

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Caliel Cardoso de Oliveira

**Ontologias como artefatos para a representação de dados abertos de pesquisa:  
uma exploração acerca de uma modelagem experimental de dados da Arqueologia  
através do modelo CIDOC Conceptual Reference Model**

Porto Alegre  
2023

Caliel Cardoso de Oliveira

**Ontologias como artefatos para a representação de dados abertos de pesquisa:  
uma exploração acerca da modelagem de dados da Arqueologia através do modelo  
CIDOC Conceptual Reference Model**

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul como requisito parcial para a aquisição do título de Mestre em Ciência da Informação.

Porto Alegre  
2023

## **UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

Reitor: Prof. Carlos André Bulhões

Vice-Reitora: Prof.<sup>a</sup> Patrícia Pranke

## **FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

Diretora: Prof.<sup>a</sup> Dr.<sup>a</sup> Karla Maria Müller

Vice Diretora: Prof.<sup>a</sup> Dr.<sup>a</sup> Ilza Maria Tourinho Girardi

## **DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

Chefe: Rene Faustino Gabriel Júnior

Chefe Substituto: Caterina Marta Groposo Pavão

## **PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

Coordenador: Prof. Dr. Thiago Henrique Bragato Barros

Coordenador Substituto: Prof. Dr. Moises Rockemback

### **CIP - Catalogação na Publicação**

Oliveira, Caliel Cardoso de

Ontologias como artefatos para a representação de dados abertos de pesquisa: uma exploração acerca de uma modelagem experimental de dados da Arqueologia através do modelo CIDOC Conceptual Reference Model / Caliel Cardoso de Oliveira. -- 2023.

174 f.

Orientador: Thiago Henrique Bragato Barros.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Programa de Pós-Graduação em Ciência da Informação, Porto Alegre, BR-RS, 2023.

1. Organização e Representação do Conhecimento. 2. Ontologias. 3. Web Semântica. 4. Dados abertos de pesquisa. 5. Dados abertos e conectados. I. Barros, Thiago Henrique Bragato, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Faculdade de Biblioteconomia e Comunicação

Departamento de Ciências da Informação

Rua Ramiro Barcelos, 2705, Bairro Santana

Porto Alegre/RS – CEP 90035-007

Telefone: 51 3308 5067

E-mail: fabico@ufrgs.br

CALIEL CARDOSO DE OLIVEIRA

**ONTOLOGIAS COMO ARTEFATOS PARA A REPRESENTAÇÃO DE DADOS ABERTOS DE PESQUISA:** uma exploração acerca de uma modelagem experimental de dados da Arqueologia através do modelo CIDOC Conceptual Reference Model

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

Porto Alegre, 28 de Abril de 2023.

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. RITA DO CARMO FERREIRA LAIPELT - Membro titular interno  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. MÁRCIO BEZERRA DA SILVA - Membro titular externo  
Universidade de Brasília<sup>1</sup>

---

Prof<sup>a</sup>. Dr<sup>a</sup>. RENATA CARDOZO PADILHA - Membro titular externo  
Universidade Federal de Santa Catarina

---

Prof. Dr. RENE FAUSTINO GABRIEL JUNIOR - Membro suplente interno  
Universidade Federal de Santa Catarina

## AGRADECIMENTOS

Agradeço, primeiramente, ao Universo por garantir o bom andamento desta pesquisa e da minha formação. A vida toma caminhos inesperados, e permanecer na Academia depois da graduação não estava nos meus planos. Mas essa reviravolta no roteiro me permitiu conhecer pessoas, me inserir melhor em minha área e perceber meu valor como pesquisador. No mais, digo o de sempre: ser humano muitas vezes não colabora, e, no meu caso, o ambiente doméstico sempre prejudicou o melhor andamento da minha ainda jovem carreira acadêmica. Mas esse é mais um passo dado e, como diz um das minhas fábulas favoritas: tudo está bem quando acaba bem.

Em segundo lugar, agradeço a meus afetos e amizades, por me darem força e ânimo para seguir em frente ao longo desta trajetória de mestrando. Tivemos muitas brigas e até alguns afastamentos, mas também pudemos trazer novas pessoas para nosso convívio e, particularmente, sinto que o mestrado me permitiu conhecer amigos que pretendo levar para o restante da vida. Agradeço em especial a meu namorado Maurício e minha melhor amiga, Bruna. Sem a parceria de ambos, não teria sobrevivido ao caos do mestrado, da pandemia e da vida doméstica que quase me consumiram. Agradeço também à Naya, que, apesar de tudo, ainda não desistiu de mim; à Rafaella, que, apesar da distância, está sempre por perto, e à sempre presente Dani, que nunca me deixa desanimar. Por fim, agradeço aos novos amigos que pude conhecer (ou manter por perto) graças à vida acadêmica: Rafael, Suelen e Náthali.

Agradeço à minha família pelo suporte que me deram durante a vida escolar e, novamente, durante a vida acadêmica. Apesar do clima doméstico infernal proporcionado por ela, não sei onde estaria agora sem vocês. Tenho a certeza de que todos, mesmo os que já não estão aqui, se orgulham muito de onde me permitiram chegar.

Em tempo, agradeço ao meu orientador, Professor Thiago, pelo imenso apoio dado durante este percurso. Sem ele para me acalmar, guiar e, bem, orientar, eu certamente teria abandonado este período de formação muito antes de chegar a este momento. Entre me incentivar, apontar locais onde poderia melhorar e resolver pepinos administrativos, o professor Thiago foi o melhor orientador que eu poderia ter tido neste mestrado. Agradeço também à professora Rita pelas contribuições oferecidas desde antes da banca de defesa: as aulas de Terminologia são algumas das memórias mais queridas que guardo do curso, oferecendo algum alento, diversão e aprendizado durante o infernal período da Pandemia. Igualmente, agradeço a todos os professores do PPGCIN, em especial às professoras Caterina e Ana, e à equipe administrativa do programa, bem como aos colegas do ORCALAB: espero que possamos trabalhar muito juntos!

Por fim, agradeço à Arte e a todos os artistas do mundo. Sem música, não estaria vivo neste dia.

O Grande Khan já folheava em seu atlas os mapas das ameaçadoras cidades que surgem nos pesadelos e nas maldições: Enoque, Babilônia, Yahoo, Butua, Admirável Mundo Novo.

Disse:

— É tudo inútil, se o último porto só pode ser a cidade infernal, que está lá no fundo e que nos suga num vórtice cada vez mais estreito.

E Polo:

— O inferno dos vivos não é algo que será; se existe, é aquele que já está aqui, o inferno no qual vivemos todos os dias, que formamos estando juntos. Existem duas maneiras de não sofrer. A primeira é fácil para a maioria das pessoas: aceitar o inferno e tornar-se parte deste até o ponto de deixar de percebê-lo. A segunda é arriscada e exige atenção e aprendizagem contínuas: tentar saber reconhecer quem e o que, no meio do inferno, não é inferno, e preservá-lo, e dar-lhe espaço.

— Ítalo Calvino, *As Cidades Invisíveis* —

**Resumo:** Trata-se de um estudo aplicado e experimental, tem como objetivo geral verificar as potencialidades das ontologias como artefatos de Representação e Organização do Conhecimento de dados científicos na área da Arqueologia. Explora bibliografia pertinente ao objetivo, apresentando referencial sobre Ontologia, Web Semântica e o movimento pelos Dados Abertos de pesquisa. Define os dados abertos de pesquisa como objeto de análise. Seleciona pacote de dados oriundo da Arqueologia como domínio temático para experimento de representação em ontologia. A partir do padrão CIDOC-CRM, modela conceitualmente o conjunto de dados selecionado. Com base na metodologia OntoForInfoScience, desenvolve um artefato ontológico funcional, contendo 22 classes oriundas da etapa de conceitualização e 20 classes nativas do padrão CIDOC-CRM. Representa, no artefato ontológico, 200 indivíduos, bem como os relacionamentos entre eles. Avalia o artefato ontológico construído a partir das etapas de verificação da metodologia escolhida. Disponibiliza o artefato criado em repositório online, e a documentação pertinente a ele como apêndices. Conclui defendendo a manutenção de repositório especializados no tratamento de dados científicos e a utilização de ontologias como ferramentas de apoio para a recuperação da informação.

**Palavras-Chave:** Representação do Conhecimento; Organização do Conhecimento; Ontologias; Dados abertos de pesquisa; Dados abertos e conectados; Web Semântica.

**Abstract:** This is an applied and experimental study, with the general objective of verifying the potential of ontologies as artifacts for the Representation and Organization of Knowledge of scientific data in the area of Archaeology. It explores bibliography pertinent to the objective, presenting references on Ontology, Semantic Web and the movement for Open Research Data. Defines open research data as the object of analysis. Select data package from Archeology as thematic domain for ontology representation experiment. Based on the CIDOC-CRM standard, it conceptually models the selected data set. Based on the OntoForInfoScience methodology, it develops a functional ontological artifact, containing 22 classes from the conceptualization stage and 20 native classes from the CIDOC-CRM standard. It represents, in the ontological artifact, 200 individuals, as well as the relationships between them. Evaluates the ontological artifact built from the verification steps of the chosen methodology. It makes available the artifact created in an online repository, and the documentation related to it as appendices. It concludes by defending the maintenance of a repository specialized in the treatment of scientific data and the use of ontologies as support tools for information retrieval..

**Keywords:** Ontologies; Knowledge Representation; Knowledge Organization; Open research data; Open and connected data; Semantic Web.



