como câmaras ou computadores. (ICOM/CIDOC CRM Special Interest Group 2017)

É muito comum referir e identificar uma instância específica de algumas classes ou categorias dentro de um determinado contexto, sendo que a este processo se chama **Denominação** (E41 Appellation). O CRM permite-nos nomear qualquer coisa ("*thing*" - termo neutro usado no CRM para as instâncias de uma qualquer classe), podendo as coisas ter vários nomes e esses nomes podem mudar com o tempo como resultado de um evento. Isso significa que o uso e a aplicação de nomes podem ser estudados em períodos de tempo. Uma coisa e o seu nome são consideradas como entidades separadas.

As diversas organizações têm diferentes sistemas de classificação, sendo que em CRM estas classificações são denominadas **Tipos** (E55 Type). Como um tipo também é um Objeto Conceitual (E28), também pode ser discutida a classificação das coisas ao longo do tempo e a história de definição e redefinição dos tipos (Oldman and CRM Labs 2014).

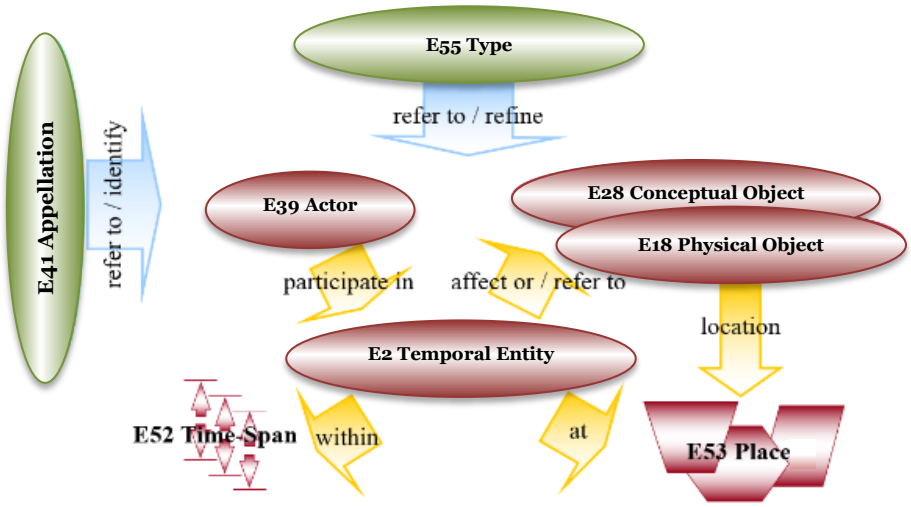


Figura 4 - Principais Entidades do CIDOC-CRM

Fonte: (Oldman and CRM Labs 2014)

2.3.2 Hierarquia de Classes e Propriedades

As hierarquias de classes e propriedade ajudam na compreensão e navegação do modelo CRM, sendo que cada uma destas apresenta uma estrutura própria.

A hierarquia de classes tem, segundo a norma, o seguinte formato:

- Cada linha tem no seu começo um identificador de classe exclusivo, sendo que este é composto por uma letra ("E") seguida de um número. Inicialmente o termo classe era denominado "Entidade", surgindo daí a letra "E" como prefixo;

- Uma série de hífens ("-") segue o identificador de classe único, indicando a posição hierárquica da classe na hierarquia;

- O nome em inglês da classe aparece à direita dos hífens;

- O índice é ordenado por nível hierárquico, de forma "primeiro em profundidade";

- Classes que aparecem em mais de uma posição na hierarquia de classes, como resultado de herança múltipla, são mostradas num tipo de letra em itálico.

Exemplos de Classes (ICOM/CIDOC CRM Special Interest Group 2017):

- **E1** CRM Entity
- **E2** - Temporal Entity
- **E3** - - Condition State
- **E4** - - Period
- **E5** - - - Event

- **E7** - - - - Activity
- **E8** - - - - - Acquisition Event
- **E9** - - - - - Move
- **E10** - - - - - Transfer of Custody
- **E11** - - - - - Modification
- **E12** - - - - - - Production
- **E79** - - - - - - Part Addition

Por sua vez, a hierarquia de propriedades, segundo a norma, tem o seguinte formato:

- Cada linha começa com um identificador de propriedade único, consistindo de um número precedido pela letra "P" (de "Propriedade").

- Uma série de hífens (“-”) segue o identificador de propriedade exclusivo, indicando a posição hierárquica da propriedade na hierarquia.

- O nome em inglês da propriedade aparece à direita dos hífens, seguido pelo nome da propriedade inversa entre parênteses.

- A classe de domínio para a qual a propriedade é declarada.

- O índice é ordenado por nível hierárquico, de forma “primeiro em profundidade” e por número de propriedades entre irmãos iguais.

- Propriedades que aparecem em mais de uma posição na hierarquia de propriedades como resultado de herança múltipla são mostradas num tipo de letra em itálico.

Exemplos de Propriedades (ICOM/CIDOC CRM Special Interest Group 2017):

Property id	**Property name**	**Entity – Domain**	**Entity - Range**
P1	is identified by (identifies)	E1 CRM Entity	E41 Appellation
P48	- has preferred identifier (is preferred identifier of)	E1 CRM Entity	E42 Identifier
P78	- is identified by (identifies)	E52 Time-Span	E49 Time Appellation
P87	- is identified by (identifies)	E53 Place	E44 Place Appellation
P102	- has title (is title of)	E71 Man-Made Thing	E35 Title
P131	- is identified by (identifies)	E39 Actor	E82 Actor Appellation

2.3.3 Modelos CRM

O CIDOC-CRM tem diversos modelos CRM que decorrem de trabalho adicional realizado em áreas para além da dos museus. A maior parte destes modelos são extensões do CIDOC. Estes modelos são apresentados em seguida.

- **CRMarchaeo** - Extensão do CIDOC-CRM criada para apoiar o processo de escavação arqueológica e as entidades e atividades relacionadas com este. Modelo criado a partir de padrões e modelos já em uso por instituições nacionais e internacionais de património cultural. Evoluiu através da análise profunda de metadados existentes em documentação arqueológica real. Aproveita os conceitos fornecidos pelo CRMsci (explicado abaixo), do qual herda a maioria dos princípios geológicos que regem a estratigrafia arqueológica, ampliando esses princípios.[4]

- **CRMba** - Extensão do CIDOC-CRM concebida para apoiar o processo de registo de evidências e mudanças sobre documentação de construções arqueológicas. Modelo baseado nos mesmos princípios do CRM do CIDOC, reutilizando, quando apropriado, partes das classes e propriedades do CIDOC-CRM. Este modelo incorpora, ainda, partes do CRMgeo, CRMsci e CRMarchaeo.[5]

- CRMdig - Ontologia e esquema RDF para codificação de metadados sobre as etapas e métodos de produção de produtos de digitalização e representações digitais sintéticas como 2D, 3D ou até mesmo modelos animados criados por várias tecnologias.[6] É uma extensão do CIDOC-CRM capaz de capturar os requisitos de modelação e consulta em relação à proveniência de objetos digitais para e-science. O CRMdig é particularmente rico para a descrição de circunstâncias físicas da observação científica que resultam em dados digitais (Doerr and Theodoridou 2011).

- CRMgeo - ontologia formal destinada a ser usada como um esquema global para integrar propriedades espaciotemporais de entidades temporais e itens persistentes. O seu objetivo principal consiste no fornecimento de um esquema consistente com o CIDOC-CRM para integrar a geoinformação usando os conceitos, definições formais, padrões de codificação e relações topológicas definidas pelo Open Geospatial Consortium (OGC) no

[4] *Website* CIDOC-CRM – consultado pela última vez a 09/01/2019 – Disponível em: http://www.cidoc-crm.org/crmarchaeo/

[5] *Website* CIDOC-CRM – consultado pela última vez a 09/01/2019 – Disponível em: http://www.cidoc-crm.org/crmba/

[6] *Website* CIDOC-CRM – consultado pela última vez a 09/01/2019 – Disponível em: http://www.cidoc-crm.org/crmdig/

GeoSPARQL. Esta ontologia introduz as classes e relações necessárias para modelar as propriedades espaciotemporais dos fenómenos do mundo real e as suas relações semânticas com as informações espaciotemporais sobre esses fenómenos que foram derivados de fontes históricas, mapas, observações ou medições (Hiebel, Doerr, and Eide 2016).

- CRMinf - Ontologia formal destinada a ser utilizada como esquema global para integrar metadados sobre argumentação e inferência em ciências descritivas e empíricas, como biodiversidade, geologia, geografia, arqueologia, património cultural, conservação, ambientes de TI de pesquisa e bibliotecas de dados de investigação. O seu propósito primordial consiste em facilitar a gestão, integração, mediação, intercâmbio e o acesso a dados acerca do raciocínio por meio de uma descrição das relações semânticas entre premissas, conclusões e atividades de raciocínio (Doerr and Stead 2015).

- CRMsci - Denominado de Modelo de Observação Científica, o CRMsci é uma ontologia formal destinada, como a ontologia anterior, a ser utilizada como um esquema global para integrar metadados sobre observação científica, medições e dados processados em ciências descritivas e empíricas. O seu propósito primordial consiste em facilitar a gestão, integração, mediação, intercâmbio e o acesso a dados de pesquisa pela descrição de relações semânticas, em particular as causais. Não é primariamente um modelo para processar os dados propriamente ditos, a fim de produzir novos resultados de investigação, mesmo que as suas representações se ofereçam para serem usadas para algum tipo de processamento (Doerr et al. 2015).

- CRMtex - É uma extensão do CIDOC-CRM criada para apoiar o estudo de documentos antigos, identificando entidades textuais relevantes e modelando o processo científico relacionado com a investigação de textos antigos e as suas características, a fim de promover a integração com outros campos de investigação do património cultural. Esta extensão do CIDOC-CRM introduz novas classes específicas mais sensíveis às necessidades das disciplinas envolvidas (papirologia, paleografia, codicologia e epigrafia) (Doerr, Felicetti, and Murano 2017). O CRMtex tem como objetivo identificar e definir de forma clara e inequívoca as principais entidades envolvidas no estudo e edição de antigos textos manuscritos e, em seguida, descrevê-los por meio de instrumentos ontológicos apropriados numa perspetiva multidisciplinar. [7]

[7] *Website* CIDOC-CRM – consultado pela última vez a 09/01/2019 – Disponível em: http://www.cidoc-crm.org/crmtex/

2.4 Aplicações do CIDOC-CRM

O CIDOC-CRM foi criado no universo dos museus, no entanto não foi apenas nesta área que já foi aplicado. Tendo isso em conta, é importante verificar o uso deste modelo no seu próprio domínio, assim como em domínios nos quais não foi criado. De seguida serão apresentados casos em que o CIDOC-CRM foi aplicado, primeiramente nos museus e outras áreas e seguidamente nos arquivos, área de estudo deste projeto.

2.4.1 Adoção do CIDOC-CRM em diversas áreas

O **Museo del Prado**, em Madrid, Espanha, é um exemplo de museu que utiliza esta ontologia para a criação de um Grafo de Conhecimento do Museu.

Segundo a página web do museu, o modelo usado tem na sua base uma ampla gama de ontologias de domínios, integrando-as num referencial ontológico comum que representa o conjunto de atividades desenvolvidas no campo museográfico, compreendidas num conjunto de técnicas, práticas e processos relacionados com o funcionamento de um museu. Estes vão desde a documentação associada aos processos de conservação da coleção, até à comunicação com os média, passando pela venda *online* de objetos da loja do Prado. O modelo ontológico é usado não apenas para gerar um conjunto de dados reutilizável, mas também para resolver o conjunto de operações e interrogações que os diferentes grupos de utilizadores podem querer executar na base de conhecimento do museu.

As principais entidades da rede semântica do Prado, Obra de Arte, Autor, Exposição e Atividade são representadas de acordo com o padrão CIDOC-CRM acima mencionado. No entanto, outros tipos de conteúdo que não são contemplados ou controlados pelo CIDOC-CRM podem ser encontrados no *site* do Museu e, por isso, foi necessário hibridizar esse padrão com o modelo FRBR (*Functional Requirements for Bibliographic Records*) e outros vocabulários (ontologias padrão) amplamente utilizados em projetos da web semântica. [8]

Também o **The British Museum** usa o CIDOC-CRM. Segundo a sua página web, esta foi a primeira organização de artes no Reino Unido a publicar a sua coleção semanticamente. Dominic Oldman, IS Development Manager, British Museum afirmou "Esta nova versão semântica fornecerá um novo grau de acessibilidade e permitirá que

[8] Website Museo del Prado - Modelo semántico digital - consultado pela última vez a 04/01/2019 - Disponível em: https://www.museodelprado.es/modelo-semantico-digital/modelo-ontologico

outros trabalhem em estreita colaboração com os dados, obtenham novas ideias e produzam aplicações inovadoras".[9]

Também o **Projeto Ariadne**, ligado ao património arqueológico, tem na sua base a utilização do CIDOC-CRM. As operações de mapeamento conceitual, isto é, as operações entre o esquema de cada base de dados arqueológica e o CIDOC-CRM, ainda estão em desenvolvimento dentro do projeto, e muitos parceiros já definem correspondências complexas entre as entidades contidas nas suas bases de dados e as classes conceptuais fornecidas pelo CIDOC-CRM. A Figura 5 mostra um exemplo retirado do mapeamento de dados de um dos esquemas arqueológicos MiBACT-ICCD (RA schema). [10]

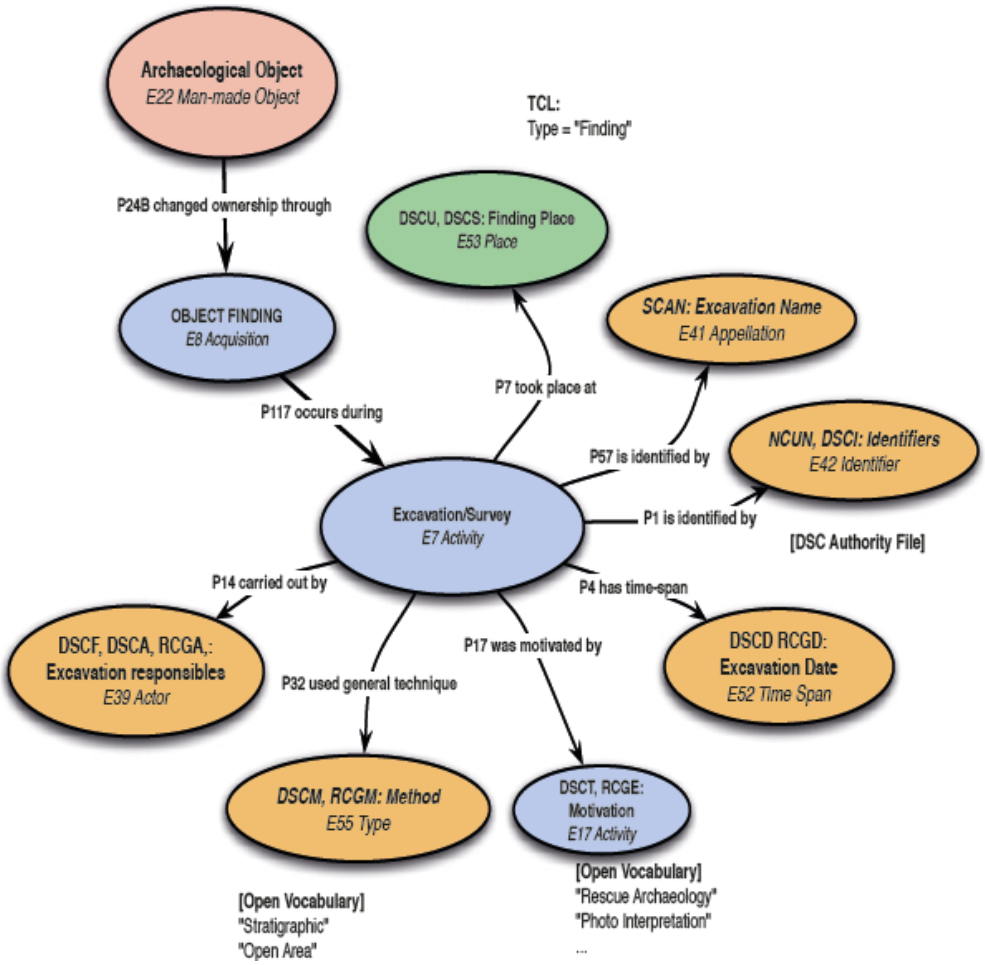


Figura 5 - Exemplo do mapeamento conceitual projeto Ariadne

Fonte: (ARIADNE 2014)

[9] Website The British Museum - consultado pela última vez a 04/01/2019 - Disponível em: https://www.britishmuseum.org/about_us/news_and_press/press_releases/2011/semantic_web_endpoint.aspx

[10] Website Ariadne - About Ariadne - consultado pela última vez a 04/01/2019 - Disponível em: http://www.ariadne-infrastructure.eu/About

2.4.2 CIDOC-CRM nos arquivos

Também nos arquivos já foram feitos alguns estudos de mapeamento para o CIDOC-CRM. Estes estudos tinham como base o padrão internacional de transmissão de metadados para descrições hierárquicas de registos arquivísticos EAD (*Encoded Archival Description*). O EAD é uma linguagem XML (*Extensible Markup Language*) usado por arquivistas em todo o mundo, que possibilita a criação de ferramentas eletrónicas de localização dentro de uma estrutura específica de dados de arquivo, compatível com a ISAD(G) (Society of American Archivists 2018).