## LISTA DE FIGURAS

<b>Figura 1:</b> diagrama relacional do banco de dados de Adaime (2021).....	77
<b>Figura 2:</b> recorte do mapa conceitual destacando o nodo “Representação do amuleto no banco de dados”.....	81
<b>Figura 3:</b> recorte do mapa conceitual com diversos exemplos de um agente realizando atividades.....	83
<b>Figura 4:</b> recorte adaptado do mapa conceitual, com a representação do processo de criação de referencial na dissertação de mestrado de Adaime (2021).....	84
<b>Figura 5:</b> recorte adaptado do mapa conceitual, com a modelagem dos sítios arqueológicos estudados por Adaime (2021) e sua representação no banco de dados criados pela autora.....	85
<b>Figura 6:</b> recorte adaptado do mapa conceitual, com a relação entre os números de identificação dos elementos estudados (amuletos) por Adaime (2021), os elementos em si e suas representações.....	86
<b>Figura 7:</b> recorte adaptado do mapa conceitual, com a cadeia de relações entre os amuletos, sua forma real, sua representação e a representação das tipologias criadas por Adaime para as formas presentes nos relatórios de escavação.....	87
<b>Figura 8:</b> recorte adaptado do mapa conceitual, expondo a teia de relações entre os amuletos, enterramentos e sítios arqueológicos, bem como suas representações dentro do banco de dados criado por Adaime (2021).....	89
<b>Figura 9:</b> recorte adaptado do mapa conceitual, expondo a teia de relações entre os sítios arqueológicos, as escavações e os relatórios de escavação, conforme apresentado por Adaime (2021).....	90
<b>Figura 10:</b> taxonomia básica do modelo CIDOC-CRM 7.1.1.....	93
<b>Figura 11:</b> uso das <i>Annotations</i> para a classe <i>MP3 Aspectos físicos do enterramento</i> .....	95
<b>Figura 12:</b> uso da ferramenta de restrição do Protégé para definição da classe <i>E2 TEMPORAL ENTITY</i> .....	96
<b>Figura 13:</b> definição formal da classe <i>MP5 Enterramento</i> .....	97
<b>Figura 14:</b> definição formal da classe <i>MP17 Representação do amuleto no banco de dados</i> .....	98
<b>Figura 15:</b> definição formal da classe <i>MP8 Forma do Amuleto</i> .....	99
<b>Figura 16:</b> definição formal classe <i>MP22 Tipo de Forma do amuleto</i> .....	100

<b>Figura 17:</b> uso da propriedade <i>P81 ongoing throughout</i> .....	102
<b>Figura 18:</b> representação do indivíduo "Representação do amuleto 115".....	103
<b>Figura 19:</b> representação do indivíduo "Gato".....	104
<b>Figura 20:</b> representação do indivíduo "Amuleto 115".....	105
<b>Figura 21:</b> diagnóstico do módulo <i>debugger</i> do Protegé.....	108

## LISTA DE QUADROS

<b>Quadro 1:</b> especificações do modelo ontológico proposto.....	73
<b>Quadro 2:</b> critérios para a avaliação do modelo proposto.....	109
<b>Quadro 3:</b> documentação para cada etapa da metodologia.....	116

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>14</b>
<b>2 PROBLEMA E JUSTIFICATIVA</b>	<b>15</b>
<b>3 OBJETIVOS DA PESQUISA</b>	<b>18</b>
3.1 Objetivo Geral	18
3.2 Objetivos Específicos	18
<b>4 REFERENCIAL TEÓRICO-METODOLÓGICO</b>	<b>19</b>
4.1 Sistemas de Organização do Conhecimento, Ontologias e seu papel na Web Semântica	19
4.1.1 Metadados e Interoperabilidade	23
4.1.2 Web Semântica	28
4.1.3 Breve histórico da evolução do termo Ontologia	33
4.1.4 As ontologias como artefatos de Organização e Representação do Conhecimento	37
4.1.5 Elementos e tipologias das ontologias	40
4.1.6 Terminologia	43
4.2 O movimento pelo Acesso Aberto e os Dados Abertos	51
4.2.1 As iniciativas de Acesso Aberto	52
4.2.2 Dados abertos de pesquisa	54
4.2.3 Repositórios de dados abertos no Brasil e no mundo	58
4.2.4 O padrão CIDOC-CRM	63
<b>5 DESCRIÇÃO DA METODOLOGIA ADOTADA E CONTEXTUALIZAÇÃO DOS DADOS DO ESTUDO</b>	<b>65</b>
5.1 Contextualização dos dados utilizados	65
5.2 Descrição da metodologia utilizada	68
<b>6 DESCRIÇÃO DOS PROCEDIMENTOS METODOLÓGICOS E APRESENTAÇÃO DOS RESULTADOS</b>	<b>70</b>
6.1 Etapa 0: avaliação da necessidade da ontologia	70
6.2 Etapa 1: especificação da ontologia	73
6.3 Etapa 2: Aquisição e Extração de conhecimento	75
6.4 Etapas 3 e 4: Conceituação e Fundamentação Ontológica	77
6.5 Etapa 5: Formalização da ontologia	92
6.5.1 Primeiro passo: criação da taxonomia geral da ontologia	93
6.5.2. Segundo passo: Definição das propriedades descritivas da ontologia	94
6.5.3. Terceiro passo: Definição formal para as classes da Ontologia	95
6.5.4. Quarto passo: definição das propriedades de dados das classes	101
6.5.5. Quinto Passo e Sexto Passo: criação das instâncias das classes e especificação das relações ontológicas	102
6.5.6. Sétimo passo: definição das propriedades das relações ontológicas	106
6.6 Etapa 6: Avaliação da ontologia	107
6.7 Etapa 7: Documentação da ontologia	116
6.8 Etapa 8: Disponibilização da ontologia	118

<b>7 CONSIDERAÇÕES FINAIS</b>	<b>119</b>
<b>REFERÊNCIAS</b>	<b>122</b>
<b>APÊNDICE A - MAPA CONCEITUAL DESENVOLVIDO DURANTE A PESQUISA</b>	<b>133</b>
<b>APÊNDICE B - QUADRO DE CLASSES CIDOC-CRM</b>	<b>134</b>
<b>APÊNDICE C - QUADRO DE CLASSES NATIVAS DO MODELO PROPOSTO</b>	<b>147</b>
<b>APÊNDICE D - QUADRO DE RELACIONAMENTOS</b>	<b>158</b>

## 1 INTRODUÇÃO

A temática da Ontologia Aplicada é fruto da intersecção entre a Filosofia, onde surge como objeto da Metafísica, com a Ciência da Computação (CC), onde suas aplicações servem para os esforços de Representação do Conhecimento (RC) para o desenvolvimento de Inteligências Artificiais (IA). A partir da década de 1990, o tema passa a se consolidar também na Ciência da Informação (CI), especialmente servindo à área da Organização do Conhecimento (OC) e no compartilhamento de dados abertos e conectados, cuja representação adequada é fundamental para o desenvolvimento de sistemas inteligentes, capazes de alcançar os objetivos idealizados com a proposta do modelo da Web Semântica. Além disso, propostas como a interoperabilidade entre sistemas, que só pode ser alcançada por meio da organização e representação adequada do conhecimento específico, dependem do desenvolvimento de artefatos para a representação e organização do conhecimento, também chamados genericamente de ontologias, produtos do esforço de cientistas da informação e engenheiros do conhecimento<sup>1</sup>. Paralela ao avanço das ontologias, outra temática que se mostra emergente nos debates da Ciência da Informação é a chamada Ciência Aberta, ou seja, o movimento pela abertura e socialização dos processos de produção científica. Ainda dentro do contexto da Ciência Aberta, emergem as iniciativas de divulgação de dados abertos, que pregam o compartilhamento responsável dos dados oriundos de pesquisas, tornando-os de fácil acesso tanto para a comunidade acadêmica quanto para o público em geral.

Nesse sentido, as ontologias, com seu potencial no que tange ao compartilhamento de dados estruturados interoperáveis e à modelagem do conhecimento especializado, se mostram artefatos úteis para a disponibilização adequada de dados de pesquisa. Dessa forma, este trabalho surge como tentativa experimental de se modelar, através do desenvolvimento de uma ontologia, os dados obtidos durante uma pesquisa da área da Arqueologia, intitulada *Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos*, desenvolvida pela pesquisadora Victoria Arroyo Adai, no ano de 2021, com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo

---

<sup>1</sup> A Engenharia do Conhecimento é uma área da Ciência da Computação, surgida para lidar com o alto grau de arbitrariedade e imprecisão que prejudicava o desenvolvimento de bases de conhecimento, dificultando a criação de sistemas automatizados e IAs (SMITH; WELTY, 2001)

(FAPESP). O modelo de representação escolhido para este experimento foi o CIDOC Conceptual Reference Model (CRM), sistema proposto pelo Conselho Internacional de Museus (ICOM) para a modelagem do domínio do Patrimônio Cultural, Arte e Arqueologia em instituições de custódia e preservação de artefatos históricos e artísticos. Por meio da representação destes dados, a partir do modelo ontológico oferecido pelo CIDOC-CRM, espera-se demonstrar como a aplicação dos princípios ontológicos para a organização do conhecimento científico pode favorecer a interoperabilidade entre diferentes sistemas e organizações, ampliando o alcance da produção acadêmica por meio da representação adequada dos dados de pesquisa.

## 2 PROBLEMA E JUSTIFICATIVA

A recuperação da informação técnico-científica vem sendo, desde o final da Segunda Guerra Mundial, uma das necessidades centrais da comunidade acadêmica, e tornou-se objeto de estudo da Ciência da Informação. Organizar o conhecimento especializado de maneira apropriada, representá-lo de maneira eficiente e recuperá-lo são preocupações da área, e entende-se que o último passo, da recuperação da informação, não pode ocorrer da maneira devida sem que as duas etapas anteriores sejam executadas adequadamente. Com o avanço da Internet e a gênese da *Web* ao longo do século XX, essas etapas passaram a levar em conta a existência dos ambientes digitais, e também tiveram de lidar com a facilitação da comunicação especializada e, portanto, com o aumento da produção científica. Nesse sentido, a aproximação entre a Ciência da Informação com a Ciência da Computação começa a ocorrer em diversas frentes, em particular no que tange às novas práticas e aos novos saberes oriundos da Representação do Conhecimento e das propostas da *Web Semântica*. Dessa forma, considerando as demandas necessárias à consolidação da *Web Semântica*, é preciso que, cada vez mais, o conhecimento esteja representado por meio de dados; e não apenas dados estruturados sintática e estaticamente, mas sim dados semanticamente enriquecidos, conectados a outros recursos *web* que permitam uma apreensão mais completa do conhecimento por eles representado.

Para que haja essa abertura e conectividade, é necessário que estes dados se apresentem da maneira o menos heterogênea possível, através da construção de páginas estruturadas e compreensíveis tanto pelo ser humano (como é a *Web de Dados*

contemporânea) quanto pela Máquina (computadores, sistemas automatizados, dentre outros). Dessa forma, vocabulários que ajudem a diminuir a ambiguidade dos termos são peça-chave para superar essa heterogeneidade, e, nesse ponto, a Ciência da Informação ainda pode oferecer muito ao desenvolvimento da *Web Semântica*, como será demonstrado em nosso referencial teórico. Detentora do saber especializado relativo à Organização do Conhecimento, a CI oferece o arcabouço técnico-teórico para a elaboração de sistemas de organização do conhecimento (SOCs) voltados para esse fim; no contexto *Web*, a intersecção com a CC aponta para o papel do cientista da informação como agente ativo, portanto, no desenvolvimento dos artefatos de *software* conhecidos como ontologias, além de todo o material que os cerca. A capacidade de representar um domínio específico do saber é, assim, uma habilidade extremamente valiosa quando se considera as potencialidades relativas ao enriquecimento semântico das páginas da *Web*, e a transformação adequada dos conceitos pertencentes a dado domínio em recursos web passa a ser uma etapa fundamental em qualquer esforço de representação do conhecimento no ambiente virtual.

Conforme mencionado no primeiro parágrafo desta seção de justificativa, a facilitação na comunicação científica, somada aos avanços da Internet e da Computação num todo levaram a um novo *boom* informacional, com uma produção imensa de dados científicos ocorrendo diariamente. Com este panorama, a necessidade de recuperação desses dados especializados se tornou mais um elemento a ser considerado quando se trata da implantação *Web Semântica* e da demanda por mais dados abertos e compartilhados. No entanto, ainda é comum entre a comunidade científica que haja equívocos quanto à maneira como os dados oriundos de suas pesquisas são disponibilizados ao público e a seus pares, e isso acaba se refletindo em uma dificuldade na recuperação destes dados tanto por atores interessados nos dados quanto por aplicações (*softwares*) que manipulam esses dados. Iniciativas como repositórios de dados abertos de pesquisa ainda se mostram incipientes no Brasil, apesar de seus avanços no cenário internacional, e mesmo orientações de caráter não-obrigatório, como o sistema de classificação de “5 Estrelas” proposto por Tim Berners-Lee ainda podem passar despercebidos pela comunidade acadêmica em geral, que nem sempre dispõe de recursos ou tempo-hábil para realizar a disponibilização adequada de seus dados.

Mediante o exposto, este trabalho surge como uma tentativa de oferecer uma contribuição ao crescente arcabouço teórico e metodológico envolvendo dados abertos



no contexto da *Web Semântica*. Através da modelagem de uma ontologia, pretende-se demonstrar como a organização e representação adequadas do conhecimento científico podem facilitar sua recuperação e subsequente uso. A escolha por este conjunto de dados em específico se deu por alguns motivos: primeiro, os dados escolhidos encontram-se disponibilizados em formato *Portable Document Format* (PDF) como um pacote de dados oriundo da pesquisa que os originou. Embora sejam de fácil compreensão e para usuários humanos, arquivos em PDF são apontados como exemplos de dados *web* de “1 estrela”, conforme o modelo de “5 Estrelas” proposto por Berners-Lee (OPEN DATA, 2012), o que quer dizer que há pouca possibilidade de utilização destes dados por aplicações e outros agentes de IA, que poderiam servir como ferramentas para novos empreendimentos científicos. Desta forma, a construção de uma ontologia que permita a descrição de maneira aberta e semanticamente enriquecida dos dados escolhidos os trará para mais perto do ideal de estruturação de conteúdo com significado desejado pela proposta da *Web Semântica*, bem como do ideal de compartilhamento de dados abertos de pesquisa para fins de reuso e validação das investigações científicas, como almejado pelo movimento da Ciência Aberta. Em segundo lugar, a escolha destes dados se deu pois defende-se, neste estudo, a importância de que pesquisas produzidas nas Ciências Humanas não sejam “deixadas de lado” pelos esforços de Representação do Conhecimento para fins de uso na *Web*: embora por vezes menos quantificáveis que aqueles dados produzidos nas Ciências Exatas e Aplicadas, sua importância como conhecimento especializado não pode ser reduzida. Também neste sentido, a combinação entre dados produzidos nas Humanidades, os quais demandam conceitualização de seu contexto para possuírem maior valor, e as ontologias, sistemas capazes de oferecer a conceitualização deste contexto, se mostrou campo fértil para o desenvolvimento de nosso estudo. Por fim, por se tratarem de dados produzidos em uma pesquisa brasileira, espera-se também que este trabalho possa contribuir para o fomento e a expansão das boas práticas de compartilhamento de dados de pesquisa e construção de modelos de representação conceitual para a *Web Semântica* no país. Relacionado ao ponto anterior, nossa intenção é também de oferecer uma contribuição, ainda que incipiente, ao panorama da representação de dados das Humanidades em repositórios de dados brasileiros, demonstrando o valor que estes dados podem oferecer e o potencial existente na manutenção de repositórios especializados neste tipo de produto de pesquisa

Apresentados os argumentos que justificam a realização deste trabalho, **o problema de pesquisa** que o motiva é: Quais os benefícios da modelagem ontológica para a representação adequada dos dados de pesquisa em repositórios institucionais?

### **3 OBJETIVOS DA PESQUISA**

A seguir, são apresentados os objetivos que se almeja alcançar com este estudo.

#### **3.1 Objetivo Geral**

Desenvolver um artefato ontológico, como um experimento, a fim de verificar as potencialidades acerca da organização, representação e recuperação de dados de pesquisa oriundos da área da Arqueologia por meio de ontologias.

#### **3.2 Objetivos Específicos**

- 1) Relacionar a bibliografia sobre em representação e organização do conhecimento com a que trabalha com gestão de dados de pesquisa;
- 2) Analisar e descrever os conceitos que fundamentam o conjunto de dados selecionado;
- 3) Modelar ontologicamente o domínio, buscando representar os conceitos básicos que o compõem, as relações que os conectam e as propriedades que os caracterizam;
- 4) Representar os dados que integram o conjunto selecionado como instâncias do domínio;
- 5) Avaliar a eficiência da ontologia desenvolvida enquanto ferramenta de representação do conhecimento especializado.

## 4 REFERENCIAL TEÓRICO-METODOLÓGICO

A seguir, expõe-se o referencial teórico que fundamenta o desenvolvimento desta pesquisa. A seção divide-se em três subseções: na primeira, explora-se o que são Sistemas de Organização do Conhecimento, em especial as ontologias, e sua relação com a *Web Semântica*, além de se discorrer brevemente sobre a Terminologia e sua importância para os esforços de Organização do Conhecimento. A segunda subseção trata das questões referentes ao movimento pelo Acesso Aberto (AA) dos dados de pesquisa, seus fundamentos e implicações. Por fim, a terceira seção abordará o histórico do movimento AA no Brasil, contextualizando o Repositório<sup>2</sup> de dados da Federação de Apoio à Pesquisa do Estado de São Paulo, onde se encontram os dados escolhidos para a realização desta pesquisa.

### 4.1 Sistemas de Organização do Conhecimento, Ontologias e seu papel na *Web Semântica*

Preocupações acerca da categorização e organização do conhecimento sempre permearam o pensamento humano. Desde a Antiguidade grega, com a transição do pensamento contemplativo acerca da natureza das coisas para o pensamento compreensivo, filosófico, que buscava entender sobre o Universo e a realidade, o conhecimento passou a servir como ponto de partida para a construção de novos saberes (MOREIRA, 2018). Para que ocorra essa geração de novos saberes, é necessário que se desenvolvam dois processos distintos: a comunicação do conhecimento (que, embora seja construído individual e tacitamente, pode ser socializado) e a organização dos saberes, processo que, apesar de considerado desde a Antiguidade, ganhou caráter de objeto científico da Biblioteconomia e da Documentação nos séculos XIX e XX, a partir do estudo do tratamento do conhecimento registrado (*Ibid.*, 2018). Nesse sentido, não se pode desconsiderar que o desenvolvimento técnico-científico massivo que ocorre desde o século passado venha gerando uma massa enorme de conhecimento, além de dados que servem como subsídio para esse conhecimento e de informações que o alteram, de forma que o tratamento e estruturação adequados do conhecimento se tornam centrais para que ele seja compartilhado e possa cumprir sua função na criação de novos conhecimentos (CINTRA *et al.*, 1994). Em resposta a esse salto na produção do

---

<sup>2</sup> <https://metabuscador.uspdigital.usp.br/>

conhecimento, alguns marcos técnico-teóricos ocorreram a partir da segunda metade do século XX, como a mecanização dos serviços de informação e a gênese da própria Ciência da Informação (GOMES, 2017).