Um documento EAD consiste em três elementos: *EAD Header* (eadheader), que é o elemento obrigatório incluindo os metadados para o documento EAD, *Front Matter* (frontmatter), que contém metadados opcionais para a ajuda à procura impressa (se for caso disso) e a *Archival Description* (archdesc), obrigatória, que fornece metadados sobre o conteúdo do arquivo e o contexto de criação. (Bountouri and Gergatsoulis 2010)

A Figura 6 mostra a estrutura do EAD, a qual é baseada numa hierarquia multinível, como acontece na descrição através da ISAD(G). A descrição do arquivo é expressa tendo em conta o *archdesc,* o qual é também o nível de descrição Fundo (level = "fonds"). Os componentes de primeiro nível são expressos através do elemento c01, definindo também o nível de descrição para cada entidade arquivística abaixo de c01, por exemplo os subfundos (nível="subfonds"), as séries (level="series") e o arquivo (level="file"). Níveis mais baixos podem seguir. (Bountouri and Gergatsoulis 2010)

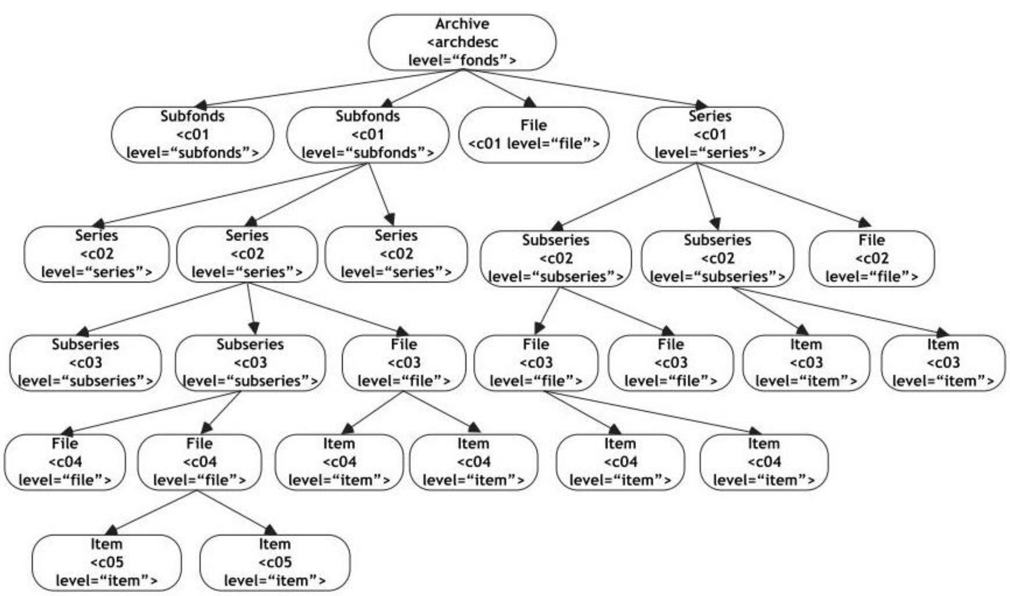


Figura 6 – Hierarquia de arquivo através de EAD

Fonte: (Bountouri and Gergatsoulis 2010)

Foram feitos estudos para a elaboração de um mapeamento de EAD para CIDOC-CRM, sendo que aqui são tidos em conta os três elementos base expressos através da EAD (*EAD Header*, *Front Matter* e *Archival Description*).

Segundo (Bountouri and Gergatsoulis 2010), com base no mapeamento das visões semânticas do arquivo para a ontologia do CIDOC-CRM, o arquivo e os seus componentes são mapeados para três hierarquias distintas de CIDOC-CRM - HPO (*Hierarchy of Physical Objects*), HIO (*Hierarchy of Information Objects*) e HLO (*Hierarchy of Linguistic Objects*). Cada uma destas hierarquias representa uma visão semântica estruturada diferente do arquivo. Além disto, a descrição arquivística é mapeada para outra hierarquia da ontologia, denominada HDEA - *Hierarchy of Documentation Elements and Attributes*).

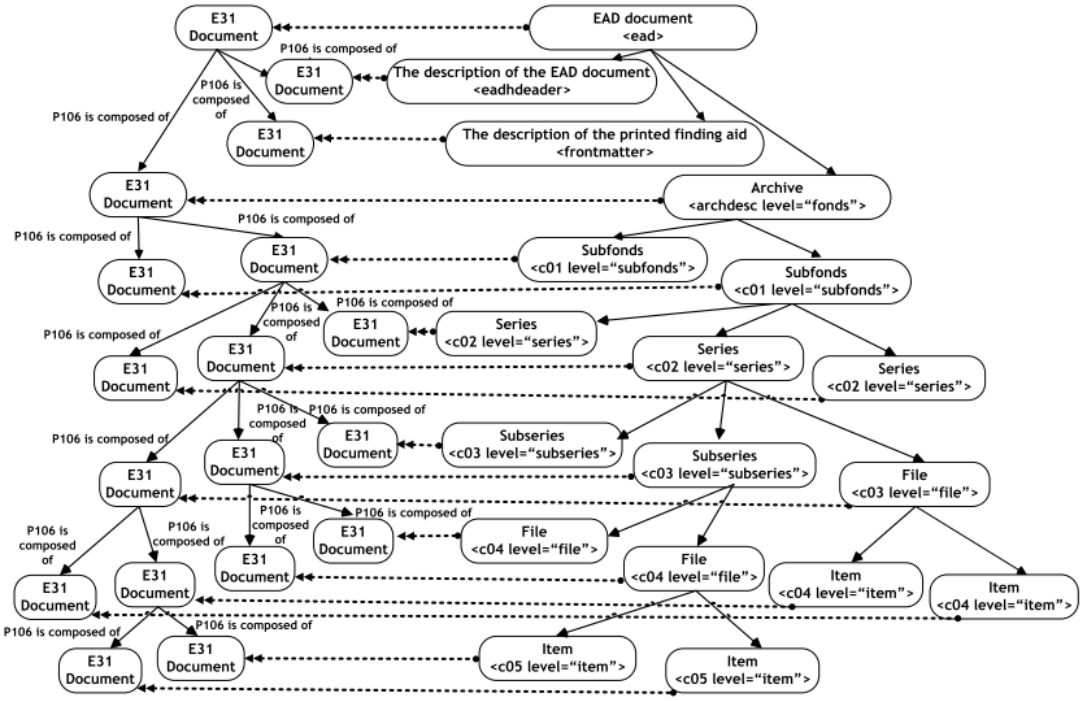


Figura 7 - HDEA - *Hierarchy of Documentation Elements and Attributes*

Fonte: (Bountouri and Gergatsoulis 2010)

Na Figura 7 pode ser observado o mapeamento da hierarquia HDEA de EAD para CIDOC-CRM, onde o conceito de documento é expresso através da classe *E31 Document*, que inclui as instâncias que são objetos imateriais definindo e documentando a realidade, como as frases de um texto, as bases de dados, etc. Todas as instâncias *E31 Document* estão ligadas entre si através da propriedade *P106 is composed of*, de modo a que a hierarquia seja estruturada.

Por sua vez, e tendo em conta o mesmo modelo de hierarquia em árvore, objetos na hierarquia HPO (*Hierarchy of Physical Objects*) são mapeados para a instância *E22 Man-Made Object*, a qual contém todos os objetos que foram criados por atividades humanas. A fim de mapear as relações hierárquicas entre as instâncias, estas são ligadas através da propriedade *P46 is composed of*.

Já a HIO (*Hierarchy of Information Objects*), onde o arquivo é considerado como um objeto que transporta informação, é utilizada a classe *E73 Information Object*, classe que inclui instâncias para os objetos imateriais. A estrutura hierárquica é formada através da utilização da propriedade *P106 is composed of*.

Por último, a HLO (*Hierarchy of Linguistic Objects*) tem como classe a *E33 Linguistic Object*, a qual expressa as classes que contém instâncias de informação que podem ser expressas num ou mais idiomas. A expressão da estrutura hierárquica entre estas instâncias é definida através da propriedade *P106 is composed of*, criando uma hierarquia que mapeia a semântica e a estrutura de árvore obtida pelo mapeamento do arquivo e dos seus componentes como um conjunto de objetos linguísticos para a ontologia. (Bountouri and Gergatsoulis 2010)

As quatro hierarquias acima apresentadas referem-se todas ao mesmo objeto, a descrição de objetos de arquivo, o que faz com que estas tenham uma mesma estrutura, diferindo apenas nos nomes das classes que aparecem nos nós da árvore. Para além disto, estas hierarquias estão relacionadas semanticamente entre si.

2.5 Análise das ferramentas *online* do ANTT

Para que se possa criar um modelo em CIDOC-CRM para o Arquivo Nacional da Torre do Tombo (ANTT) é, desde logo, fundamental compreender como é que este arquivo funciona atualmente e como é que as normas de descrição arquivística são utilizadas. Com isto será possível comprovar como é que as descrições das normas podem vir a ser tornadas em CIDOC-CRM e quais os campos que necessitam de ser refinados para que este novo modelo tenha uma descrição mais flexível, ampla e fina dos objetos culturais. Para que isto aconteça será analisado o *website* e o portal de pesquisa do Arquivo Nacional da Torre do Tombo.

2.5.1 Análise do *website* da Torre do Tombo

A Torre do Tombo tem uma página *web*, que pode ser consultada em http://antt.dglab.gov.pt/, que tem como objetivo mostrar a instituição aos utilizadores, desde a sua história, informações úteis, notícias, serviços de pesquisa presenciais, até sondagens relativas à opinião dos utilizadores.

Figura 8 - Página inicial do *website* da Torre do Tombo

Fonte: *Website* Arquivo Nacional da Torre do Tombo

Disponível em: http://antt.dglab.gov.pt/

Como se pode verificar através da página inicial, há ainda um documento em destaque, o horário de atendimento, como chegar à Torre do Tombo e o que fazer para visitar este Arquivo Nacional.

Neste *website* estão acessíveis diversos tipos de conteúdos, sendo a maioria de carácter informativo e noticioso. Com isto pretende-se dizer que esta página informa o utilizador acerca dos diversos documentos existentes na Torre do Tombo e outros assuntos relacionados com esta instituição, não indo ao pormenor da descrição arquivística dos documentos.

No que diz respeito às pesquisas por documentos existentes na Torre do Tombo, os resultados obtidos cingem-se a notícias relacionadas com os registos dos documentos em questão. Como se pode observar na Figura 9, com o resultado de uma pesquisa sobre o

"*Apocalipse do Lorvão*", os resultados limitam-se a notícias (relativas ao Registo Memória do Mundo UNESCO e aos Mosteiros de Alcobaça e Lorvão), exposições virtuais e a referência aos fundos e coleções (instituições monásticas e conventuais I).

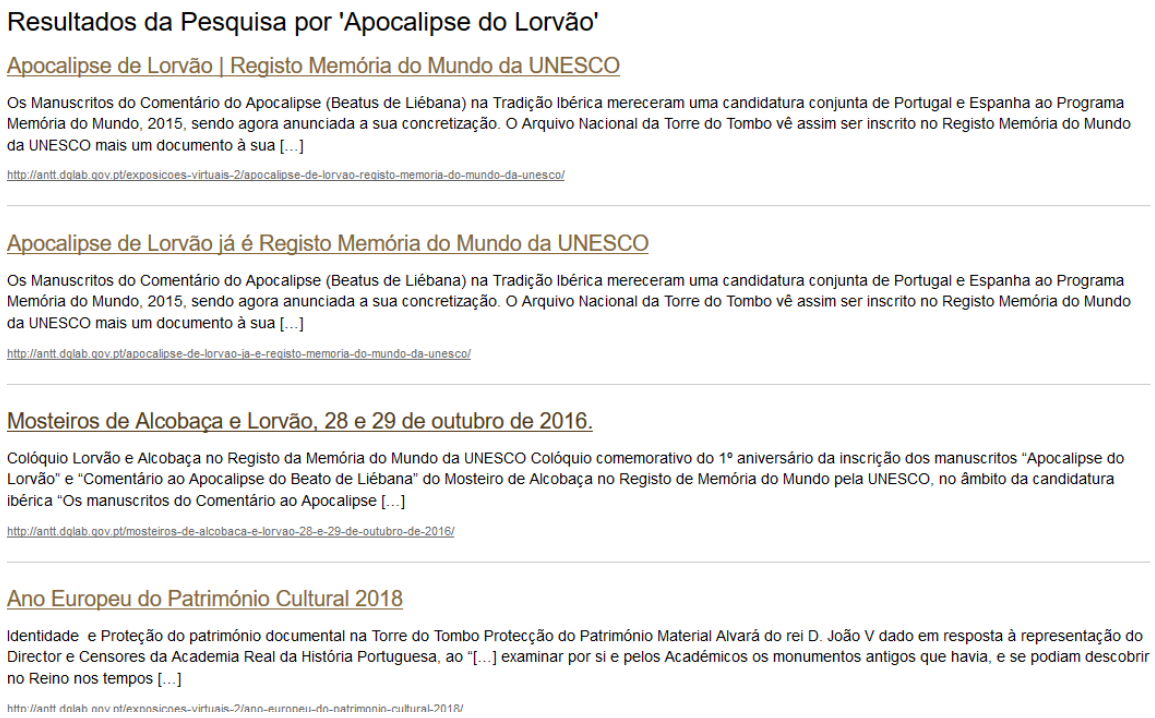


Figura 9 - Resultados da Pesquisa por "Apocalipse do Lorvão"

Fonte: *Website* Arquivo Nacional da Torre do Tombo

Disponível em: http://antt.dglab.gov.pt/?s=Apocalipse+do+Lorv%C3%A3o

2.5.2 Análise do Portal de Pesquisa da Torre do Tombo

No Portal de Pesquisa da Torre do Tombo, *Digitarq*, que pode ser consultado em https://digitarq.arquivos.pt/, é possível elaborar pesquisas através da *web*, semelhantes às que poderiam ser feitas no próprio local. Aqui pode aceder-se, segundo a sua página inicial, a todos os documentos que se encontram digitalizados, podendo fazer-se requisições de certos documentos para consulta e leitura na própria Torre ou solicitar a cópia de documentos em formato digital, entre outras funcionalidades.

Na Figura 10 pode-se observar a página inicial deste portal, onde é permitido ao utilizador fazer as pesquisas por documentos, com o exemplo de uma pesquisa simples, onde se inserem os termos a pesquisar e o período de tempo a que se pretende

restringir os documentos. Aqui são ainda apresentadas as pesquisas mais frequentes, os documentos mais vistos e os documentos recentes.