A expressão *Organização do Conhecimento*, por sua vez, foi cunhada em 1989, durante a fundação da *International Society for Knowledge Organization* (ISKO), explica Dahlberg (2006), embora sua origem seja mais antiga, derivando-se da *Organization of Knowledge* proposta pelo bibliotecário Henry Evelyn Bliss na década de 1930. Dahlberg ainda afirma que se considerarmos a Organização do Conhecimento como uma nova ciência, então seu objeto de estudo é o conhecimento em si, definido pela autora como “[...] a certeza razoavelmente bem-fundamentada, de forma subjetiva e objetiva, que alguém possui acerca da existência de um fato ou assunto [...]” (DAHLBERG, 2006, p.12, tradução nossa). Nesse sentido, Moreira (2018) afirma, com base em Brascher e Café (2008), que a Organização do Conhecimento é aplicada sobre unidades do pensamento (ou conceitos), e que visa à criação de modelos de mundo que se constituam como abstrações da realidade. A esses modelos dá-se o nome de Sistemas de Organização do Conhecimento (SOCs), dentre os quais se destacam dicionários, taxonomias, tesouros, ontologias, rede semânticas<sup>3</sup>, dentre outros (SCHISSL; SHINTAKU, 2012)

Para que possa ser feita a construção desses SOCs, a literatura apresenta diferentes formas de se compreender um campo do conhecimento. De acordo com Moreira (2018), pode-se fazê-lo a partir da epistemologia do campo, ou seja, de suas fundamentações teóricas; pode-se fazê-lo a partir de sua metodologia, ou seja, dos métodos e técnicas dos quais ele faz uso; e pode-se fazê-lo ainda por meio dos conceitos que ele emprega e da rede de inter-relacionamento que estabelece entre eles. Ainda segundo o autor, a Ciência da Informação entende o conhecimento sob um prisma objetivo, ou seja, focado naquele conhecimento intelectual, que é produto da mente humana e que é registrado em documentos (o “conhecimento objetivado”) (MOREIRA, 2018). O acesso a esse conhecimento se dá, portanto, através da representação adequada dos conceitos que o compõem em SOCs, num ciclo de atividades que integram a disciplina chamada de Representação do Conhecimento (MOREIRA, 2018). As representações são a forma pela qual o ser humano organiza o mundo, nomeando e simbolizando as entidades que o cercam e que ele é capaz de perceber; se estas

---

<sup>3</sup> Redes semânticas são “conjuntos de termos que representam conceitos, modelados como os nós em uma rede de tipos de relacionamentos variáveis.” (ZENG, 2008, tradução nossa).

representações estiverem equivocadas, a comunicação dos saberes entre os indivíduos se dará de maneira incorreta (ALMEIDA, 2020).

Sendo uma das preocupações da Ciência da Informação a representação adequada do conteúdo dos documentos (ou seja, do conhecimento objetivado), o desenvolvimento de SOCs é objeto central da área, e depende do diálogo da mesma com outras disciplinas, como a Terminologia e a Ontologia Aplicada (ALMEIDA, 2020; SCHIESSL; SHINTAKU, 2012). Da Terminologia, derivam-se as linguagens documentárias, que são vocabulários convencionados que extraem os conceitos-chaves dos documentos, traduzindo-os para linguagens que sejam compreendidas por sistemas de informação digitais, visando a captura do sentido daquilo que se está buscando representar em um termo (SCHIESSL; SHINTAKU, 2012). Da Ontologia Aplicada derivam-se os princípios que levarão à classificação (ou seja, à observação, seleção e representação das características principais das entidades) e à categorização (ou seja, à análise do estado das coisas a partir de certos critérios) das entidades que deverão ser representadas (ALMEIDA, 2020). Esse diálogo entre diferentes áreas no qual se baseia o desenvolvimento dos SOCs pode ser representado pela definição que Moreira (2018, p.102, grifo nosso) traz acerca do conceito:

De modo geral, os SOC referem-se a **estruturas terminológicas que enumeram conceitos, expressos [em] termos selecionados, bem como suas diversas relações**. Os SOC são utilizados para organizar e representar o conhecimento com vistas à sua recuperação e uso e suas funções variam conforme o grau de complexidade e a finalidade de sua aplicação.

Referente à complexidade e finalidade de cada tipo de SOC, tanto Moreira (2018) quanto Schiessl e Shintaku (2012) apontam para a norma ANSI/NISO Z39.19-2005 (R2010), que define diferentes tipos de SOCs. Desta forma, em ordem de crescente complexidade, os autores citam as listas de termos<sup>4</sup>, os anéis de sinônimos<sup>5</sup>, as taxonomias, os tesouros e as ontologias como exemplos de SOCs. Estas últimas merecem destaque particular por algumas razões.

Inicialmente, é importante apontar que seu maior grau de complexidade é devido às diferentes relações que as ontologias são capazes de representar: enquanto outros

---

<sup>4</sup> Embora use a expressão “*term lists*” como categoria para diferentes SOCs, Zeng (2008, tradução nossa) define as listas num geral como “conjuntos limitados de termos em alguma ordem sequencial”.

<sup>5</sup> Anéis de sinônimos são definidos como “conjuntos de termos que são considerados equivalentes para fins de recuperação”.(ZENG, 2008, tradução nossa).

SOCs são capazes de representar apenas relações hierárquicas ou associativas simples, as ontologias permitem todo tipo de representação de relacionamentos, pois representam todos os conceitos e propriedades em um domínio (BIAGETTI, 2021; SCHIESSL; SHINTAKU, 2012). Além disso, as ontologias são regidas por axiomas, expressões lógicas sempre verdadeiras que regem o domínio do conhecimento sendo ali representado, o que permite sua integração em sistemas de informação para propósitos de inferência automatizada (*Ibid.*, 2012). Por fim, e devido às duas características apresentadas anteriormente, as ontologias são elemento fundamental da *Web Semântica*, o atual estado de interconectividade de dados que rege a *Web* contemporânea (*Ibid.*, 2012; ALMEIDA, 2020).

Ainda cabe discorrer rapidamente sobre a distinção que existe entre as disciplinas da Organização do Conhecimento e da Representação do Conhecimento. Embora ambas estejam costumeiramente relacionadas, Moreira (2018) nos lembra que esta associação nem sempre ocorre, pois nem toda representação do conhecimento passa, ao menos conscientemente, por um processo prévio de organização. Além disso, o autor também aponta que a Representação do Conhecimento é objeto comum de análise em trabalhos na Ciência da Computação, em particular no domínio do desenvolvimento de IAs, onde seu entendimento é muito similar ao que lhe é dado pela Ciência da Informação. Essa perspectiva é corroborada por Almeida (2020), que posiciona a Representação do Conhecimento como uma disciplina própria, voltada para o desenvolvimento de IA na Computação. Com base na literatura, Moreira (2018) afirma que a Representação do Conhecimento pode ser compreendida como uma série de compromissos ontológicos, isto é, uma associação de pontos de vista consensuais, de comum acordo entre os atores responsáveis pelas atividades de ROC, acerca da realidade e de quais aspectos dela devem ser levados em consideração no momento do esforço de Representação. Esses compromissos são necessários, conforme já foi e seguirá sendo reiterado neste trabalho, pois são as representações que tornam o conhecimento acessível; a falta de entendimento comum que seja acordado entre os responsáveis pela representação implicará em um entendimento falho daquele conhecimento que está sendo representado (ALMEIDA, 2020). Nesse sentido, entende-se que as discussões acerca de qual ponto de vista tomar acerca da Organização e Representação do Conhecimento envolvem questões de Ontologia Aplicada; isto é, do estudo dos entes no mundo e de quais aspectos deles deve-se levar em conta quando do esforço de representá-los (ALMEIDA,

2020). A distinção entre a Ontologia enquanto disciplina e as ontologias enquanto artefatos representacionais é importante e será tratada na seção 4.1.3 do presente trabalho. A seguir, abordaremos as questões referentes aos metadados, à Web Semântica e ao papel que as ontologias têm em sua construção.

#### 4.1.1 Metadados e Interoperabilidade

Os metadados são elementos importantes no contexto da *Web Semântica* e, portanto, desta pesquisa. Seu papel como elementos estruturadores e enriquecedores do conhecimento em contextos digitais lhes dão papel de destaque quando se fala do desenvolvimento de sistemas inteligentes na *Web Semântica* e da interoperabilidade almejada por esse modelo, razão pela qual cabe trazer algumas definições acerca dos mesmos.

O termo *metadata* (metadados) surge na década de 1960, no contexto da Ciência da Computação, utilizado de maneira ampla para se referir a um dado que descreve ou representa características de outro dado (MAYERNIK, 2020). Uma definição mais clássica, mas que Sayão (2010) considera um tanto simplista, é a de que metadados seriam “dados sobre dados”, expressão que deixaria de lado as implicações mais importantes acerca dos usos e da importância que os metadados trazem consigo tanto na Ciência da Informação quanto na Computação contemporâneas. A expressão é também tida como sendo a mais comum por Mayernik (2020), que aponta que diversos pesquisadores trazem definições mais ou menos amplas, baseadas tanto nas exigências que levam ao uso de metadados quanto às funções que estes ocupam. Uma definição simples, mas geralmente compreensiva acerca de metadados, é aquela oferecida pela NISO<sup>6</sup>, que os descreve como “informações estruturadas que descrevem, explicam, localizam, ou de outra forma tornam mais fácil de recuperar, usar ou gerenciar um recurso informacional” (NISO, 2004, p.4, tradução nossa). O entendimento apresentado pela NISO é também o de que os metadados serão compreendidos de maneiras distintas dependendo da área do conhecimento com a qual se está lidando; na CC, serão entendidos como aquela informação compreensível por máquinas, enquanto, na CI, podem se referir a esquemas formais para a descrição de recursos, quer sejam digitais ou não (*Ibid.*, 2004). Consoante a essa distinção por área do conhecimento, Mayernik aponta

---

<sup>6</sup> National Information Standards Society (<https://www.niso.org/>)

que tanto definições literárias quanto normativas de metadados tendem a focar nos metadados sob a perspectiva do seu uso, adequado àquela especialidade (*Ibid.*, 2020).

De fato, o uso dos metadados em seu sentido amplo é tão pervasivo em diversas áreas do conhecimento que Arakaki e Arakaki (2020) apontam que os metadados são vistos ora como produto final de uma atividade científica, ora como processo da atividade em si. Há relação entre os metadados-processo e os metadados-produto, explica Mayernik (2020), já que os primeiros frequentemente surgem durante o desenvolvimento dos segundos, ou como prática constante para situações em que ainda não existem metadados-produto finalizado que possam ser utilizados. Estes metadados-produto frequentemente se manifestam na forma de padrões para a descrição de recursos informacionais, tais como linguagens de marcação como MARC, JSON e XML<sup>7</sup>, algumas das quais abordaremos ainda nesta pesquisa. Já os metadados não-estruturados, segue Meyernik (2020), seriam toda aquela documentação que, de maneira mais ou menos informal, cumprem as funções de descrição para fins de gerenciamento, preservação, etc. dos recursos informacionais; embora construída com menos rigor, essa documentação não estruturada pode trazer informações valiosas acerca de um recurso informacional, inclusive algumas que seriam difíceis ou impossíveis de descrever em uma linguagem mais formal.

Quer sejam estruturados ou não, finalizados como produto ou em construção através de processos, o ponto central na natureza dos metadados é de que eles são criados para serem usados, de maneira deliberada, pelos usuários humanos de um sistema ou por aplicações automatizadas, desde que estas sejam capazes de compreender e trabalhar com a camada extra de informação que os metadados oferecem (MAYERNIK, 2020). Considerando este ponto central em sua existência, Sayão (2010) afirma que é consenso na literatura que se dividam os metadados em três grandes categorias, a saber:

- Metadados descritivos: que trazem elementos úteis à descrição de um recurso informacional para fins de descoberta e identificação;
- Metadados estruturais: que documentam os padrões de composição (tais como ordenamento de páginas, etc.) que devem ser seguidos para que um

---

<sup>7</sup> *Machine Readable Cataloging, JavaScript Object Notation e Extensible Markup Language*, respectivamente.



recurso informacional complexo não se desorganize quando do momento de sua recomposição e uso;

- Metadados administrativos: que fornecem informações importantes para a gestão dos recursos em si, desde questões técnicas (como formato do arquivo) até legais (licenças de uso, etc.).

Outras tipologias de metadados são ainda apresentadas por Arakaki e Arakaki (2020, p.5) que, além das três categorias trazidas por Sayão, ainda apresentam as seguintes definições<sup>8</sup>:

- Metadados de autenticação: informações que possibilitam a identificação, integridade, legitimidade de um recurso informacional;
- Metadados de preservação: relacionados com informações de preservação e conservação dos recursos informacionais;
- Metadados de proveniência: estão relacionados às informações de procedência, fornecem dados sobre entidades, criação e modificações e seus relacionamentos;
- Metadados técnicos: estão relacionados a como um sistema funciona, fornecem informações do sistema ou do recurso;
- Meta-metadados (Metametadata): correspondem às informações sobre o registro criado, ou informações da criação de um conjunto de dados;
- Metadados de direitos: estão relacionados às informações sobre propriedade e direitos autorais;
- Metadados de acesso e uso: são informações de como um recurso informacional foi acessado e utilizado, como restrições de circulação e acesso, registros de exposições, entre outros;
- Linguagens de marcação (Markup languages): integram metadados e sinalizações para outros recursos estruturais ou semânticos.

---

<sup>8</sup> A definição que os autores trazem para “metadados administrativos” é um pouco diferente da de Sayão, consistindo em “[metadados] usados para gerenciar e administrar coleções e recursos informacionais, para auxiliar na tomada de decisão e na manutenção dos registros e recursos informacionais. Fornecem informações sobre a origem e a manutenção de um objeto” (ARAKI; ARAKAKI, 2020, p.5)

Essas tipologias de dados devem ser consideradas quando do momento da seleção de que padrão de metadados deve ser utilizado para a atividade que se está executando (ARARAKI; ARARAKI, 2020).

As atividades que fazem uso de metadados são das mais diversas, e a NISO traz as cinco principais:

- A descoberta de novos recursos, já que os metadados em ambientes digitais cumprem função similar à dos catálogos nas coleções analógicas, identificando os recursos e permitindo sua localização a partir da descrição de critérios importantes para sua caracterização (NISO, 2004). Este uso dos metadados é particularmente útil para agrupar recursos similares e separar recursos muito diferentes, fornecendo certo grau de personalização às atividades que um usuário pode realizar na *Web* ao aproximá-lo de recursos que possam ser do seu interesse (*Ibid.*, 2004; *Id.*,2017);
- A organização de recursos eletrônicos, por meio do uso de linguagens de marcação e metadados estruturais na construção de páginas da *Web*. Como veremos na próxima subseção desta pesquisa, a inclusão destes metadados em uma camada extra das páginas da *Web* permite a aplicações inteligentes identificar recursos *Web* distintos, o que leva à construção de enormes índices contendo informação relevante para um usuário ou grupo de usuários (NISO, 2004; *Id.*,2017);
- A identificação digital, alcançada pela atribuição de um identificador ao recurso informacional tais como um *Digital Object identifier* (DOI) ou um *Uniform Resource Identifier* (URI). Estes conjuntos de metadados servem como identificadores persistentes, que se mantêm atribuídos ao recurso informacional mesmo que seu “endereço” na *Web* seja alterado (NISO, 2004). A distinção entre URI e o “endereço” que um recurso pode ter na *Web* será tratada em mais detalhes na próxima subseção desta pesquisa;
- O arquivamento e a preservação, relacionada à atividade de identificação, já que consiste, dentre outras práticas, na atribuição de metadados que permitam o acesso a um recurso informacional mesmo que ocorram mudanças em sua hospedagem ou localização; outros metadados importantes a esta atividade são aqueles que carregam informações que permitam verificar alterações no recurso, como dados pertinentes à sua criação (NISO, 2004). A atividade de preservação

está dentre as mais deliberadas e intencionais se tratando da gestão de recursos informacionais, já que os objetos digitais sobrevivem apenas se houver esforços para se preservar os ambientes tanto digitais (*software*) quanto físicos (*hardware*) dos quais os recursos dependem, defende Sayão (2010). Nesse sentido, os metadados atuam permitindo tanto para que se tenha noção de possíveis alterações que o recurso possa ter sofrido ao longo de sua existência quanto para contextualizar informações importantes para sua compreensão e uso, que podem vir a se perder com a passagem do tempo (*Ibid.*, 2010);

- A interoperabilidade, alcançada via o enriquecimento dos recursos informacionais com metadados que permitam sua compreensão por diferentes usuários (quer sejam humanos ou não) com o mínimo de informação perdida ou significado desentendido. São os metadados que vão permitir a troca entre sistemas com estruturas de conhecimento distintas, fornecendo algum parâmetro, alguma linguagem comum para que esses sistemas possam fazer um intercâmbio de informações de maneira eficiente (NISO, 2004; *Id*, 2007).

A temática da interoperabilidade é significativa para este estudo, e por isso cabe discorrer um pouco mais acerca da mesma. A norma ISO (*International Organization for Standardization*) 25964-2 define interoperabilidade como a “capacidade de dois ou mais sistemas ou componentes de trocarem informações e utilizarem as informações que foram trocadas” (ISO, 2013, sem paginação, tradução nossa). Embora preocupações acerca do intercâmbio de informações sejam anteriores à existência da Internet, é no contexto dela e dos diversos sistemas de gerenciamento de informações e dados que surgem com ela que a interoperabilidade passa a ocupar posição de destaque em diferentes áreas do conhecimento (ZENG, 2018). De fato, a literatura observa que, desde a década de 1980, houve uma mudança de foco nos estudos sobre a interoperabilidade, partindo de problemas de sistema (ou seja, incompatibilidades de *hardware*) para problemas de sintaxe (diferenças de codificação), destes para problemas de estrutura (variação entre esquemas e modelos de dados), até chegar nos problemas de semântica mais contemporâneo (inconsistências quanto ao significado dos termos) (*Ibid.*, 2018). Neste sentido, os problemas de caráter semântico são particularmente importantes para as atividades de ROC, já que o conteúdo de um recurso informacional reflete algum conhecimento objetivo sobre a realidade, o qual é representado no dito recurso (WEISS,

2021). Finalmente, a problemática da interoperabilidade em suas quatro camadas (sistêmica, sintática, estrutural e semântica) é fundamental para se pensar o modelo da Web Semântica, do qual trataremos a seguir.

#### 4.1.2 Web Semântica

A Web Semântica é um modelo proposto pelo teórico e cientista da computação, Tim Berners-Lee, amplamente considerado como pai da *Web* como a conhecemos atualmente, para descrever um paradigma futuro da Internet, onde o conteúdo presente nas páginas da *Web* seria legível não apenas para humanos, mas também para as máquinas, permitindo a estas extraírem significado daquilo que está ali descrito e operassem de forma mais ou menos autônoma por meio de programas ou aplicativos automatizados (BERNERS-LEE; HENDLER; LASSILA, 2001). Esta camada de significado seria inserida no código-fonte de cada página da Web, através da representação do conteúdo presente naquela página na forma de recursos web, descritos formalmente através de vocabulários de especificação, as ontologias. Nesse sentido, as ontologias permitiriam aos aplicativos realizar operações de inferência, ou seja, retirar das relações entre um conceito “A” e outro conceito “B” informações que o aplicativo ainda não detinha sobre “A”, ou mesmo sobre um terceiro conceito “C” que esteja relacionado a algum dos dois primeiros; são também as ontologias que descrevem as propriedades dos conceitos e os axiomas (ou verdades lógicas) que regem o domínio do conhecimento onde “A”, “B” e “C” estão inseridos (SANTAREM SEGUNDO; CONEGLIAN, 2016), como já explicado neste trabalho. Através dessa autonomia oferecida pela contextualização dos recursos (SILVA; MARTINS; SIQUEIRA, 2018), os programas e aplicações automatizadas são capazes de obter resultados mais precisos em suas consultas, encontrando conteúdo significativo pertinente à busca mesmo que este esteja disperso em diferentes páginas, contanto que estejam conectadas e incluídas na delimitação do domínio onde está sendo realizada a busca.