Figura 10 - Página inicial do Portal de pesquisa da Torre do Tombo

Fonte: Portal de Pesquisa Torre do Tombo

Disponível em: https://digitarq.arquivos.pt/

Para além da pesquisa simples, visível na página inicial, este portal permite ainda a elaboração de uma pesquisa avançada, a qual possibilita o uso de critérios adicionais. Este refinamento pode ser feito através de diversos campos, os quais são apresentados na Figura 11.

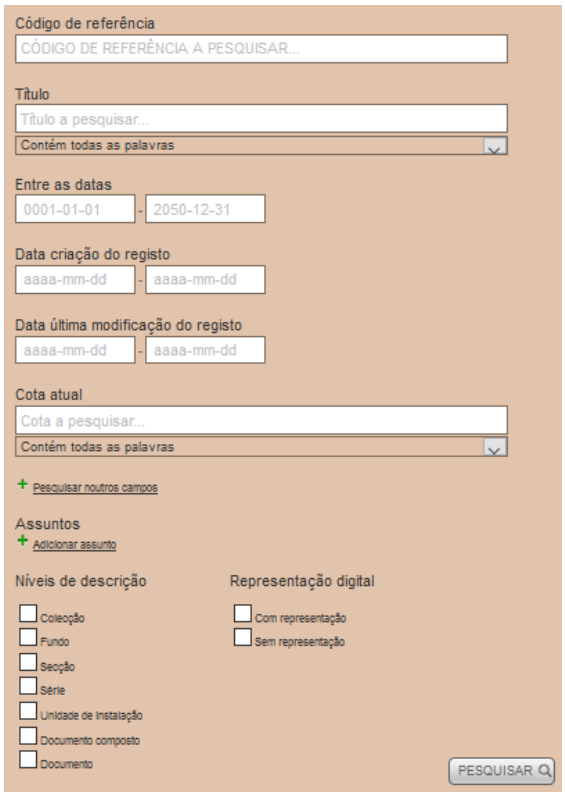


Figura 11 - Pesquisa avançada Portal de pesquisa da Torre do Tombo

Fonte: Portal de Pesquisa Torre do Tombo

Disponível em: https://digitarq.arquivos.pt/asearch

Para além dos dois tipos de pesquisa, esta página web tem também alguns destaques que ajudam na pesquisa de documentos, serviços em-linha, uma sala de referência e leitura virtual, e ajuda, onde é disponibilizado um conjunto de tutoriais para ajudar os utilizadores a interagirem com o Arquivo Nacional da Torre do Tombo.

Após a compreensão de todas as funcionalidades existentes neste portal, foram elaboradas algumas pesquisas para observar como é que este portal responde à pesquisa por um determinado objeto cultural.

Um dos documentos pesquisados foi o "*Apocalipse do Lorvão*", um documento pertencente ao fundo "*Mosteiro de Lorvão*". Aquando da pesquisa simples no portal, este dá um total de quatro resultados, os quais dizem respeito ao fundo e unidade de descrição acima referidos e dois documentos compostos, nos quais está referido o termo "*Lorvão*" em dois campos distinto, nas unidades de descrição relacionadas, da "*Bíblia*", e no âmbito e conteúdo, na "*Correspondência de Fernando Bissaya Barreto*". Este segundo

documento composto não tem a sua informação tratada arquivisticamente, de acordo com o mencionado no resultado da pesquisa.

Figura 12 - Resultados Pesquisa Simples no *Digitarq*

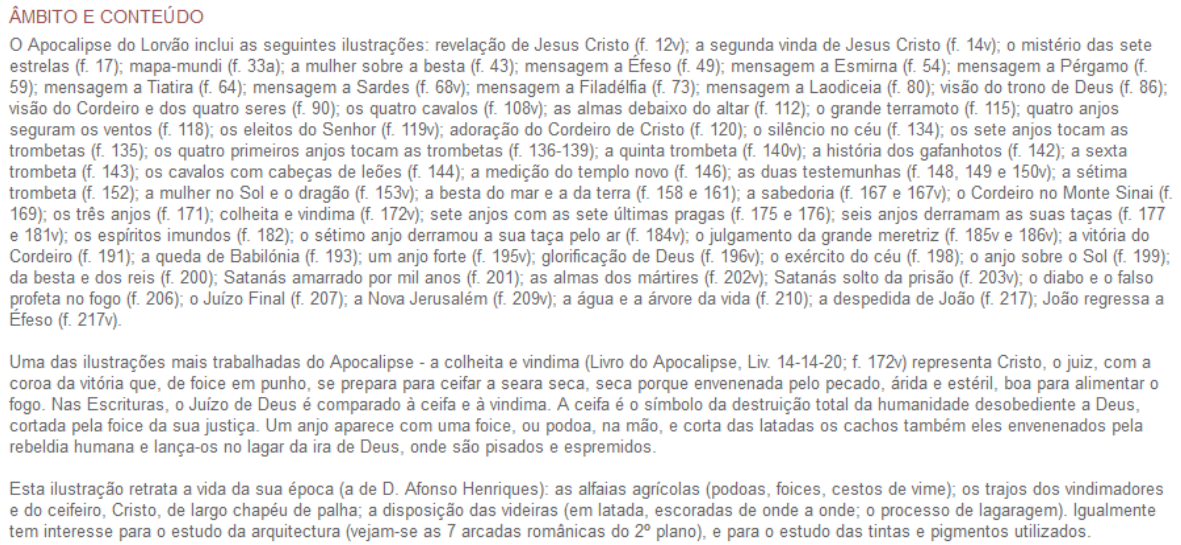
Fonte: Portal de Pesquisa Torre do Tombo

Disponível em: https://digitarq.arquivos.pt/results?t=Apocalipse+do+Lorv%c3%a3o

Depois de elaborada a pesquisa, foi selecionada a unidade "*Apocalipse do Lorvão*", onde foi possível observar, do lado direito, a hierarquia deste registo e, ao centro, os diversos campos de descrição arquivística. Estão ainda visíveis a representação digital dos documentos e os serviços disponíveis em relação a este.

Observando a descrição arquivística desta unidade pode-se constatar que alguns dos campos existentes são bastante densos e longos, o que dificulta a sua leitura, compreensão e extração de dados que possam ser relevantes para a compreensão dos documentos. Isto acontece no caso das Notas de Publicação e no Âmbito e Conteúdo. Na Figura 13 pode-se comprovar que o campo do Âmbito e Conteúdo, o qual está repleto de informação relevante, perde pela maneira como está organizado. Isto acontece uma vez que a informação aqui presente se torna confusa com a constante referência a folhas em que se encontram as ilustrações que vão sendo referidas, entre parênteses. Tendo em conta que existe a representação digital deste documento, uma das opções para melhor compreensão deste texto seria a utilização de hiperligações que permitissem ao leitor observar a

ilustração se assim o quisesse, através do título da mesma, o qual se encontra no campo do âmbito e conteúdo, sem ter de procurá-la em todo o códice.



ÂMBITO E CONTEÚDO

O Apocalipse do Lorvão inclui as seguintes ilustrações: revelação de Jesus Cristo (f. 12v); a segunda vinda de Jesus Cristo (f. 14v); o mistério das sete estrelas (f. 17); mapa-mundi (f. 33a); a mulher sobre a besta (f. 43); mensagem a Éfeso (f. 49); mensagem a Esmirna (f. 54); mensagem a Pérgamo (f. 59); mensagem a Tiatira (f. 64); mensagem a Sardes (f. 68v); mensagem a Filadélfia (f. 73); mensagem a Laodiceia (f. 80); visão do trono de Deus (f. 86); visão do Cordeiro e dos quatro seres (f. 90); os quatro cavalos (f. 108v); as almas debaixo do altar (f. 112); o grande terramoto (f. 115); quatro anjos seguram os ventos (f. 118); os eleitos do Senhor (f. 119v); adoração do Cordeiro de Cristo (f. 120); o silêncio no céu (f. 134); os sete anjos tocam as trombetas (f. 135); os quatro primeiros anjos tocam as trombetas (f. 136-139); a quinta trombeta (f. 140v); a história dos gafanhotos (f. 142); a sexta trombeta (f. 143); os cavalos com cabeças de leões (f. 144); a medição do templo novo (f. 146); as duas testemunhas (f. 148, 149 e 150v); a sétima trombeta (f. 152); a mulher no Sol e o dragão (f. 153v); a besta do mar e a da terra (f. 158 e 161); a sabedoria (f. 167 e 167v); o Cordeiro no Monte Sinai (f. 169); os três anjos (f. 171); colheita e vindima (f. 172v); sete anjos com as sete últimas pragas (f. 175 e 176); seis anjos derramam as suas taças (f. 177 e 181v); os espíritos imundos (f. 182); o sétimo anjo derramou a sua taça pelo ar (f. 184v); o julgamento da grande meretriz (f. 185v e 186v); a vitória do Cordeiro (f. 191); a queda de Babilónia (f. 193); um anjo forte (f. 195v); glorificação de Deus (f. 196v); o exército do céu (f. 198); o anjo sobre o Sol (f. 199); da besta e dos reis (f. 200); Satanás amarrado por mil anos (f. 201); as almas dos mártires (f. 202v); Satanás solto da prisão (f. 203v); o diabo e o falso profeta no fogo (f. 206); o Juízo Final (f. 207); a Nova Jerusalém (f. 209v); a água e a árvore da vida (f. 210); a despedida de João (f. 217); João regressa a Éfeso (f. 217v).

Uma das ilustrações mais trabalhadas do Apocalipse - a colheita e vindima (Livro do Apocalipse, Liv. 14-14-20; f. 172v) representa Cristo, o juiz, com a coroa da vitória que, de foice em punho, se prepara para ceifar a seara seca, seca porque envenenada pelo pecado, árida e estéril, boa para alimentar o fogo. Nas Escrituras, o Juízo de Deus é comparado à ceifa e à vindima. A ceifa é o símbolo da destruição total da humanidade desobediente a Deus, cortada pela foice da sua justiça. Um anjo aparece com uma foice, ou podoa, na mão, e corta das latadas os cachos também eles envenenados pela rebeldia humana e lança-os no lagar da ira de Deus, onde são pisados e espremidos.

Esta ilustração retrata a vida da sua época (a de D. Afonso Henriques): as alfaias agrícolas (podoas, foices, cestos de vime); os trajos dos vindimadores e do ceifeiro, Cristo, de largo chapéu de palha; a disposição das videiras (em latada, escoradas de onde a onde; o processo de lagaragem). Igualmente tem interesse para o estudo da arquitectura (vejam-se as 7 arcadas românicas do 2º plano), e para o estudo das tintas e pigmentos utilizados.

Figura 13 - Âmbito e Conteúdo "Apocalipse do Lorvão"

Fonte: Portal de Pesquisa Torre do Tombo

Disponível em: https://digitarq.arquivos.pt/details?id=4381091

Com este exemplo pudemos observar que, nos campos de descrição que têm uma quantidade de texto elevada, na transformação para o CIDOC-CRM será necessário atender aos conteúdos existentes e analisá-los para extração da informação adicional, a inserir na descrição CIDOC-CRM. Deste modo conseguir-se-á ter acesso a entidades e relações que anteriormente não eram explícitas, por estarem no meio de campos textuais descritivos.

3. Coleção Torre do Tombo

A Torre do Tombo tem em sua posse uma grande quantidade de objetos culturais que foram sendo guardados ao longo da sua existência, o que faz com que a sua coleção seja bastante vasta e rica. Para se saber como e o que mapear, é fundamental conhecer esta coleção de uma maneira aprofundada.

Com o intuito de se saber os documentos que se encontram na posse deste arquivo, é necessário, desde logo, fazer uma análise da sua base de dados. Para que esta análise seja feita, foram analisadas características dos registos e dos seus campos que permitissem fazer uma primeira caracterização das descrições de arquivo, portanto dos registos de metadados que aqui se encontram. Com base nestas características serão feitas perguntas à base de dados, segundo os quais se verá se estas são as características mais adequadas ou se necessitam de ser reformuladas.

Tendo em conta os objetivos delimitados no início do projeto, foram tidas em conta características que pudessem vir a ter mais relevância para a análise dos registos existentes no arquivo. Estas características foram, na sua maioria, selecionadas tendo em conta um conhecimento geral que já se tinha dos registos existentes, pois sem este conhecimento seria difícil fazer uma seleção adequada.

Entre as características selecionadas para fazer uma análise exploratória do conteúdo dos arquivos, encontram-se as seguintes:

- Número de caracteres (de modo representativo) - os metadados existentes têm um número bastante diversificado de caracteres, dependendo do campo de descrição ISAD(G).

- Níveis de descrição - existem vários níveis de descrição, devendo haver uma seleção representativa dos diversos níveis, uma vez que as descrições diferem em completude, quantidade e profundidade em função do nível descritivo em que se posicionem.

- Arquivos de origem - existem vários arquivos a nível nacional, devendo haver um conhecimento da distribuição dos registos por estes.

Com estas características, as quais foram tidas em conta para a análise da base de dados, será possível determinar quais os critérios que vão ser usados para a escolha da amostra a usar no treino de algoritmos para extração de entidades e propriedades, no EPISA.

3.1 Caracterização da coleção

A análise e caracterização seguintes são realizadas sobre a base de dados que contém as descrições, ou registos de metadados, relativos aos documentos pertencentes ao Arquivo Nacional da Torre do Tombo. Esta base de dados tem uma grande quantidade de registos associados, devido à diversidade de documentos que pertencem ao Arquivo Nacional.

Através da análise da base de dados foi possível saber a quantidade de registos existentes nesta. Estes registos encontram-se divididos de acordo com os diversos arquivos do país, uma vez que esta base de dados comporta todos os arquivos a nível nacional.

Os arquivos estão identificados pelos seus códigos de referência, sendo o Arquivo Nacional da Torre do Tombo o que tem mais documentos em sua posse, com um total de 2.207.470 registos. A figura 14 mostra a distribuição dos registos por arquivo e está apresentada em escala logarítmica, dada a diferença de ordem de grandeza no número de registos que cada arquivo possui. Há 54 registos sem código de referência disponível, o que leva a que não se saiba em que arquivo é que se encontram estes registos.

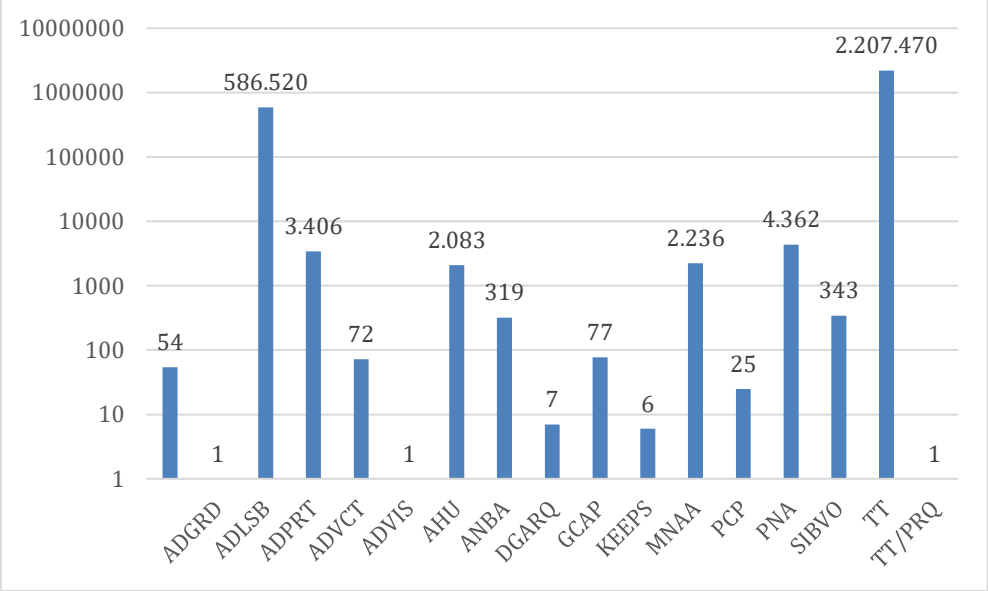


Figura 14 - Número de registos por arquivo

Fonte: Base de Dados *Digitarq*

Os registos que se encontram na base de dados estão divididos em diferentes níveis de descrição. De entre os níveis existentes, o que se destaca, por existir em maior número, é o nível de Documentos Compostos, com um total de 1.743.883, seguido dos Documentos, com 803.588, e Unidades de Instalação, com 223.272. Por sua vez, os que se encontram

em menor número são os sub-sub-fundos, apenas 64, seguidos das Coleções, 107, e das sub-sub-secção, 330. Há ainda um total de 1.813 registos que não têm qualquer nível de descrição registado.

Na Figura 15, que também se encontra em escala logarítmica, podem ser observados todos os níveis de descrição existentes e o número de registos que cada um deles comporta. A coluna de unidade de instalação encontra-se a uma cor diferente dos restantes níveis de descrição, devido ao facto de que este não se encontra na hierarquia de níveis existentes na ISAD(G). No entanto, este é considerado pela TT como um nível de descrição que permite o atalho para ter descrições para grupos reunido por localização. Apesar disto isto não significa que o que está nessas unidades não tenha descrição, uma vez que numa caixa pode estar metade de uma série e metade de outra, sendo que cada série tem a sua descrição e a caixa também.

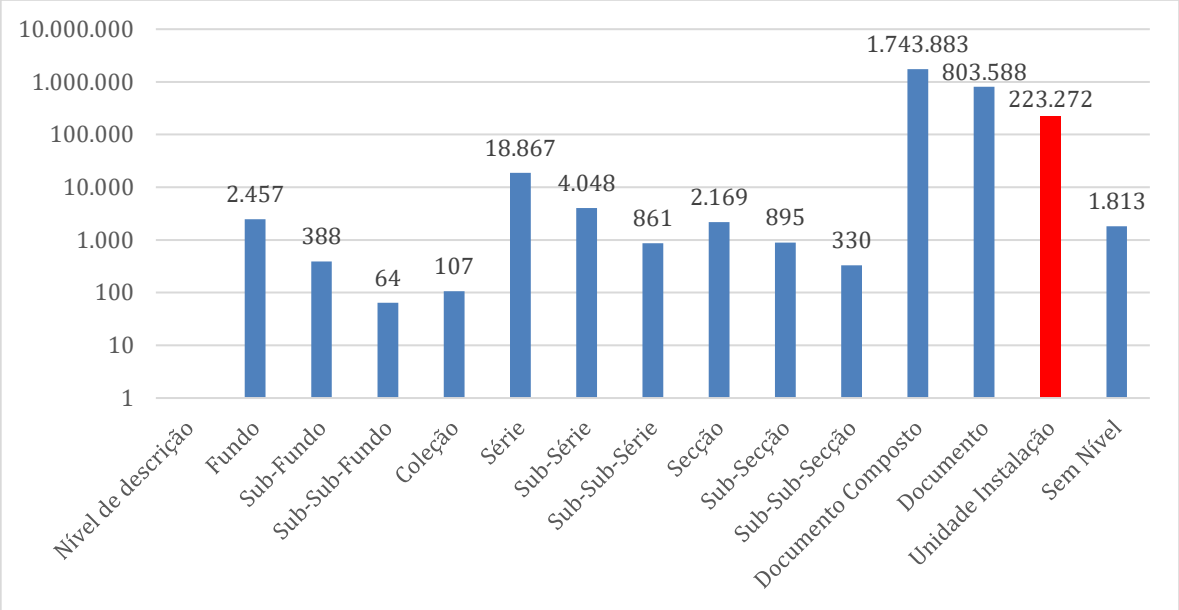


Figura 15 - Distribuição dos níveis de descrição

Fonte: Base de Dados *Digitarq*

Foram ainda analisados os números de caracteres que existem nos campos descritivos, entre os quais se encontram a História Custodial, a História Biográfica e o Âmbito e Conteúdo.

De acordo com a Figura 16 estes três campos têm uma grande quantidade de caracteres associados, sendo que o que tem um maior número máximo de caracteres é o campo do Âmbito e Conteúdo, com um total de 64.223 caracteres, seguido da História Biográfica, com 21.942, e da História Custodial, com 17.832.

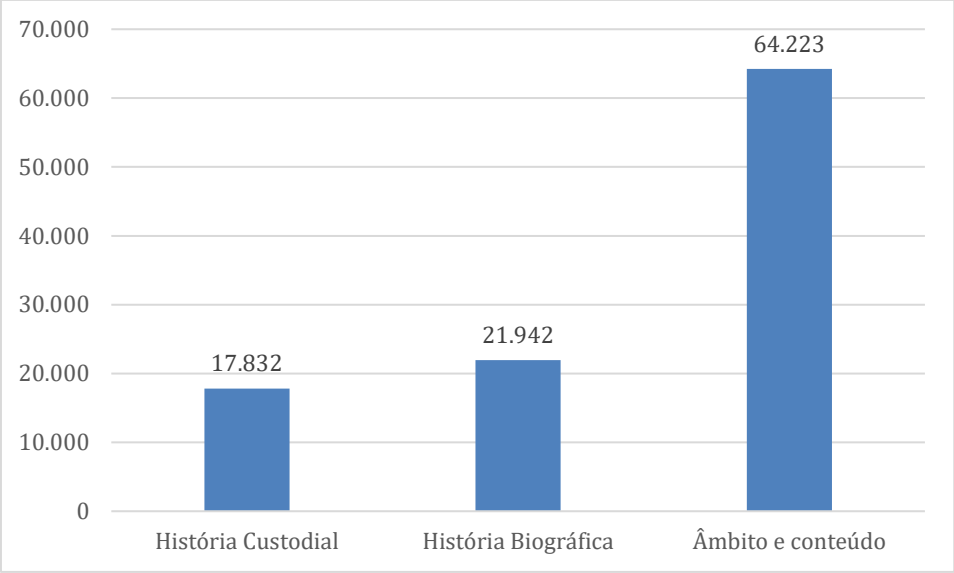


Figura 16 - Número máximo de caracteres em campos descritivos

Fonte: Base de Dados *Digitarq*

Relativamente ao número de caracteres médios destes campos verificou-se que este é baixo. De acordo com o Figura 17, a História Biográfica tem uma média de apenas 3 caracteres. Isto deveu-se ao facto de muitos registos não terem qualquer tipo de informação descrita nestes campos. Esta grande diferença de caracteres levou a que se fizesse uma nova média tendo em conta os registos não vazios.

A Figura 16 mostra os resultados destas duas perguntas e a disparidade entre as médias quando se incluem, ou não, os registos com os caracteres nestes campos. O campo em que isto é mais notório é na História Biográfica, onde a média é de 3 caracteres, quando são tidos em conta todos os registos, e 794 quando só são tidos em conta os registos com a presença de caracteres.

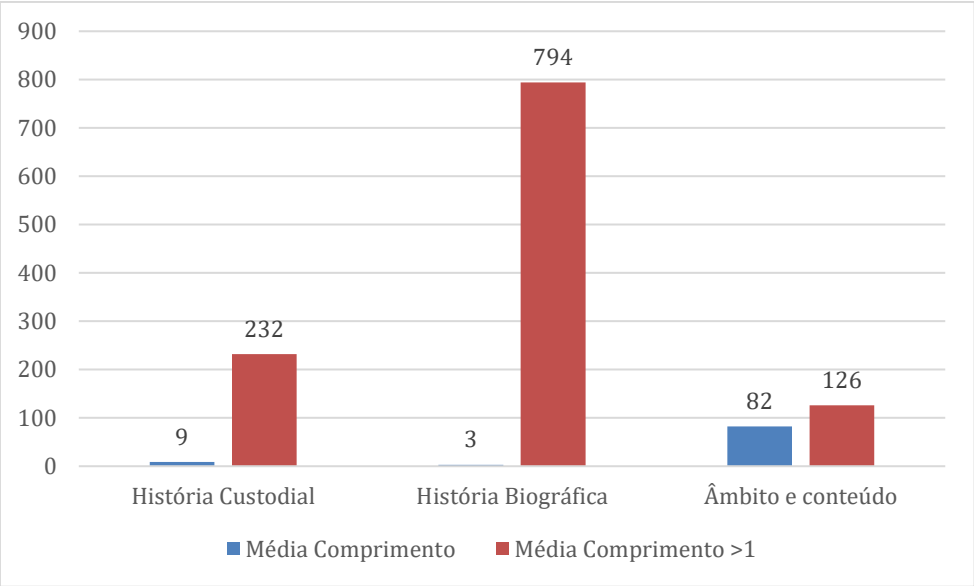


Figura 17 - Número médio de caracteres em campos descritivos

Fonte: Base de Dados *Digitarq*

4. Modelo CIDOC-CRM para o ANTT

Com o intuito de renovar a estrutura de dados utilizada até então, a Torre do Tombo começou a ter em conta modelos de dados que pudessem suportar os campos existentes na descrição dos modelos do ICA (ISAD(G) e ISAAR(CPF)) e que permitissem uma estruturação em grafo de conhecimento. Foram assim tidos em conta dois modelos, o RiC-CM (*Records in Context Conceptual Model*) e o CIDOC-CRM.

O RiC-CM apenas se encontra disponível na sua versão preliminar e destina-se a ser usado estritamente por arquivos. Este modelo caracteriza-se pela transformação de uma descrição multinível, característica da descrição em arquivo, numa descrição multidimensional. Com este novo tipo de descrição o modelo, em vez de ser baseado numa hierarquia, passa a ser representado na forma de grafo ou rede. O modelo multidimensional possibilita a descrição do fundo, mas também da existência do fundo num contexto alargado, em relação a outros fundos.

O RiC-CM é um modelo que permite a descrição de documentos e o ambiente nos quais são criados, acumulados, usados e geridos, de uma maneira que capta e expressa as realidades contextuais complexas de forma mais completa do que através de uma única estrutura hierárquica. (International Council on Archives 2016)

Como este modelo não está disponível na forma de uma ontologia e ainda não se encontra estável ou formalmente estabelecido como norma, o Arquivo Nacional da Torre do Tombo optou pela utilização do CIDOC-CRM, modelo previamente apresentado, criado no âmbito dos museus e que se encontra estável.

Com o modelo de dados a utilizar no presente projeto selecionado, e tendo em conta as normas de arquivo necessárias para o desenvolvimento do novo modelo de dados para o Arquivo Nacional da Torre do Tombo, passou-se ao estudo de mapeamento entre as normas de arquivo (ISAD(G) e ISAAR(CPF)) e o CIDOC-CRM.

Ao longo do presente capítulo serão tidos em conta os procedimentos utilizados para a criação do modelo CIDOC-CRM para o Arquivo Nacional da Torre do Tombo. Será aqui explicitada a ontologia criada para representar os registos existentes na Torre do Tombo, sendo utilizados exemplos de documentos descritos com a norma ISAD(G), presentes no *Digitarq*, portal do ANTT.

4.1 ISAD(G) para CIDOC-CRM

Para descrever os documentos presentes na sua coleção, o Arquivo Nacional da Torre do Tombo tem vindo, ao longo da sua existência, a elaborar descrições arquivísticas tendo em conta as normas ISAD(G) e ISAAR(CPF). Com vista num novo modelo de dados em grafo, e a conformidade ao CIDOC-CRM, é fundamental compreender quais as classes e propriedades deste modelo que podem vir a representar os campos existentes nas descrições arquivísticas utilizadas atualmente.

Para que o mapeamento entre as normas aplicadas neste projeto seja possível é necessário, primeiramente, saber quais os metadados que são efetivamente utilizados nas descrições arquivísticas realizadas pela Torre do Tombo e como é que estes são apresentados. Para este efeito foram selecionados e analisados alguns documentos fornecidos pela Torre do Tombo. Entre estes documentos encontram-se catorze do *Digitarq* e quatro das Orientações para a Descrição Arquivística (ODA). Os documentos selecionados têm diferentes níveis de descrição, de modo a que se possa fazer uma análise abrangente.

Na Tabela 1 pode-se observar um exemplo de registo de metadados analisado, denominado "*Mosteiro do Salvador de Grijó*" [11]. Este registo faz parte de um conjunto de registos que foram analisados no *Digitarq*. Na primeira coluna da tabela estão os campos da ISAD(G) que foram utilizados ao longo de todos os registos selecionados no *Digitaq*. Por sua vez, na segunda coluna da tabela estão os metadados que descrevem o documento acima referido. Notar que nem todos os campos existentes estão presentes na descrição deste documento. Isto acontece uma vez que os campos das ISAD(G) são relativos ao conjunto de todos os registos analisados. Quando um determinado campo não é utilizado na descrição em causa, esse campo aparece como "N/A" neste exemplo. Já a terceira coluna da tabela corresponde ao número de registos em que o campo em questão foi preenchido. Com isto pretende-se ilustrar quais são os campos mais utilizados ao longo deste conjunto de descrições.

De entre os campos apresentados neste conjunto de registos, os que representam a Zona da Identificação na ISAD(G) são os que se mostraram sistematicamente preenchidos. Isto deve-se ao facto de esta ser a zona de preenchimento obrigatório na descrição de objetos de arquivo.

[11] *Website Digitarq* - consultado pela última vez a 06/06/2019 - Disponível em: https://digitarq.arquivos.pt/details?id=4380804

Apesar de esta não ser uma amostra representativa dos registos existentes no *Digitarq*, a análise destes permitiu a compreensão do tipo de metadados que se encontram em cada um dos campos ISAD(G) e como é que estes são apresentados.