Historicamente, a chamada *Web* 1.0 surgiu como um sistema que fazia uso da Internet para realizar a consulta e a atualização de documentos, conforme planejada por Tim Berners-Lee e outros pesquisadores (LAUFER, 2015). Esses documentos eram, efetivamente, páginas estáticas, conectadas uma à outra por hipertexto: nessa visualização, cada documento (ou página, que poderia conter diferentes tipos de dado)

seria um nó na grande teia (*web*) que ligava os servidores por meio da Internet (*Ibid.*, 2015). Esse sistema se baseia em três elementos básicos, de acordo com Carlos Laufer (2015): os *Uniform Resource Locator* (URLs), que servem como “endereços” dos recursos, identificando onde fica cada nó na rede; a *Hypertext Markup Language* (HTML) uma linguagem de código que permite a marcação dos documentos e, portanto, a construção das páginas web; e o *HyperText Transfer Protocol* (HTTP), o protocolo de comunicação que estabelece a comunicação entre a aplicação-cliente e o servidor onde se localiza dado nó da rede (ou seja, dada página da *Web*). Também são importantes os *links* e *hiperlinks*, ou seja, caminhos que fazem a ligação entre os diferentes nós da *Web* (*Ibid.*, 2015).

Como já explicado, a *Web* era, neste momento histórico, projetada especificamente para a compreensão por parte dos seres humanos: tratava-se de um “enorme repositório de hiperdocumentos” (PEIS *et al.*, 2003), do qual os seres humanos eram capazes de extrair sentido via a compreensão da sintaxe, da estrutura daquilo que lhes era apresentado na tela de seu computador (LAUFER, 2015). Esse cenário muda com o surgimento da *Web* Programável, no qual passa-se a considerar a necessidade de que as aplicações possam realizar operações básicas sem a condução humana, mas sim em resposta às demandas do usuário, por meio de tarefas que operam em ambientes de programação como *Hypertext Preprocessor* (PHP), *Java*, etc. (*Ibid.*, 2015). Para que essas aplicações possam executar essas tarefas, elas são desenvolvidas em três camadas, ditas de apresentação (ou seja, o navegador *web* que o usuário humano utiliza), de lógica de negócios (ou seja, a camada intermediária que executa o programa com as ordens solicitadas pelo navegador *web*) e a de armazenamento (ou seja, aquela que contém os dados a serem processados pelo programa executado e exibidos de volta para o usuário que fez a consulta no navegador) (*Ibid.*, 2015). Essa facilitação nas operações na *Web* obtida com o uso das aplicações e programas automatizados levou a um aumento na produção e circulação de dados, inaugurando a chamada *Web* 2.0, ou *Web* Social. Nesse contexto, as interações entre os usuários são mais frequentes, e o compartilhamento de conteúdo e a indexação coletiva<sup>9</sup> daquilo que está sendo produzido

---

<sup>9</sup> Sistemas coletivos de indexação, baseados em folksonomias e tagueamento, consistem em práticas que permitem aos usuários do conteúdo na *Web* gerar termos significativos, para fins de descrição categorização ou comentários dos conteúdos digitais que consomem ou produzem. Esse sistemas geralmente se baseiam no tripé “usuário - objeto informacional - termo”, conectando tanto os recursos *web* quanto os próprios usuários (RAFFERTY, 2018).

são as normas que imperam no ambiente virtual, legitimando e naturalizando a *Web* como elemento da vida cotidiana (GIL, 2014).

Apesar dos avanços em direção a uma maior autonomia das aplicações para a execução de operações na *Web*, seu ecossistema ainda é composto majoritariamente de conjuntos de dados estanques, desconexos entre si. E, com a multiplicação dos produtores de conteúdo, que potencialmente podem ser quaisquer indivíduos conectados à Internet, esses dados existem em formatos extremamente heterogêneos, o que prejudica esforços de interoperabilidade e conectividade entre aplicações de origens distintas (LAUFER, 2015). Nessa *Web* de Dados, tornou-se comum a prática de se embutir metadados durante o desenvolvimento das páginas; ou seja, era comum a inclusão de uma camada de dados acerca dos próprios dados contidos naquela página, a ser consumida pelas aplicações para que estas pudessem extrair mais sentido das páginas que estavam acessando (*Ibid.*, 2015). No entanto, a estruturação dos dados em uma página da *Web* é um processo comumente feito sem padronização, com a inclusão de metadados nos mais diversos formatos e das mais diversas origens, o que dificulta a interoperabilidade por parte das aplicações que consomem estes metadados (LAUFER, 2015). É prevendo esse cenário que Berners-Lee e seus colegas propuseram o modelo da *Web* Semântica, ou seja, um ambiente *Web* padronizado, propício à abertura dos dados e à interconectividade e interoperabilidade, alcançado a partir da integração de ontologias, agentes inteligentes e serviços semânticos (ISOTANI; BITTENCOURT, 2015). Oferecendo páginas da *Web* carregadas de semântica apropriadamente estruturada, isto é, que tragam em sua composição dados padronizados e contextualizados, enriquecidos pelo uso de metadados, e que se conectam a outros dados, os nós que compõem a *Web* se tornam cada vez mais interligados, e as aplicações são capazes de compreender e realizar operações cada vez mais complexas de maneira mais rápida e eficiente (LAUFER, 2015).

Para que esses dados abertos e conectados se tornem cada vez mais a regra no desenvolvimento de páginas da *Web*, alguns princípios de arquitetura devem ser considerados em seu processo de estruturação. Esses quatro princípios foram introduzidos pelo próprio Tim Berners-Lee, e ficaram conhecidos como *Principles of Linked Data*, de acordo com Heath e Bizer (2011). São eles:

- Utilizar *Uniform Resource Identifiers* (URIs) como nomes para as coisas;

- Utilizar URIs sob o protocolo *Hyper Text Transfer Protocol Secure* (HTTPS), que permitam que as pessoas busquem por esses nomes;
- Oferecer informações úteis quando alguém busca por um URI, através do uso de padrões como *Resource Description Framework* (RDF) e *Protocol and RDF Query Language* (SPARQL);
- Incluir *links* para outras URIs, de forma que o usuário possa encontrar mais recursos de seu interesse.

Os URIs são similares aos URLs, apresentados no começo desta subseção, mas diferenciam-se por não serem o “endereço” onde tal recurso está localizado na *Web*, e sim a própria representação do recurso em si. Esta distinção é sutil mas significativa: o uso de URIs implica na possibilidade de se representar conceitos do mundo real, incluindo conceitos abstratos, em uma linguagem compreensível por máquinas, além de apenas apontar para páginas na *Web* e conteúdo digital, como no caso dos URLs (HEATH; BIZER, 2011). O protocolo HTTPS, como já explicado também no começo desta subseção, é o caminho pelo qual um usuário da *Web* acessa os recursos da mesma, e portanto é imprescindível para que os URIs sejam acessáveis, no processo que Heath e Bizer (2011) chamam de “dereferenciação”. O terceiro princípio faz sentido quando consideramos a já mencionada heterogeneidade de dados que existem na *Web*, além da grande diversidade de aplicações e programas que precisam fazer uso dos mesmos. Nesse sentido, padrões que facilitem a representação dos dados estruturados e a consulta pelos mesmos, como no caso dos padrões RDF e SPARQL, respectivamente, são fundamentais para o desenvolvimento da *Web Semântica* (HEATH; BIZER, 2011).. Por fim, o quarto princípio defende a ampla utilização de *hyperlinks* para fazer a ligação entre os recursos na *Web*. Heath e Bizer (2011) explicam que estes *hyperlinks*, diferentes dos que existiam na *Web* até então, costumemente apresentam tipologias, de acordo com os dois recursos que estão ligando e com o vocabulário no qual se baseiam; porventura de os *hyperlinks* serem baseados nas relações entre os recursos, o *hyperlink* “é amigo de” poderia ser usado para conectar dois URIs representando duas pessoas, por exemplo.

O padrão de metadados *Resource Description Framework* (RDF) permite fazer afirmações sobre quaisquer recursos na *Web* por meio de uma estrutura em tripla: <sujeito>; <predicado>; <objeto>; essas triplas podem ter por objeto um valor literal, caso

estejam expressando propriedades de um recurso, bem como podem representar relações entre dois recursos, como exemplificado anteriormente, embora o RDF em si não seja capaz de apontar essa distinção (LAUFER, 2015). Também é comum a visualização de triplas RDF em grafos (HEATH; BIZER, 2011). Com o tempo, o padrão RDF foi expandido para o RDFS (ou RDF *Schema*), que permite especificar classes para os recursos e apontar as diferenças caso uma tripla represente uma propriedade ou relação entre dois recursos; dessa forma mais triplas, além das que estão formalmente declaradas, podem ser inferidas pelas aplicações capazes de ler e compreender o esquema (*Ibid.*, 2015). O passo seguinte ao RDFS foi a criação e aplicação de vocabulários controlados que restringissem ainda mais os conceitos e as propriedades e relações entre eles dentro de dado domínio, permitindo sua representação sem ambiguidades no ambiente *Web*. Esses vocabulários são as ontologias, que serão discutidas na seção seguinte deste trabalho; no entanto, cabe mencionar aqui a existência da *Ontology Web Language* (OWL). De acordo com Laufer (2015), a OWL é o vocabulário recomendado pelo *World Wide Web Consortium* (W3C)<sup>10</sup> para aplicação de semântica em páginas da *Web*, e consiste em uma linguagem de especificação de ontologias baseada na existência de três entidades: classes, instâncias e propriedades, pautadas por três tipos de afirmações: terminológicas (chamadas de *Tbox*, que descreve relacionamentos entre as classes), assertivas (ou *Abox*, que descreve os indivíduos, incluindo suas propriedades, relacionamentos e a que classes pertencem) e de função (*Rbox*, que traz informações acerca das propriedades, restringindo e pautando como elas funcionam e podem ser utilizadas).