Tabela 1 - Exemplo de registo do *Digitarq*

Campos ISAD(G)	Metadados	N.º Ocorrências
Nível de descrição	Fundo	14
Código de referência	PT/TT/MSGR	14
Título/ Nome	Mosteiro do Salvador de Grijó	14
Tipo de título	Atribuído	14
Datas de produção	1302 a 1833	12
Datas predominantes	N/A	1
Datas descritivas	N/A	3
Dimensão e suporte	51 liv., 8 mç. perg., papel	14
Extensões	N/A	2
História administrativa/ biográfica/ familiar	O Mosteiro do Salvador de Grijó era masculino, situava-se na antiga Terra e comarca da Feira. Aderiu à Ordem de Santo Agostinho. Esteve sujeito à jurisdição ordinária do Porto. Aderiu à reforma do Mosteiro Santa Cruz de Coimbra e foi unido à Congregação do mesmo nome. (CONTINUA)	6
História custodial e arquivística	Em 1833, o inventário do extinto Mosteiro refere três cartórios: o cartório (cujos documentos transitaram depois, na sua maioria, para o Arquivo da Torre do Tombo e para o Arquivo Distrital do Porto), o cartório eclesiástico com documentos da freguesia de Grijó (com livros de visitação ao Mosteiro, registos de testamentos, audiências, e despesas eclesiásticas, registos de termos de culpados, de ordens, de certidões de baptismo, registos de baptismos, do crisma, de casamentos, de óbitos) (CONTINUA)	9
Fonte imediata	Em 1972, a 23 de Maio, a Crónica em duas partes, de	5

de aquisição ou transferência	1634, foi comprada ao Dr. João Martins da Silva Marques, director da Torre do Tombo. Em 1974, a 2 de Maio, foi comprado, no Porto, um manuscrito no leilão Soares e Mendonça, o "Index de todos os breves, doações e mais papéis que estão em todos os armários e sacos do cartório do mosteiro de Grijó, feito no ano de 1622".	
Âmbito e conteúdo	Contém cartas régias (incluindo cartas de confirmação), documentos pontifícios, contratos, prazos, cartas de escambo, de quitação, posses, renúncias, tombos (contêm cópias autênticas do século XVIII), o tombo das rendas e direitos dos cónegos, jurisdição eclesiástica e privilégios dos pontífices (contém cópias de documentos do séculos XII e seguintes) jurisdição secular e privilégios dos reis e dos príncipes (contém cópias de documentos dos séculos XII e seguintes) (CONTINUA)	14
Avaliação e seleção	N/A	1
Ingressos adicionais	N/A	2
Sistema de organização	Ordenação numérica específica para cada tipo de unidade de instalação (livros e maços).	5
Condições de acesso	Contém documentos sujeitos a autorização para consulta e a horário restrito.	6
Condições de reprodução	N/A	3
Cota atual	N/A	4
Cota antiga	N/A	1
Caraterísticas físicas e requisitos técnicos	N/A	1
Idioma e escrita	Latim e Português	7
Instrumentos de pesquisa	ARQUIVO NACIONAL DA TORRE DO TOMBO - [Base de dados de descrição arquivística]. [Em linha]. Lisboa: ANTT, 2000- . Disponível no Sítio Web e na Sala de Referência da Torre do Tombo. Em actualização permanente. Índice (inventário) dos livros de diversos conventos,	7

	ordens militares e outras corporações religiosas guardados no Arquivo da Torre do Tombo, conventos diversos, caderneta 3 (Santo Elói a Teatinos) (C 270), f. 51. (CONTINUA)	
Unidades de descrição relacionadas	Portugal, Arquivo Distrital de Braga. Portugal, Arquivo Distrital do Porto, Convento de São Salvador de Grijó - Vila Nova de Gaia. Portugal, Arquivo Municipal de Vila Nova de Gaia. Portugal, Biblioteca Nacional. (CONTINUA)	5
Notas	N/A	3
Notas de publicação	"Documentos Medievais Portugueses". Lisboa: Academia Portuguesa de História, 1958- . 2 vol.; 38 cm. V. 1, t. 1: "Documentos Régios: documentos dos Condes Portucalenses e de D. Afonso Henriques A.D. 1095-1185. 1962. "Ordens religiosas em Portugal: das origens a Trento: guia histórico". Dir. Bernardo de Vasconcelos e Sousa. Lisboa: Livros Horizonte, 2005. ISBN 972-24-1433-X. p. 182-183.	4
Existência e localização de cópias	N/A	4
Data de criação	07/04/2011 00:00:00	14
Última modificação	02/02/2017 10:33:34	14

Fonte: *Website Digitarq* - consultado pela última vez a 06/06/2019 - Disponível em: https://digitarq.arquivos.pt/details?id=4380804

Ao analisar os diversos registos, foi possível comprovar que os registos que se encontram no *Digitarq* apenas contêm os campos da ISAD(G), enquanto os registos da ODA contêm os campos das ISAD(G) e ISAAR(CPF). Isto deve-se ao facto de os documentos das ODA serem exemplos específicos do uso de registos de autoridade, enquanto no *Digitarq* se faz apenas a descrição dos documentos, sem referir de forma normalizada as entidades que estão ligadas ao registo, não aplicando a ISAAR(CPF).

Tendo em conta que o que se pretende mapear consiste nos registos que se encontram na base de dados atual da Torre do Tombo, os registos que se tiveram em conta para a elaboração deste mapeamento são os que se encontram no *Digitarq*. Isto aconteceu devido ao facto de a ISAD(G) estar mais relacionada com as propriedades, um dos principais focos deste trabalho, como será demonstrado em seguida.

Foram, assim, considerados os campos mais utilizados no *Digitarq*, para se começar a elaborar uma tabela com as possíveis correspondências entre as normas ISAD(G) e CIDOC-CRM. Na tabela elaborada é possível compreender que há classes que podem ser diretamente relacionadas com a norma ISAD(G), enquanto outras não, visto que não têm uma relação direta com o pretendido.

Classes do CIDOC-CRM mostraram-se diretamente relacionadas com os campos da ISAD(G), uma vez que no CIDOC-CRM foi possível encontrar classes que pretendem representar conceitos que também estão presentes na descrição arquivística e que têm a mesma finalidade.

A Tabela 2 tem exemplos de campos da ISAD(G) que estão diretamente relacionados com as classes CIDOC-CRM. Entre estes campos encontram-se alguns dos campos mais utilizados no *Digitarq*, os quais são utilizados com o mesmo propósito na identificação de objetos de museu e de arquivo. Para se ver a totalidade dos campos, consultar o Anexo 3.

Tabela 2 - ISAD(G) para CIDOC-CRM

ISAD(G)	CIDOC-CRM
Código de referência	E42 Identifier
Título	E35 Title
Data (s)	E52 Time-Span
Nível de Descrição	E55 Type
Dimensão e Suporte	E54 Dimension / E57 Material
Cota Atual	E42 Identifier
Idioma e Escrita	E56 Language
Data de criação	E52 Time-Span
Última Modificação	E52 Time-Span

Fonte: Autoria Própria

Por outro lado, houve classes do modelo CIDOC-CRM que não se mostraram apropriadas para representar os registos ISAD(G). Isto deveu-se ao facto de nenhuma

classe CIDOC-CRM conseguir exprimir o que a norma ISAD(G) define, o que faz com que se perca o significado das descrições. Houve apenas classes que permitiam colocar alguns campos das descrições arquivísticas todas dentro de uma mesma classe, o que faz com que não haja diferenciação entre os diversos campos.

A Tabela 3 inclui exemplos de campos ISAD(G) que podem estar relacionados com as classes do CIDOC-CRM, no entanto, ao contrário do que acontece na tabela anterior, estes campos do CIDOC-CRM não se mostram apropriados para representar o que a ISAD(G) pretende. Isto deve-se ao facto de estes campos serem específicos dos arquivos e não estarem presentes na descrição de objetos de museu. Isto é indicativo de que será necessário ter em conta extensões que permitam a criação de significado para os objetos culturais dos arquivos. As extensões ao modelo CIDOC-CRM permitem manter a essência do modelo e ampliar o seu significado para a área dos arquivos, objetivo principal deste trabalho.

Tabela 3 - ISAD(G) para CIDOC-CRM

ISAD(G)	CIDOC-CRM
História administrativa/biográfica/familiar	E62 String
História custodial e arquivística	E62 String
Fonte imediata de aquisição ou transferência	E62 String
Âmbito e conteúdo	E62 String
Condições de acesso	E62 String
Instrumentos de pesquisa	E62 String
Notas de publicação	E62 String

Fonte: Autoria Própria

Com um maior conhecimento do CIDOC-CRM, e com as classes a utilizar para um primeiro mapeamento já selecionadas, começou-se por elaborar uma representação em CIDOC-CRM de um registo existente no *Digitarq*. Este registo, proposto pela equipa do Arquivo Nacional da Torre do Tombo, foi o Apocalipse do Lorvão. Este documento, "considerado pela UNESCO como um dos mais belos documentos da civilização medieval ocidental" [12], tem um dos registos mais completos da coleção da Torre do Tombo.

[12] *Website* RTP Ensina - consultado pela última vez a 24/04/2019 - Disponível em: http://ensina.rtp.pt/artigo/apocalipse-do-lorvao-raridade-do-seculo-xii/

A Figura 18 é uma representação deste registo com recurso a classes e propriedades do CIDOC-CRM em que cada cor corresponde a uma zona da ISAD(G), conforme apresentado na legenda à direita do esquema.

Ao longo deste mapeamento foram utilizadas as classes e propriedades que haviam sido estudadas anteriormente e que estão presentes nas tabelas anteriores. Apesar de este ser um primeiro mapeamento, foi possível comprovar como é que as classes e propriedades se podem relacionar e perceber o aspeto geral do grafo.

Para além deste mapeamento foram feitos outros, relativos a outros registos presentes no *Digitarq*, de modo a compreender se este mapeamento seria possível para os diversos registos e níveis de descrição presentes na coleção completa.

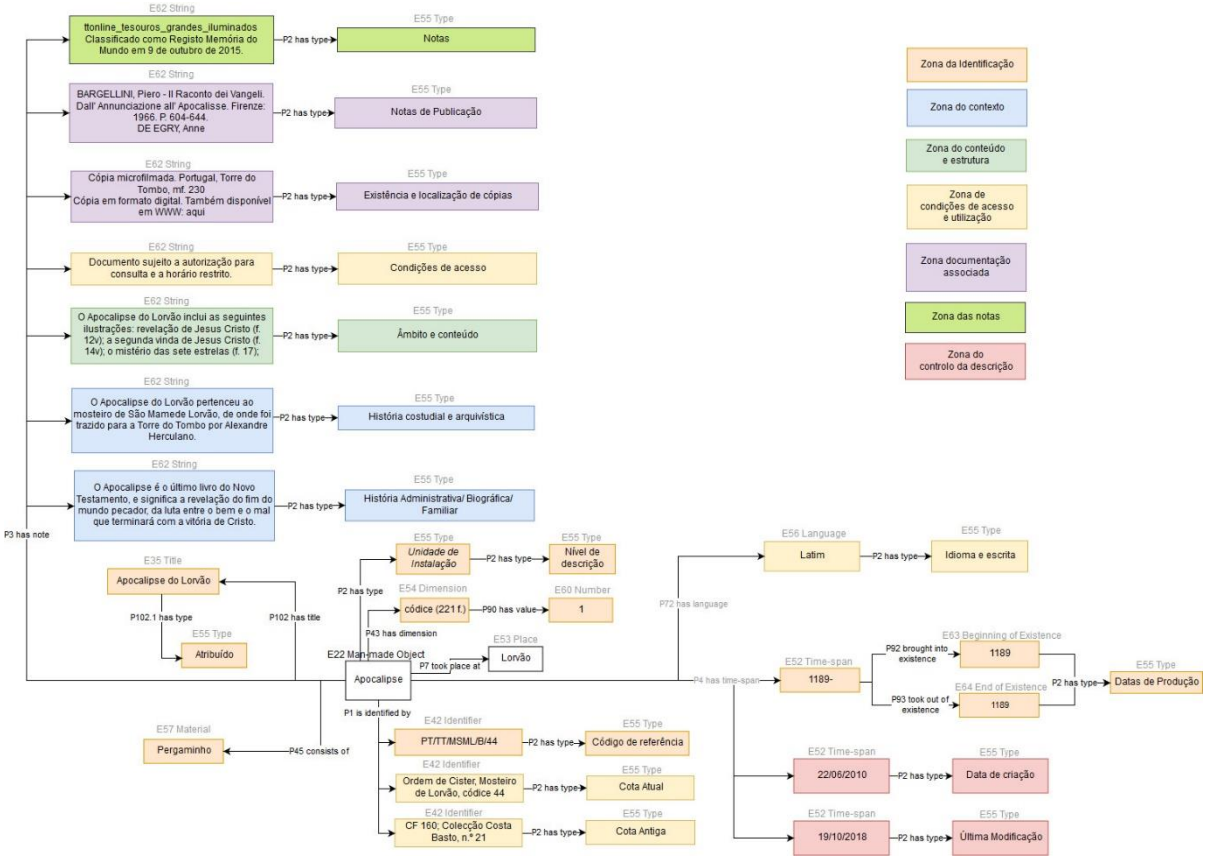


Figura 18 – Primeiro mapeamento CIDOC-CRM

Fonte: Autoria própria

Com um mapeamento já elaborado, e discutido com os especialistas, procedeu-se à criação da ontologia para os arquivos. Esta ontologia inclui classes e propriedades do

CIDOC-CRM e será testada com indivíduos correspondentes aos registos existentes para testar o mapeamento da descrição arquivística para o CIDOC-CRM, sem perder o significado de cada uma das zonas de descrição existentes na ISAD(G).

Ao longo do desenvolvimento do presente modelo as classes e as propriedades existentes neste primeiro mapeamento foram sendo adequadas e mapeadas, de modo a que a ontologia e o grafo de conhecimento pudessem estar de acordo, sendo que algumas escolhas da ontologia foram influenciadas por este alinhamento. Tudo isto foi sendo executado de maneira a nunca descurar o significado do que se pretende representar.

4.2 Ontologia

As ontologias desempenham um papel fulcral nos cenários de interoperabilidade e integração semântica, uma vez que podem representar um domínio e expressar a semântica de maneira formal. Como resultado, as ontologias são preferidas em relação a outros esquemas, devido à sua capacidade de descrever domínios particulares de interesse e expressar a sua riqueza semântica. (Bountouri and Gergatsoulis 2011).

Tendo em conta a análise anteriormente feita do CIDOC-CRM e das normas de arquivo, começou-se por analisar a ontologia CIDOC-CRM. Foi tida em conta a versão 6.2 da ontologia, uma vez que é esta que se encontra disponível no *website* oficial do CIDOC-CRM[13].

Ao analisar esta ontologia, foi possível concluir que aqui está representada a sua hierarquia de classes e propriedades, podendo estas ser *Data Properties* ou *Object Properties*.

Como nesta ontologia estão representadas todas as classes e propriedades existentes no CIDOC-CRM optou-se, no começo, por se fazer uma ontologia apenas com as classes e propriedades estritamente necessárias para que a ontologia para os arquivos fizesse sentido e representasse este domínio. Com isto, foi mais fácil compreender como é que o CIDOC-CRM se encontra estruturado.

Para se criar esta ontologia foram-se criando as classes fundamentais para o campo dos arquivos, havendo, por isso, a necessidade de criar as suas superclasses, de modo a manter a integridade da ontologia original.

[13] *Website* CIDOC-CRM - consultado pela última vez a 03/05/2019 - Disponível em: http://new.cidoc-crm.org/Version/version-6.2

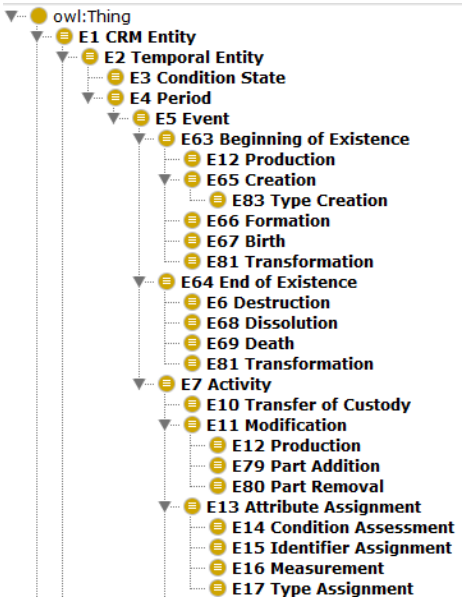


Figura 19 – Classes CIDOC-CRM V6.2

Fonte: Ontologia CIDOC-CRM V6.2

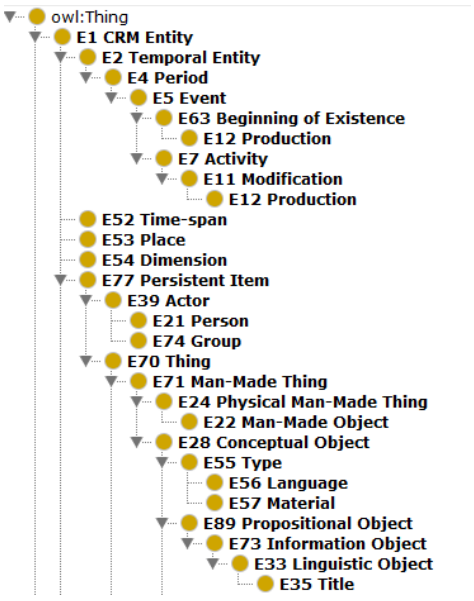


Figura 20 – Classes CIDOC-CRM necessárias

Fonte: Autoria Própria

Nas figuras 19 e 20 podem-se observar as classes do CIDOC-CRM presentes na ontologia CIDOC-CRM v6.2 e da que foi criada para os arquivos. Na Figura 19, estão representadas as classes presentes na hierarquia de classes CIDOC-CRM, na sua versão original, enquanto na Figura 20 estão presentes as classes que foram consideradas necessárias para que a ontologia mantivesse o seu significado e representasse a descrição arquivística.