Um último aspecto a tratar em relação à *Web Semântica* e à estruturação dos dados concerne aos padrões para consulta dos mesmos. Levando em conta que se trata de um modelo de abstrações acerca dos conceitos, as triplas produzidas em RDF e a partir de suas expansões necessitam de uma notação formal que permita seu entendimento por máquinas, dentre os quais se destacam notações como RDF/XML, *Turtle* e JSON-LD<sup>11</sup> (LAUFER, 2015). Uma vez traduzidas para uma dessas linguagens notacionais, é necessária uma linguagem de busca que permita fazer consultas à base de conhecimento, e a linguagem mais amplamente utilizada para esse fim é a SPARQL, que é utilizada pela maioria dos bancos de triplas (HEATH; BIZER, 2011). Laufer (2015)

---

<sup>10</sup> <https://www.w3c.br/>

<sup>11</sup> *Extensible Markup Language*, *Turtle* e *JavaScript Object Notation for Linked Data*, respectivamente.



explica que a SPARQL é uma linguagem formal cujas variáveis iniciam pela pontuação <?>, e podem assumir qualquer posição na tripla RDF (sujeito, predicado ou objeto), retornando como resultado da busca o valor presente naquela posição. Embora possa estar embutida nas aplicações que permitem a visualização das triplas ou grafos RDF, também é comum que páginas da *Web* ofereçam pontos de acesso chamados SPARQL *endpoints*, que aceitam as consultas a partir do navegador do usuário e retornam os resultados via protocolo HTTPS (*Ibid*, 2015). Como explicado no começo do parágrafo, as triplas RDF precisam ser primeiro traduzidas para uma linguagem de notação, para que no momento da consulta esta linguagem seja especificada e os resultados voltem no formato da linguagem selecionada. (*Ibid.*, 2015).

A próxima subseção irá discorrer sobre o tópico central deste trabalho: as ontologias. Como adiantado até aqui, ontologias são um artefato representacional relevante nos sistemas de organização do conhecimento contemporâneos, possuindo a função de conceptualizar e restringir universos de conhecimento específicos em uma linguagem formal (ALMEIDA, 2020), o que lhes confere extrema importância no contexto da expansão da *Web* Semântica. No entanto, a origem do termo Ontologia e das discussões acerca da compreensão e representação adequada de tudo aquilo que existe é muito anterior às Ciências da Computação e da Informação, e na verdade tem suas origens na Filosofia.

#### 4.1.3 Breve histórico da evolução do termo Ontologia

O termo *Ontologia* é, em si, relativamente recente, tendo sido criado por filósofos alemães do século 15 que uniram o prefixo grego *onto-* (ser) e o sufixo grego *-logos* (palavra); Sowa (2001) explica que a intenção desses filósofos era formalizar uma expressão que denotasse o estudo do “ser-em-si”, distinguindo este do estudo sistemático dos seres que era empreendido nas ciências naturais. Apesar de o termo ter sido criado apenas no século 15, a preocupação pelo estudo do “ser-enquanto-ser” (ou seja, o ente em suas propriedades e características únicas, independentemente do contexto no qual o ser está inserido) remonta ao filósofo grego Aristóteles, que entendia esta preocupação como parte do objeto de estudo da Metafísica (SCHIESSL; BRÄSCHER, 2012). Outro filósofo grego importante para o estudo do “ser-enquanto-ser” foi Porfírio que, estudando os escritos de Aristóteles, desenvolveu sua “árvore”, um esquema em diagrama que expõe a hierarquia entre as diferentes categorias a partir da diferenciação entre as mesmas (SOWA, 2001). Ainda segundo Sowa (2001), as

contribuições tanto de Aristóteles quanto de Porfírio para o desenvolvimento da disciplina da Ontologia (e para a Organização do Conhecimento) são utilizadas até hoje, tais como:

- O conceito de **Categoria**, ou seja, o conceito mais alto que denota tudo o que pode ser tido como propriedade, como **predicado** acerca de algum ente (por *ente* entenda-se tanto coisas materiais, quanto conceitos imateriais, quanto indivíduos, etc.);
- As noções de **Gênero** e **Diferenciação**. Diferenciação é o ato de apontar as características que distinguem os entes dentro das categorias, separando-os e dando origem a novos gêneros (ou subgêneros), cujos entes podem ainda ser diferenciados novamente, até que se chegue no ente mais específico possível, o **indivíduo**;
- A ideia de **Herança**, ou seja, que todos os entes em um subgênero possuem “suas características próprias + as características daquele gênero + as características do gênero imediatamente acima daquele”. Dito de outra forma: toda subclasse herda as características da classe ao qual está subordinada, numa relação que permite levar ao último ponto, a seguir;
- A capacidade de **Inferência** a partir de silogismos. O silogismo é uma afirmação lógica; por exemplo, se fazemos as afirmações “Todo homem é humano”, “Todo humano é mortal” e “Pedro é homem”, é possível inferir a partir do que foi afirmado que Pedro, por ser homem (e, portanto, humano), é também mortal. Os silogismos aristotélicos são divididos entre afirmativos e negativos, particulares ou universais, e foram ainda expandidos por filósofos medievais.

Dessa forma, pode-se notar que o que rege o pensar em Ontologia é a problemática das categorias e dos atributos que servem para demarcá-las. De acordo com Almeida (2020), três sistemas de categorias merecem destaque quando se trata de Ontologia, visto sua influência nas teorias contemporâneas. Inicialmente, o autor destaca o já mencionado modelo aristotélico, além do modelo kantiano e do modelo husserliano. O modelo kantiano, elaborado por Immanuel Kant (1724-1804), desacreditava na capacidade humana de captar a realidade plenamente utilizando dos sentidos, e por isso buscava determinar quais são as categorias dos sistema cognitivo humano; essas

estruturas cognitivas ditariam o conhecimento prévio (*a priori*) que os seres humanos teriam a respeito dos entes na realidade. Nesse sentido, Kant diferencia entre as proposições analíticas, na qual o sujeito já possui predicados implícitos que o caracterizam, e as proposições sintéticas, construídas a partir da apreensão das características, pautada pelas categorias internalizadas pelos sujeitos. Dessa forma, “[a] teoria de Kant lida com categorias do entendimento humano, e não com entidades da realidade como em Aristóteles” (ALMEIDA, 2020, p. 73). Já a Ontologia de Edmund Husserl (1859-1938) é baseada na fenomenologia, ou seja, na adoção de uma postura que se entende transcendental para descrever a consciência pura, e não as percepções do mundo natural (*Ibid.*, 2020). Dessa forma, estudam-se as categorias como sendo parte do reino das essências formais, ou seja, da Ontologia e da Lógica. A contribuição de Husserl para a aplicação da Ontologia seria oferecer a divisão entre Ontologia Formal e Ontologia Material (ou Regional); desta divisão surgem os termos Ontologia de Alto Nível, que se refere a ontologias genéricas, criadas para fins filosóficos, e Ontologia de Domínio, que se refere àquelas ontologias elaboradas por especialistas para a modelagem de um dado domínio do conhecimento (*Ibid.*, 2020).

Outros teóricos da Ontologia contemporânea são apresentados por Almeida (2020), tais como Hartmann, Crisholm e Johansson; em comum entre eles há o aprofundamento das bases teóricas trazidas por Aristóteles, Kant e Husserl, demonstrando a importância dos três filósofos para o desenvolvimento da Ontologia enquanto disciplina.

Enquanto disciplina filosófica, a Ontologia (e os princípios ontológicos que regem cada teoria da área) foi apropriada tanto por cientistas da informação quanto da computação, dando origem aos artefatos informacionais como o que esta pesquisa almeja desenvolver. Nesse sentido, Almeida (2020) explica que, desde a década de 1960, as pesquisas em bancos de dados deram origem a três tipos de modelos de dados: o modelo de rede, o modelo hierárquico e o modelo relacional; esses três modelos, por sua vez, seriam complementados pelos modelos semânticos da década de 1970, a partir dos quais a modelagem conceitual de domínios nos bancos de dados viria a se desenvolver. Almeida (2020) também aponta que é na década de 1960 que ocorre o primeiro uso do termo *ontologia*, pelo cientista da computação G.H. Mealy, que define uma ontologia como uma estrutura de termos formalmente organizada.

A partir dos anos 1980, torna-se amplamente aceito entre os pesquisadores da IA que, para um desenvolvimento apropriado de sistemas computacionais baseados em conhecimento, seria necessária antes a “conceptualização do domínio” para o qual aquele sistema estaria sendo desenvolvido (VICKERY, 1997). Nesse momento histórico, o que se entendia por “conhecimento” equivalia ao uso comum do termo, ou seja, ao acúmulo de saberes tidos como verdadeiros e utilizados por seres humanos para a realização de tarefas; com a necessidade da conceptualização do domínio, no entanto, essa perspectiva muda, e se torna necessário não apenas traduzir o pensamento humano para a máquina, mas sim oferecer estruturas de mundo que se assume serem verdadeiras, no contexto das operações que os sistemas devem realizar (*Ibid.* 1997; ALMEIDA, 2020). Conhecimento, a partir desse momento, passa a significar aquilo que pode ser representado em uma linguagem formal e que, ao ser tomado como verdade pelo sistema, permite que se realizem operações de maneira efetiva, racional e sistemática (SANTOS; COSTA, 2015).