Tendo em conta que existe um grande número de classes nestas ontologias, estão aqui representadas amostras das classes existentes; notar que existem mais classes no modelo original CIDOC-CRM do que no modelo que está a ser criado para os arquivos. No modelo CIDOC-CRM original existe um total de 169 classes, existindo apenas 40 classes no modelo para os arquivos, sendo que as classes são um subconjunto do CIDOC-CRM com a criação de classes extra.

Quanto às propriedades existentes no modelo CIDOC-CRM, estas são, na sua maioria, do tipo *Object Properties*, havendo poucas *Data Properties*, como se pode observar nas Figuras 21 e 22 que representam parte das *Object Properties,* que são em grande número, 277 propriedades no total, e a totalidade das *Data Properties,* apenas 8 propriedades.

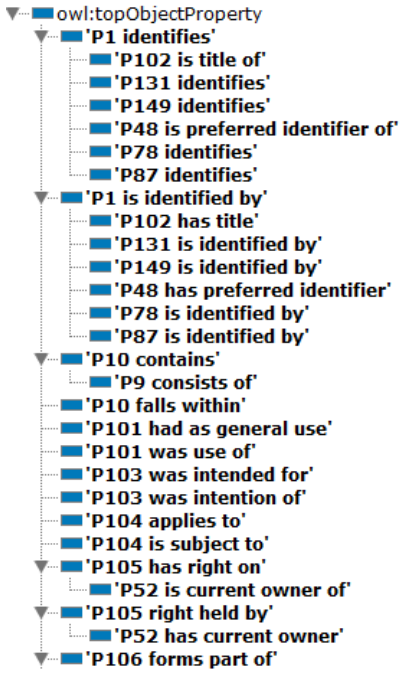


Figura 21 – *Object Properties* CIDOC-CRM V6.2

Fonte: Ontologia CIDOC-CRM V6.2

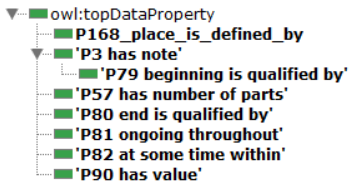


Figura 22 – *Data Properties* CIDOC-CRM V6.2

Fonte: Ontologia CIDOC-CRM V6.2

Após a análise das propriedades deste modelo, foi possível concluir que, dado o domínio dos arquivos, as propriedades existentes se demonstraram insuficientes para a representação dos diversos campos das normas de descrição arquivísticas. Como tal, foi fundamental ter em conta a criação de novas propriedades que permitissem a descrição dos campos em falta. A criação das propriedades passou, principalmente pelo tipo *Data*, uma vez que no CIDOC-CRM quase não existe este tipo de propriedades. Dada a essência dos campos existentes na ISAD(G), este é o tipo de propriedade que se verificou essencial conceber.

Na Figura 23 podem ser observadas as *Data Properties* que se criaram inicialmente, entre as quais se encontram as propriedades relativas aos campos das ISAD(G) que têm uma grande quantidade de caracteres, como o caso das histórias, âmbito e conteúdo e notas de publicação. Foram ainda criadas propriedades para os diferentes tipos de datas, de modo a se distinguir as mesmas.

Tendo em conta que alguns títulos de documentos existentes no Arquivo Nacional da Torre do Tombo têm uma grande dimensão, optou-se por colocar a propriedade *P102 has*

title como *Data Properties*, ao contrário do que acontece na ontologia do CIDOC-CRM, onde esta é considerada uma *Object Property*.

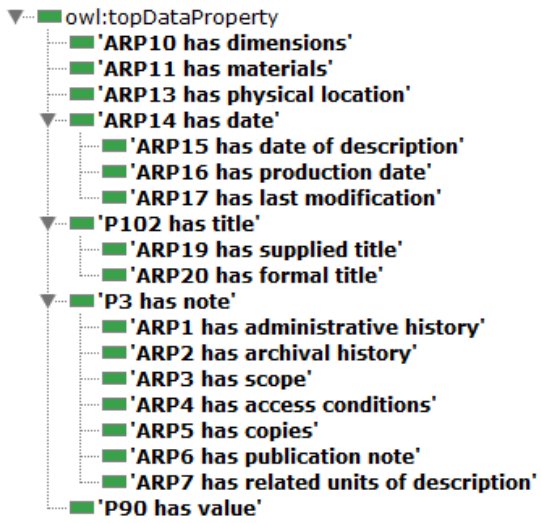


Figura 23 – *Data Properties* para os arquivos

Fonte: Autoria Própria

Nesta figura é ainda possível observar que as novas propriedades criadas têm uma designação com prefixo diferente das propriedades que já vêm do CIDOC-CRM. Isto deve-se ao facto de se querer diferenciar o que já vem do modelo original do que foi criado para o modelo dos arquivos. As propriedades que vêm do CIDOC-CRM são antecedidas por um prefixo "P", como no caso de *P3 has note*, enquanto as propriedades criadas para os arquivos são antecedidas por um prefixo "ARP", como em *ARP1 has administrative history*. Este foi o prefixo escolhido, uma vez que a extensão se destina a arquivos, passando estas propriedades a ser *Archival Properties*. Verificou-se que este prefixo não colide com outras extensões já propostas para o CIDOC-CRM (Doerr et al. 2016).

Para além de se criar *Data Properties*, foi também necessária a criação de uma classe e de *Object Properties*. A classe criada, *ARE1 Level of description*, tem como objetivo identificar os diversos níveis de descrição existentes na ISAD(G), conceito que não existe no âmbito dos museus. Assim, como nas propriedades anteriormente criadas, esta classe tem um prefixo para a distinguir das classes existentes no CIDOC-CRM, neste caso o prefixo é "*ARE*", referente a *Archival Entity*.

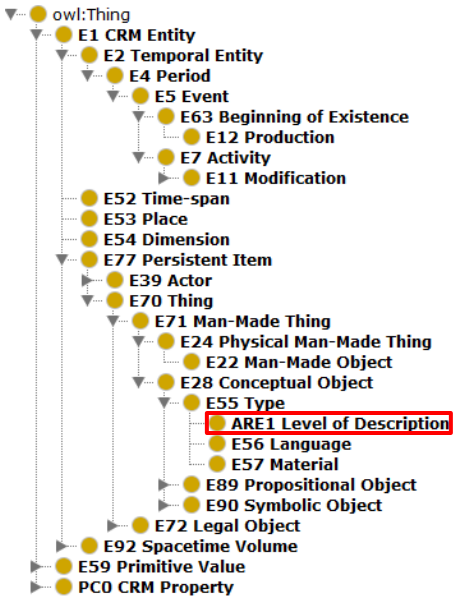


Figura 24 – Classe *ARE1 Level of Description*

Fonte: Autoria Própria

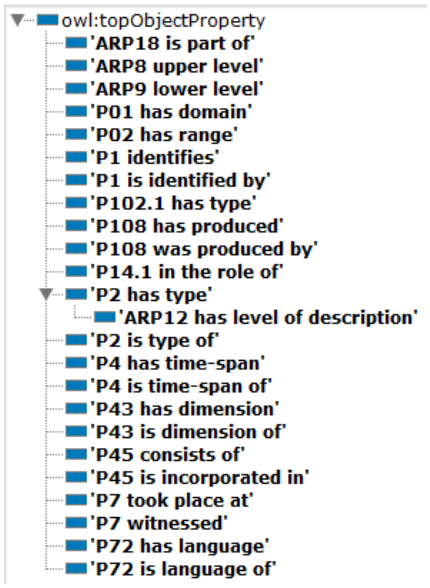


Figura 25 – *Object Properties*

Fonte: Autoria Própria

Devido à necessidade de representar a hierarquia existente nos níveis de descrição, foram também criadas *Object Properties* que permitissem relacionar este mesmos níveis. Assim, foram criadas as propriedades *ARP8 upper level*, *ARP9 lower level* e *ARP18 is part of*.

As duas primeiras propriedades, *ARP8 upper level* e *ARP9 lower level*, foram criadas com o propósito de se definir a hierarquia de níveis de descrição. Estas duas propriedades, inversa uma da outra, pretendem retratar os níveis de descrição possíveis acima ou abaixo de um determinado nível de descrição. Na Figura 26 pode ser observado o exemplo do nível Coleção, o qual tem o nível *Fundo* como *ARP8 upper level* e os níveis Documento Composto, Série e Documento Simples como *ARP9 lower level*.

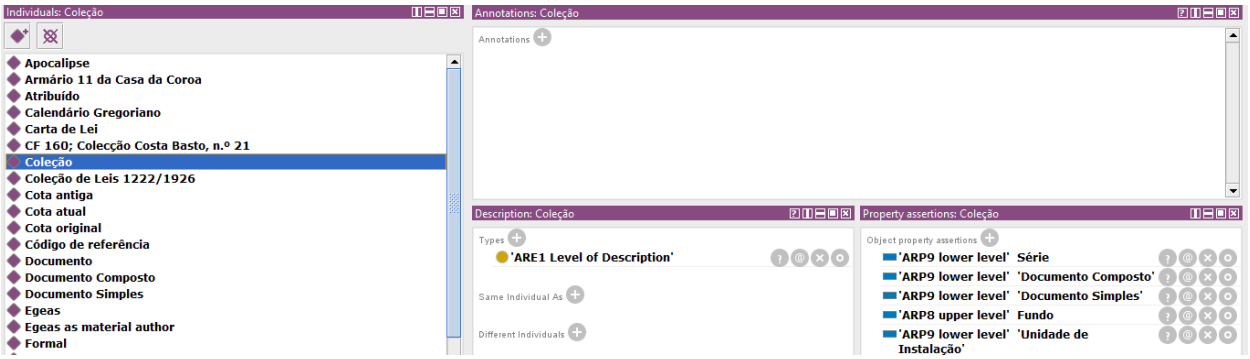


Figura 26 – Exemplo de Nível de descrição

Fonte: Autoria Própria

Já a última propriedade, *ARP18 is part of*, foi criada para indicar que um registo faz parte de outro, como se pode ver na Figura 27. Neste exemplo pode-se constatar que o indivíduo *Calendário Gregoriano* faz parte do indivíduo *Maço 3*, o qual é um registo que está no nível diretamente acima do *Calendário Gregoriano*.

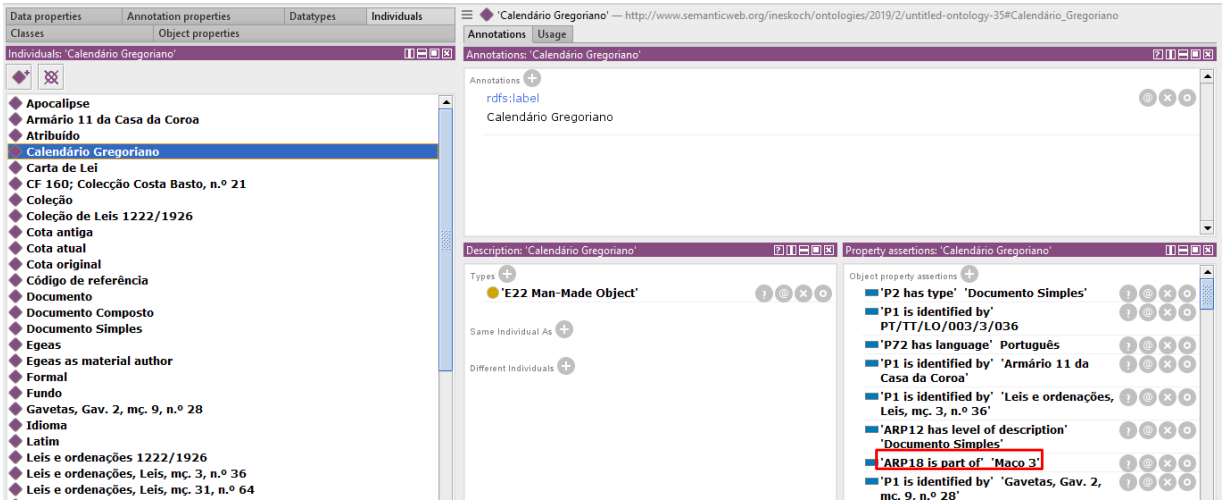


Figura 27 – Exemplo *ARP18 is part of*

Fonte: Autoria Própria

Com a ontologia base já criada, e tendo em conta todos os desenvolvimentos, optou-se pela criação de uma outra ontologia. Esta nova ontologia, denominada *ArchOnto*[14], é baseada na anteriormente criada e teria importada a ontologia original do CIDOC-CRM v6.2.

Com a importação desta ontologia, estarão disponíveis todas as classes e propriedades originárias do CIDOC-CRM, sem a necessidade de as criar e atualizar, uma vez que quando a ontologia for modificada, esta será atualizada só com a atualização do documento que está a ser importado.

Nesta nova ontologia apenas será necessário proceder à criação das extensões, de modo a que a descrição de objetos de arquivos seja elaborada de acordo com o que havia sido estudado e decidido até então.

Para além da ontologia do CIDOC-CRM foi também importada uma outra ontologia, denominada *Data Object*, a qual foi criada para facilitar a validação dos dados existentes no grafo.

[14] Disponível em: https://github.com/feup-infolab/archontology

4.2.1 Relações n-árias

As relações que vão estando presentes na ontologia do CIDOC-CRM são, na sua grande maioria, relações entre dois indivíduos, sendo estas representadas através de propriedades *Object Property*, que captam relações binárias. No entanto, com o desenvolvimento da ontologia para os arquivos, apareceu a necessidade de representar relações não binárias em alguns casos.

Este foi um dos aspetos que foi mais debatido dentro da equipa do projeto, uma vez que envolvia uma solução apropriada para a ontologia, mas também para o grafo de conhecimento.

O primeiro contacto com uma relação intrinsecamente n-ária surgiu com a representação do papel de um determinado autor, numa dada obra. Esta relação demonstrou desde início que seria mais complexa do que todas as relações expressas até então. Para se representar esta relação, tanto a nível da ontologia, como a nível de grafo, foi necessário compreender como é que esta poderia ser apresentada de maneira legível e de fácil compreensão.

Este tipo de relações mostrou-se essencial para a representação de relações que não são binárias, uma vez que com a utilização de relações binárias se perderia uma parte da relação que se quer representar. Por exemplo, no caso da definição do papel de um autor numa dada obra, este facto só se consegue expressar através de uma relação ternária, uma vez que se tirarmos um dos indivíduos envolvidos, a relação resultante deixa de representar o que é pretendido.

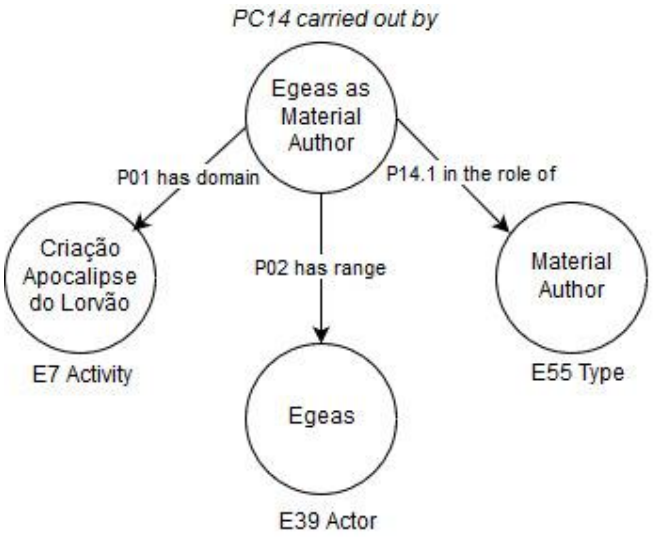


Figura 28 – Exemplo relação n-ária

Fonte: Autoria Própria

A solução apresentada na Figura 28 cria uma nova classe (*PC14 carried out by*) para ligar os três elementos pretendidos - a criação do documento, o seu autor e o papel deste na criação do documento.

Esta nova classe, que vai conter objetos que podem ser considerados indivíduos artificiais, foi criada para corporizar a relação ternária. Como indivíduos artificiais entende-se um qualquer indivíduo que não corresponde a nenhum conceito diretamente existente na realidade que se está a modelar. Por exemplo, a *PC 14 carried out by,* que se encontra na figura acima apresentada, tanto pode ter *Egeas as Material Author*, como *Egeas as Illustrator* como indivíduos.

Foi necessário, também, proceder à criação de duas propriedades que permitissem expressar a relação entre a nova classe criada e a atividade da criação do documento e o seu autor. Esta solução para a representação de uma relação não binária foi criada tendo em conta uma modelação proposta pelo CIDOC-CRM[15].

4.2.2 Validação de dados

A Torre do Tombo foi, ao longo dos anos, criando uma grande quantidade de metadados que são a descrição dos seus objetos culturais. Com a criação de um novo modelo de dados, surge a necessidade de validar os dados que serão inseridos neste novo modelo.

Esta validação partiu, principalmente, da necessidade de haver uma maior clareza no código que incorpora a ontologia[16]. Para o código do grafo, a ontologia passa a ser uma fonte de parametrização, tentando passar o máximo de informação para a validação.

Com a presente validação, o objetivo incide na possibilidade de separar os dados que necessitam de ser validados dos que não necessitam, usando o modelo para fazer esta mesma separação.

Tendo em conta que nas ontologias os dados inseridos podem ter duas naturezas, valores (oriundos das *Data Properties*) ou objetos (oriundos das *Object Properties*), é necessário analisar como é que cada um destes pode ser validado. Para isso foram feitas experiências, onde os dados eram utilizados ora como objetos, ora como valores. Concluiu-

[15] Website CIDOC-CRM consultado pela última vez a 04/06/2019 – Disponível em: http://www.cidoc-crm.org/sites/default/files/Roles.pdf

[16] Freitas, Nuno. "*ArchGraph: Design of a vertical prototype infrastructure for semantic archives.*"MSc, Faculdade de Engenharia da Universidade do Porto, 2019 (em preparação)

se aqui que em certos campos, devido aos conteúdos respetivos e à especificação necessária dos mesmos, a maneira mais fácil e eficaz de os validar é através da mudança da natureza desses dados.

Com isto, alguns dos dados que inicialmente eram considerados valores de *Data Properties* passaram a ser tratados como atributos de indivíduos, incluindo datas e *strings* (Figura 29). Isto deveu-se ao facto de ao serem considerados literais a validação de dados passar pela especificação do contradomínio da propriedade e da sua validação. Com a especificação do contradomínio, que no caso das *Object Properties* é uma classe que define as características dos indivíduos respetivos, a validação passa a ser feita tendo em conta as suas parametrizações, as quais podem estar dependente de ficheiros de autoridade ou de expressões regulares.

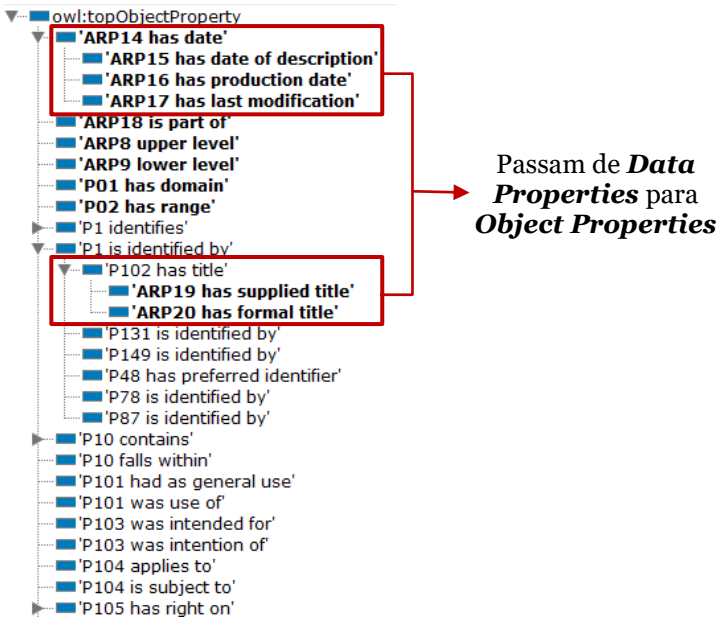


Figura 29 - *Data Properties* que passam a *Object Properties*

Fonte: Autoria Própria

Por sua vez, nas *Data Properties* o contradomínio é, por exemplo, um valor literal e a sua validação torna-se mais complexa. Isto deve-se ao facto de, quando se pretende diferenciar os dados inseridos num mesmo tipo de dados literal, o código necessário para expressar a validação é mais complexo do que quando se utiliza um tipo de uma classe. Isto acontece porque todos os campos com o mesmo tipo de dados fica automaticamente com o mesmo tipo de validação. Por exemplo, quando um valor é considerado uma *String*, todas são validadas da mesma maneira, sejam estas o nome de um autor ou uma história

arquivística ou custodial. Fazer a diferenciação deste tipo de valores requer código mais complexo do que o que está sistematicamente associado a cada classe.

Para captar este tipo de validação criou-se uma nova ontologia a ser importada pelo modelo de dados de arquivo. Nesta nova ontologia, denominada *Data Object*, (ver Figura 30) encontra-se a hierarquia de dados que se pretendem validar. Nesta hierarquia são tidas em conta as diferentes tipologias de dados necessárias para a elaboração da validação.

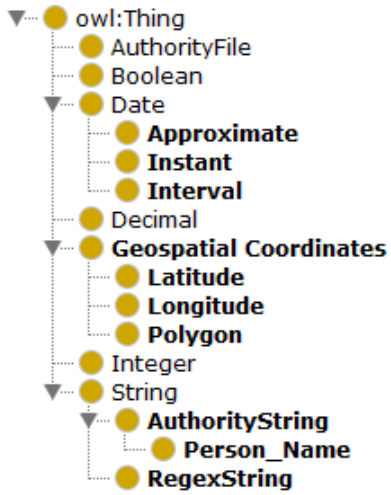


Figura 30 - *Data Object*

Fonte: Autoria Própria

Com a criação desta hierarquia de dados todos os valores que são considerados *String*, por exemplo, passam a ser validados de maneira singular. Isto permite incorporar regras provenientes de ficheiros de autoridade ou expressões regulares que servem como parâmetros de validação do valor inserido numa propriedade em que esta classe é contradomínio. Sempre que se pretender especificar um determinado tipo de *String* surge apenas a necessidade de criar subclasses que permitam a validação desse tipo. Pode-se então dizer que a validação de dados passa pelas classes que estão presentes na ontologia, as quais servem para encapsular tipos de valores e respetiva validação.

Para cada classe teremos *Data Properties* para guardar valores e *Data Properties* para guardar o ficheiro de autoridade.

A validação de *Data Properties* como *Object Properties* neste caso é a mais adequada, uma vez que a base de dados onde está inserida a ontologia é uma base de dados de grafo, a qual não têm *schema* e, como tal, não é tipificada, sendo necessário inferir o tipo de dados.

Com a importação das ontologias do CIDOC-CRM e da *Data Object*, em conjugação com extensões criadas para o arquivo, a ontologia *ArchOnto* encontra-se formulada. No entanto, com a importação destas duas ontologias, surgiram inconsistências que anteriormente não ocorriam. Isto deveu-se ao facto de a ontologia criada inicialmente não ter em conta as restrições que o CIDOC-CRM impõe relativamente ao facto de um objeto apenas poder ser considerado físico ou conceptual. Com isto, um mesmo objeto não pode ter as duas condições simultaneamente, o que na primeira ontologia acontecia.

Inicialmente, as classes criadas para o modelo dos arquivos não tinham em conta quaisquer tipos de restrições, pois estas, muitas vezes, tinham em conta classes que não foram necessárias criar para expressar o domínio dos arquivos.

Com estas inconsistências foi necessário proceder à criação de dois indivíduos para um mesmo registo, pois só com isto é que se poderia designar que um objeto tem características físicas e conceptuais.

Como se pode observar na Figura 31, quando um individuo é físico – *E22 Man-Made Object* – as propriedades utilizadas passam pela identificação do próprio documento físico, como por exemplo, os elementos que descrevem a localização do documento na Torre do Tombo.

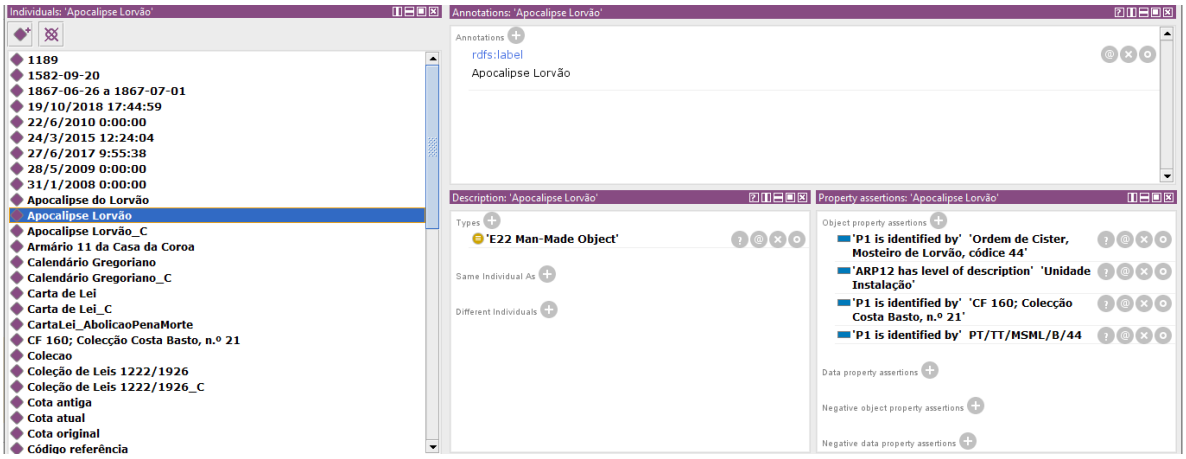


Figura 31 - Apocalipse do Lorvão como objeto físico

Fonte: Autoria Própria

Por sua vez, quando um objeto é considerado conceptual - *E33 Linguistic Object* - são considerados todos os campos que correspondem à descrição arquivística, os quais são referentes ao registo do documento. Entre estes campos encontram-se os campos mais descritivos, como o Âmbito e Conteúdo, a História Custodial e a História Biográfica. Para

além destes, estão também presentes os campos referentes ao idioma em que se encontra o documento, o seu título e as diferentes datas do documento, como a data da sua criação, da sua descrição e da última modificação (Figura 32).

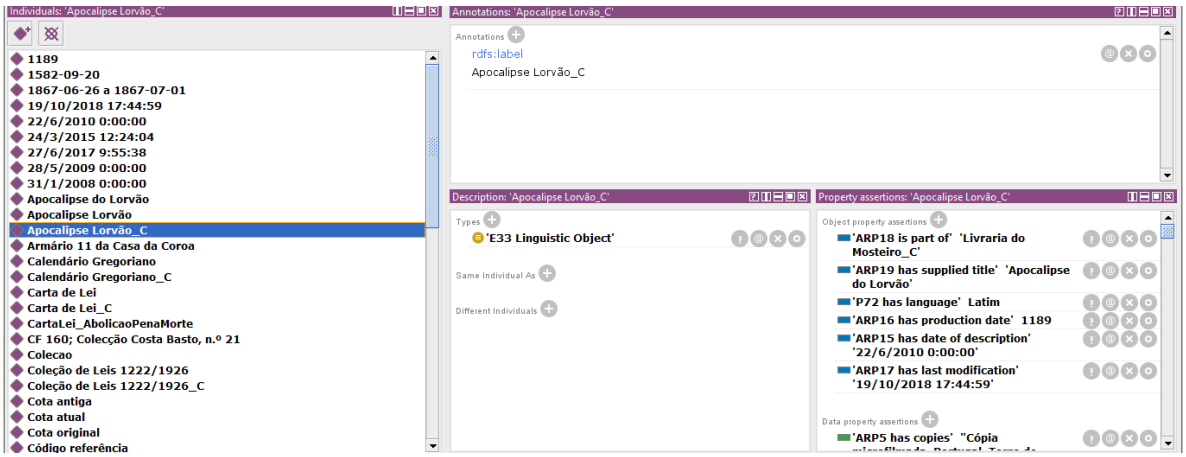


Figura 32 - Apocalipse do Lorvão como objeto conceptual

Fonte: Autoria Própria

4.3 Extração de metadados de descrições ISAD(G)

Ao longo do trabalho realizado ficou claro que alguns dos campos existentes na descrição dos objetos culturais da Torre do Tombo tinham uma grande quantidade de informação a eles associada. Tendo isso em conta, concluiu-se que havia dados que podiam vir a ser extraídos, e representados com recurso a classes CIDOC-CRM que permitissem a existência de descrições que até agora não existiam, devido à natureza da ISAD(G). Este foi um dos objetivos que se teve em conta ao longo da elaboração do novo modelo de dados.

Com este objetivo em vista, foram analisados os campos da ISAD(G) onde seria possível a extração de dados para as classes CIDOC-CRM. Com esta extração de dados poder-se-á povoar a ontologia proposta, acrescentado pessoas, lugares e eventos que anteriormente estariam incluídos em campos de descrição arquivística.

De entre os campos da ISAD(G), os campos da História Custodial e Arquivística, História Administrativa/Biográfica/Familiar e Âmbito e Conteúdo são aqueles que vieram a mostrar-se com conteúdo relevante para ser extraído e formar novas classes.

No caso da História Custodial e Arquivística, um dos exemplos mais claros ocorre no "*Apocalipse do Lorvão*", quando, neste campo, se refere o percurso que este documento

percorreu para chegar à Torre do Tombo: "*O Apocalipse do Lorvão pertenceu ao mosteiro de São Mamede Lorvão, de onde foi trazido para a Torre do Tombo por Alexandre Herculano*". Neste exemplo, é possível retirar vários metadados que podem ser considerados entidades independentes, como o caso do Alexandre Herculano (Pessoa), o mosteiro de São Mamede Lorvão e Torre do Tombo (Lugares) e o ato de levar o Apocalipse do Lorvão de um lugar para o outro (Evento).

Com a extração destes metadados será possível, por exemplo, fazer uma descrição do autor Alexandre Herculano e da sua obra, fazendo a ligação entre este autor e o que se encontra na sua ficha na Biblioteca Nacional de Portugal. Assim sendo, poder-se-á fazer uma destilação das descrições de arquivo com o intuito de afirmar autores, lugares ou eventos de um outro meio e ligá-los a um objeto de arquivo, como no exemplo da História Custodial e Arquivística do "*Apocalipse do Lorvão*".

Os registos de metadados mais atomizados e representados em grafo podem passar a ser incorporados em qualquer sistema *Linked Open Data* (Dados Ligados Abertos) e a ser interoperáveis com as instituições parceiras da DGLAB, como o caso da Biblioteca Nacional de Portugal.

5. Conclusões

O presente projeto visou a criação de um modelo de dados que permitisse a transformação dos dados de descrição arquivística em *Linked Open Data*. Para isso, foi tido como base o CIDOC-CRM, modelo criado na área dos museus e já estudado noutras áreas.

O novo modelo de dados para os arquivos, representado através de ontologias, foi um passo essencial para se compreender como é que o CIDOC-CRM pode ser articulado com as normas de arquivo e os registos existentes atualmente na Torre do Tombo. Com a criação deste modelo, será viável a migração dos registos criados por este arquivo para o CIDOC-CRM. A experimentação do modelo apenas foi elaborado com base em registos para os quais havia conteúdos disponíveis nos vários campos, os quais se mostraram relevantes e com campos com descrições ricas.

O desenvolvimento do novo modelo de descrição arquivística permitiu conhecer melhor no que é que consistem as normas de arquivo e como é que estas são aplicadas no Arquivo Nacional da Torre do Tombo.

5.1 Desafios do projeto

Ao longo da construção da ontologia para os arquivos foram encontrados alguns desafios. Estes surgiram, principalmente, pelo facto de a ontologia ter sido criado para os museus e, por isso, não incorpora todas as classes e propriedades necessárias para a correta representação dos campos existentes nas ISAD(G) e ISAAR(CPF).

O primeiro desafio encontrado foi a grande diversidade de classes e propriedades que esta ontologia possui. Muitas destas, por descreverem conteúdos específicos dos museus, não se apropriam ao modelo para os arquivos. Tendo isso em conta, utilizou-se o modelo CIDOC-CRM como base para a criação de um novo modelo de dados para arquivo.

O segundo desafio identificado foi a quase inexistência de *Data Properties*, as quais se mostram essenciais para a representação dos campos das ISAD(G) e manter o significado desses campos, não passando todos a *String*. A maior parte das propriedades existentes eram do tipo *Object Properties*.

Com estes desafios em vista e as suas soluções, foi elaborado um artigo em colaboração[17], que está aceite para a conferência TPDL 2019 (*Theory and Practice of Digital Libraries 2019*).

Para além destes desafios, foi ainda encontrado um outro. Isto deveu-se ao facto de não fazer parte dos conteúdos de desenvolvimento da ontologia e do grafo de conhecimento, mas sim de outro aspeto relativo ao desenrolar do EPISA.

Este desafio consistiu no acesso tardio à base de dados completa do *Digitarq*, de modo a que pudesse fazer uma análise a fundo das descrições existentes em ISAD(G). Isto também se deveu ao esforço necessário para interpretar o CIDOC-CRM ter sido subestimado, requerendo a construção da ontologia mais esforço que o estimado.

Apesar de este desafio ter surgido, foi possível contorná-lo de modo a que houvesse um primeiro contacto com as descrições dos registos presentes no *Digitarq*, fazendo algumas questões numa base de dados teste e enviando as mesmas para os elementos da DGLAB, de modo a que estes as fizessem diretamente na base de dados e me enviassem os resultados. Assim foi possível fazer uma análise mais superficial da base de dados.

5.2 Trabalho Futuro

O estudo de mapeamento entre as normas ISAD(G) e ISAAR(CPF) e o CIDOC-CRM foi desenvolvido ao longo do período de dissertação, o qual foi, aproximadamente, de seis meses, sendo apenas exequível uma análise e um primeiro mapeamento das atuais normas de arquivo para o novo modelo de dados.

Como o EPISA ainda se encontra na sua fase inicial, há ainda bastante trabalho que necessita de ser desenvolvido após a elaboração desta dissertação. Tendo em conta o que foi desenvolvido até então, denota-se a necessidade de, primeiramente, e após uma análise profunda da base de dados, verificar se os critérios de seleção da amostra que foram escolhidos, no início do novo modelo de descrição arquivística, são os mais indicados, refinando os mesmos. Assim será possível fazer a seleção dos registos que servirão de amostra para o povoamento do modelo.

Com o objetivo de haver uma maior granularidade da informação, tendo em conta a natureza de eventos que o CIDOC-CRM apresenta na sua estrutura, será necessário um refinamento da ontologia. A ontologia elaborada nesta dissertação teve em vista a

[17] Com o Nuno Freitas e equipa do INESC TEC do EPISA

migração dos campos existentes na ISAD(G), os quais têm uma estrutura descritiva, para o CIDOC-CRM. No entanto, pretende-se que no futuro a descrição arquivística deixe de ter textos descritivos de grande dimensão, como o que acontece agora, passando os conteúdos (relatados em documentos) ou a história dos próprios documentos a ser eventos.

6. Referências Bibliográficas

Almeida, Maria José De, and Lucília Runa. 2018. “ICON Project: Content Integration in Portuguese National Archives Using CIDOC-CRM.” In , 1–12. Lisboa, Portugal.

ARIADNE. 2014. “The Way Forward to Digital Archaeology in Europe.” European Commission under the Community’s Seventh Framework Programme. http://ariadne-infrastructure.eu/About.

Börstler, Jürgen. 2001. “Experience with Work-Product Oriented Software Development Projects.” *Computer Science Education* 11 (2): 111–33. https://doi.org/10.1076/csed.11.2.111.3840.

Bountouri, Lina, and Manolis Gergatsoulis. 2010. “Mapping Encoded Archival Description to CIDOC CRM.” *First Workshop on Digital Information Management*, no. March 2011: 8–25.

Bountouri, Lina, and Manolis Gergatsoulis. 2011. “The Semantic Mapping of Archival Metadata to the CIDOC CRM Ontology.” *Journal of Archival Organization* 9 (3–4): 174–207. https://doi.org/10.1080/15332748.2011.650124.

Cordeiro, Edson dos Santos. 2003. “Modelagem Descritiva Iterativa e Incremental de Processo de Software: Uma Experiência Em Uma Microempresa de Desenvolvimento de Software.” Universidade Federal De Santa Catarina.

Direcção Geral De Arquivo. 2007. *Orientações Para a Descrição Arquivística*. 2ª ed. Lisboa: Direção Geral de Arquivos.

Doerr, Martin, Achille Felicetti, Sorin Hermon, Gerald Hiebel, Athina Kritsotaki, Anja Masur, Paola Ronzino, Wolfgang Schmidle, Maria Theodoridou, and Despoina Tsiafaki. 2016. “Definition of the CRMarchaeo: An Extension of CIDOC CRM to Support the Archaeological Excavation Process.” Forth - Institution of Computer Science.

Doerr, Martin, Achille Felicetti, and Francesca Murano. 2017. “Definition of the CRMtex: An Extension of CIDOC CRM to Model Ancient Textual Entities.” Forth - Institution of Computer Science.

Doerr, Martin, Athina Kritsotaki, Yannis Rousakis, Gerald Hiebel, and Maria Theodoridou. 2015. “Definition of the CRMsci: An Extension of CIDOC-CRM to Support Scientific Observation.” Forth - Institution of Computer Science.

Doerr, Martin, and Stephen Stead. 2015. “CRM Inf : The Argumentation Model.” Forth - Institution of Computer Science, no.27 .

Doerr, Martin, and Maria Theodoridou. 2011. “CRM Dig : A Generic Digital Provenance Model for Scientific Observation.” *Proceedings of TaPP’11: 3rd, USENIX Workshop on the Theory and Practice of Provenance*.

Hiebel, Gerald, Martin Doerr, and Øyvind Eide. 2016. “CRMgeo: A Spatiotemporal Extension of CIDOC-CRM.” *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-016-0192-4.

ICOM/CIDOC CRM Special Interest Group. 2017. *Definition of the CIDOC Conceptual Reference Model*. 6.2.2.

International Council on Archives. 2002. “ISAD(G): General International Standard Archival Description.” Estocolmo, Suécia: International Council on Archives.

International Council on Archives. 2004. “ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families.” Paris: International Council on Archives.

International Council on Archives. 2016. “Records in Context: A Conceptual Model for Archival Description.” International Council on Archives.

Lima, João Alberto de Oliveira. 2008. “Modelo Genérico de Relacionamentos Na Organização Da Informação Legislativa e Jurídica.” Universidade de Brasília.

Oldman, Dominic, and CRM Labs. 2014. “The CIDOC Conceptual Reference Model (CIDOC-CRM): Primer.”

Santos, Hercules Pimenta dos. 2016. “Modelo CIDOC CRM: Interoperabilidade Semântica de Informações Culturais.” *Brazilian Journal of Information Studies: Research Trends*, 56–62.

Society of American Archivists. 2018. *Encoded Archival Description (EAD)*. Chicago.

Anexos

Anexo 1 - Conceitos ISAD(G)

- Acesso (*access*) - Possibilidade de utilizar documentação de um fundo, geralmente sujeita a regras e condições.

- Autor (*author*) - Pessoa singular ou colectiva responsável pelo conteúdo intelectual de um documento. Não confundir com produtor.

- Avaliação (*appraisal*) - Processo pelo qual se determina o prazo de conservação de documentos de arquivo.

- Colecção (*collection*) - Conjunto de documentos reunidos artificialmente em função de qualquer característica comum, independentemente da sua proveniência. Não confundir com fundo.

- Custódia (*custody*) - A responsabilidade pela conservação de documentos de arquivo, baseada na sua guarda física. A custódia nem sempre implica a propriedade legal ou o direito de controlar o acesso aos documentos.

- Fundo (*Fonds*) - Conjunto de documentos de arquivo, independentemente da sua forma ou suporte, organicamente produzido e/ou acumulado e utilizado por uma pessoa singular, família ou pessoa colectiva, no decurso das suas actividades e funções.

- Ingresso adicional (*accrual*) - Aquisição de documentos de arquivo complementares de uma unidade de descrição já custodiada por um serviço de arquivo.

- Instrumento de descrição (*finding aid*) - Termo genérico que se aplica a qualquer instrumento de descrição ou de referência, elaborado ou recebido por um serviço de arquivo, com vista ao controlo administrativo ou intelectual dos documentos de arquivo.

- **Organização** (*arrangement*) – Conjunto de operações intelectuais e físicas que consistem na análise, estruturação e ordenação dos documentos de arquivo, e seu resultado.

- Peça (*Item*) - A mais pequena unidade arquivística intelectualmente indivisível, por exemplo: carta, memorando, relatório, fotografia, registo sonoro.

- Pessoa colectiva (*corporate body*) – Organização ou grupo de pessoas identificado por um nome próprio e que atua, ou pode actuar, como uma entidade.

- Ponto de acesso (*access point*) – Nome, termo, palavra-chave, expressão ou código que pode ser utilizado para pesquisar, identificar e localizar uma descrição arquivística.

- Processo (*File*) - Unidade organizada de documentos agrupados, quer para utilização corrente pelo seu produtor, quer no decurso da organização arquivística, por se referirem a um mesmo assunto, actividade ou transacção. Um processo é geralmente a unidade básica de uma série.

- **Produtor** (*creator*) - A pessoa colectiva, família ou pessoa singular que produziu, acumulou e/ou conservou documentos de arquivo no decurso da sua actividade. Não confundir com coleccionador.

- Proveniência (*provenance*) - Relação entre os documentos de arquivo e as pessoas colectivas ou singulares que os produziram, acumularam e/ou conservaram e os utilizaram no decurso de suas atividades.

- Série (*Series*) - Conjunto de documentos organizados de acordo com um sistema de arquivagem e conservados como uma unidade, por resultarem de um mesmo processo de acumulação, do exercício de uma mesma actividade, por terem uma tipologia particular, ou devido a qualquer outro tipo de relação resultante do processo de produção, recepção ou utilização. É também designada como série documental (*records series*)

- Subfundo (*Subfonds*) - Subdivisão de um fundo compreendendo um conjunto de documentos relacionados que corresponde a subdivisões administrativas da agência ou instituição produtora ou, quando tal não é possível, correspondendo a uma divisão geográfica, cronológica, funcional ou agrupamentos de documentos similares. Quando o organismo produtor tem uma estrutura hierárquica complexa, cada seção tem tantas subdivisões subordinadas quantas forem necessárias, de modo a refletir os níveis da estrutura hierárquica da unidade administrativa subordinada primária.

- Suporte (*medium*) - Material sobre o qual a informação é registada (por exemplo: argila, papiro, papel, pergaminho, filme, fita magnética).

- Tipo de documento (*form*) - Uma classe de documentos que se distingue com base em características comuns, físicas (por exemplo: aguarela, desenho) e/ou intelectuais (por exemplo: diário, livro diário, borrador).

- Título (*title*) - Palavra, frase, caracter ou grupo de caracteres que designa uma unidade de descrição.

- Título atribuído (*supplied title*) - Título dado pelo arquivista a uma unidade de descrição que não apresente um título formal.

- Título formal (*formal title*) - Título que aparece proeminente ou explicitamente na documentação de arquivo descrita.

Anexo 2 - Conceitos ISAAR(CPF)

- Descrição arquivística (*Archival description*) - A elaboração de uma representação exacta de uma unidade de descrição e das partes que a compõem, caso existam, através da recolha, análise, organização e registo de informação que sirva para identificar, gerir, localizar e explicar a documentação de arquivo, assim como o contexto e o sistema de arquivo que a produziu. Este termo também se aplica ao resultado desse processo.

- Documento de arquivo (*Record*) - A informação de qualquer tipo, registada em qualquer suporte, produzida ou recebida e conservada por um organismo ou pessoa, no exercício das suas competências, ou actividades.

- Pessoa colectiva (*Corporate body*) - O organismo ou grupo de pessoas identificado por um nome próprio e que age, ou pode agir, como uma entidade. Pode incluir um indivíduo agindo enquanto pessoa colectiva.

- Ponto de acesso (*Access point*) - O nome, termo, palavra-chave, expressão ou código que pode ser utilizado para pesquisar, identificar e localizar descrições arquivísticas, incluindo registos de autoridade.

- Produtor (*Creator*) - A pessoa colectiva, família ou pessoa singular que produziu, acumulou e/ou conservou documentos de arquivo no decurso das suas actividades. Não confundir com coleccionador.

- Proveniência (*Provenance*) - A relação entre os documentos de arquivo e as pessoas colectivas ou singulares que os produziram, acumularam e/ou conservaram e os utilizaram no decurso das suas actividades.

Anexo 3 - Mapeamento ISAD(G) para CIDOC-CRM

	Digitarq	**Classes CIDOC-CRM**
1. Identificação	Nível de descrição	E55 Type
	Código de referência	E42 Identifier
	Referência	E42 Identifier
	Código da entidade detentora	E42 Identifier
	Código do país	E42 Identifier
	Tipo de título	E55 Type
	Título	E35 Title
	Título paralelo	E35 Title
	Data extrema inicial	E52 Time-span
	Certeza da data inicial	E55 Type
	Data extrema final	E52 Time-span
	Certeza da data final	E55 Type
	Datas de acumulação	E52 Time-span
	Datas predominantes	E52 Time-span
	Datas descritivas	E52 Time-span
	Dimensão e suporte	E54 Dimension / E57 Material
	Extensões (designação)	E62 String
	Extensões (número)	E60 Number
	Entidade detentora	E74 Group
	Identificador da entidade detentora	E42 Identifier
2. Contexto	Produtor	E37 Mark
	Identificador da entidade produtora	E42 Identifier
	Autor intelectual	E21 Person
	Autor material	E21 Person
	Colaborador	E21 Person
	Destinatário	E21 Person
	História administrativa/biográfica/familiar	E62 String
	Localidade	E53 Place
	Estatuto Legal	E62 String

	Funções, ocupações e actividades	E62 String
	Mandatos/fontes de autoridade	E62 String
	Contexto geral	E62 String
	História custodial e arquivística	E62 String
	Fonte imediata de aquisição ou transferência	E62 String
3. Conteúdo e estrutura	Âmbito e conteúdo	E62 String
	Assunto	E62 String
	Tradição documental	E62 String
	Tipologia documental	E62 String
	Marcas	E62 String
	Monogramas	E62 String
	Selos	E62 String
	Inscrições	E62 String
	Assinaturas	E62 String
	Avaliação e selecção	E62 String
	Eliminação	E62 String
	Datas de eliminação	E52 Time-span
	Ingressos adicionais	E62 String
	Sistema de organização	E62 String
4. Condições de acesso e utilização	Condições de acesso	E62 String
	Condições de reprodução	E62 String
	Cota atual	E42 Identifier
	Cota original	E42 Identifier
	Cota antiga	E42 Identifier
	Idioma e escrita	E56 Language
	Características físicas e requisitos técnicos	E62 String
	Instrumentos de pesquisa	E62 String
5· Documentação associada	Existência e localização de originais	E62 String
	Existência e localização de cópias	E62 String
	Unidades de descrição relacionadas	E62 String
	Notas de publicação	E62 String
6. Notas	Notas	E62 String

7. Controlo da descrição	Título (Notas do arquivista)	E35 Title
	Nota (Notas do arquivista)	E62 String
	[Autor] (notas do arquivista)	E21 Person
	[data] (notas do arquivista)	E52 Time-span
	Regras ou convenções	E62 String
	Data de criação	E52 Time-span
	Criado por	E21 Person
	Alterado por	E21 Person
	Última modificação	E52 Time-span
	Notas de migração	E62 String
	Nota de edição	E62 String
8. Relações	Roda AIP ID	E42 Identifier