Dessa forma, Currás (2010) explica que a apropriação dos princípios da Ontologia pelos teóricos da Computação surge do desejo de transformar a realidade natural de dado domínio do conhecimento humano em uma representação codificada e estruturada, legível e operável por máquinas. Adotam-se esses princípios para a modelagem conceitual, a fim de se entender o mundo de maneira estruturada, o que levou à construção de sistemas baseados na representação do domínio e na lógica, permitindo a operacionalização de inferências automáticas por parte dos sistemas (ALMEIDA, 2020). Essa apropriação dos princípios ontológicos começou a ocorrer já na década de 1980, em especial pelos estudiosos da IA, antes de se tornar amplamente aceita na década de 1990, com teóricos como Guarino e Gruber<sup>12</sup>, e em especial a partir de conferências realizadas desde o ano de 1996 (CURRÁS, 2010; ALMEIDA, 2020).

A apropriação das ontologias pela CI também se dá já na década de 1990, conforme o tema avança na Engenharia do Conhecimento, na Ciência da Computação, para a Representação e Organização do Conhecimento (ROC), e na Ciência da Informação (VICKERY, 1997). É interessante notar que a literatura aponta que, na Computação, as ontologias eram vistas com certo ar de ineditismo, o que não

---

<sup>12</sup> Nicola Guarino e Thomas Gruber são dois eminentes pesquisadores das áreas de Representação do Conhecimento e Engenharia Ontológica, contribuindo com uma produção teórica e metodológica reconhecida tanto na Ciência da Informação quanto na Computação, desde a década de 1990 (CAFÉ; SANTOS; BARROS, 2015)

correspondia totalmente com a realidade, visto que a CI já fazia uso dos princípios ontológicos da Filosofia nas atividades de organização da informação, ainda que sem o uso do termo “ontologia” em si (ALMEIDA, 2014). Temas como a Teoria da Classificação tem na Ontologia aristotélica a base de suas indagações, e mesmo problemáticas de aspecto mais técnico, como a recuperação da Informação, esbarram em questões de natureza ontológica (ALMEIDA., 2020). Vickery (1997) afirma que eram notáveis as semelhanças entre as ontologias que estavam sendo propostas e os já existentes sistemas de organização da informação, como tesouros, classificações bibliográficas e léxicos conceituais, a maioria dos quais passava despercebida pelos teóricos da Computação. Em contrapartida, houve também certa resistência por parte dos teóricos da CI em relação à adoção das ontologias, pois tanto se criticava as tentativas de reexaminar como se davam os processos de classificação quanto se questionava se o termo “ontologia” não seria apenas outro nome para práticas com as quais a CI já se ocupava há décadas, apenas aplicadas a tecnologias mais modernas (*Ibid.*, 1997; *Ibid.*, 2020). Ainda assim, a temática passou a ganhar espaço entre os acadêmicos da CI e, conforme a CI foi se aproximando das questões que surgiram com o avanço da Internet, o papel das ontologias no modelo da Web Semântica também influenciou para que se desse mais atenção a elas (*Ibid.*, 2020).

#### **4.1.4 As ontologias como artefatos de Organização e Representação do Conhecimento**

Tendo exposto como se deu a consolidação e apreensão do termo “Ontologia” pela Computação e pela CI, cabe explicar melhor como as ontologias são entendidas em ambas as áreas. Nesse sentido, são trazidas diferentes definições de teóricos das duas disciplinas.

Maurício de Almeida define as ontologias como artefatos representacionais formais, compostos por uma taxonomia de entidades e relacionamentos (ALMEIDA, 2020). Já Vickery (1997) as descreve como uma forma de organização que utiliza categorias semânticas para representar os conceitos significativos de dado domínio, incluindo definições sobre cada conceito e explicitando os relacionamentos existentes entre os conceitos. Sowa (2001) afirma que uma ontologia (enquanto produto resultante da aplicação da disciplina Ontologia) é um catálogo de tipos de coisas que supõe-se existirem em um dado domínio, a partir da perspectiva de alguém que utiliza certa linguagem para se referir a este domínio. Ding e Foo (2002) se referem às ontologias

como modelos de referência de domínios de aplicação, que possuem o propósito de melhorar a consistência e a reusabilidade da informação, além da interoperabilidade entre sistemas e o compartilhamento do conhecimento. Borst (1997 *apud* BRANDÃO; LUCENA, 2002) define uma ontologia como uma “especificação formal e explícita de uma conceitualização compartilhada”, definição que, explicam Almeida e Bax (2003), engloba os atributos de ser legível por computadores (“formal”), trazer componentes explicitamente definidos (“explícita”), trazer consigo conhecimento consensualmente validado (“compartilhada”) e modelar conceitos do modelo real (“conceitualização”). Guarino (1998 *apud* MACULAN; AGANETTE, 2017), define a ontologia como um artefato constituído por um vocabulário usado para descrever certa realidade, contendo predicados explícitos e aceitos (ou seja, que se assumem como verdadeiros) para os elementos que compõem seu vocabulário. Consonante a essa pluralidade de definições, que variam entre o teórico e o aplicado, Gruber (2007) afirma que o uso do termo denota dois significados: refere-se tanto a uma teoria de mundo estruturado quanto a um componente dos sistemas de engenharia do conhecimento.

Como artefato de *software*, as ontologias consistem em uma junção entre os modelos representativos de um domínio específico e uma linguagem lógica, resultando em sistemas automatizados, capazes de levar em conta a semântica dos modelos do domínio para a realização de operações de inferência (SOWA, 2001; ALMEIDA, 2020). Para os cientistas da Computação, em particular aqueles preocupados com o desenvolvimento de IAs, sua utilidade estaria na capacidade de permitir um tipo de raciocínio automatizado básico, baseado no compartilhamento de estruturas conceituais claramente formalizadas e da realização de inferências a partir das relações entre estes conceitos formais; a partir disto, surge a definição do artefato ontologia como uma “especificação formal e explícita de uma conceitualização compartilhada” (GRUBER, 1993b *apud* SALES, CAFÉ, 2009).

Desta forma, as ontologias seriam ferramentas para a representação do conhecimento, pois traduzem uma realidade específica, delimitada por um domínio (uma área do conhecimento, por exemplo) para uma linguagem formal e estruturada (ou seja, compreensível por máquinas), que possa ser compartilhada e reutilizada sem o risco de ambiguidades (SCHIESSL; BRÄSCHER, 2012). Como será exposto a seguir, no cerne da Ontologia enquanto disciplina estão os sistemas de categorias e seus vários processos: entender e descrever todos os entes em um dado universo, expor suas características,

diferenciá-los de acordo com o que lhes é único, ter condições de afirmar algo sobre o indivíduo mais particular de uma subclasse a partir do que é verdadeiro para sua superclasse. Embora esse processo não seja extremamente complicado para a mente humana, torna-se difícil replicá-lo para um sistema automatizado, que não possui sentidos com os quais observar o mundo real, tampouco capacidade de discernimento para apontar as diferenças naquilo que se observa, e muito menos conhecimentos *a priori* acerca de qualquer coisa. É nesta lacuna que atuam as ontologias, servindo como mecanismos de especificação parcial do conhecimento, representando apenas um determinado domínio e estipulando as leis de coerência interna que permitirão às máquinas realizar operações lógicas envolvendo os modelos representativos que compõem aquele domínio (MORAIS; AMBRÓSIO, 2007).

No contexto da Ciência Informação, costuma-se associar as ontologias a outros SOCs, como taxonomias e tesouros. Segundo Santos (2014, p.23), o artefato ontologia pode ser entendido como um Sistema de Organização do Conhecimento pois “[atua] em um domínio específico, representando seus conceitos e a relação entre eles”, e permite a visualização estruturada desse domínio. No contexto da CI, portanto, as ontologias se caracterizam como artefatos de representação de conteúdo, que visam à diminuição da ambiguidade na linguagem natural e, através de semântica formal, a facilitar a recuperação da informação (SANTOS; CORRÊA; LAPA, 2013). Para Almeida (2020), a importância das ontologias na CI é consonante com a integração cada vez maior das atividades de Organização do Conhecimento com ambientes digitais, os quais são regidos pela lógica formal e pelas linguagens estruturadas, e se relaciona a duas subáreas da CI, sendo estas a Teoria da Classificação e a Recuperação da Informação. Em relação à primeira, a aplicação dos princípios ontológicos sempre foi parte integral de atividades como a classificação, catalogação e outras que compõem a Organização do Conhecimento como um todo. Já em relação à Recuperação da Informação, o autor aponta que os avanços tecnológicos que levaram à migração do conhecimento registrado em meios físicos para o meio digital levaram à emergência das ontologias como solução para a necessidade de se traduzir vocabulários controlados em uma linguagem formal, compreensível por máquinas e utilizável no meio virtual (*Ibid*, 2020).

O estudo de Santos (2014) demonstrou como se dá a apropriação das ontologias pela CI no contexto brasileiro, através de revisão bibliográfica na Base de Dados de

Artigos de Periódicos em Ciência da Informação (BRAPCI<sup>13</sup>). A autora encontrou uma prevalência de trabalhos de caráter teórico sobre aqueles de caráter prático, além de obter outros dados acerca de como as ontologias são conceptualizadas e percebidas dentro da área: quanto à natureza conceitual das ontologias, a maioria dos autores as trata como instrumentos de conceitualização do conhecimento; quanto aos componentes que integram as ontologias, a autora observou uma predominância das relações e classes como componentes mais mencionados; finalmente, quanto ao uso, destacam-se os usos para representação do conhecimento, seguido do uso para a recuperação da informação.

#### 4.1.5 Elementos e tipologias das ontologias

Nesta seção, expõem-se algumas características das ontologias propriamente ditas, tais como os elementos que as compõem e as diferentes tipologias pelas quais podem ser agrupadas. Acerca dos elementos que as compõem, embora haja diferentes metodologias para a elaboração das ontologias, Morais e Ambrósio (2007) apontam alguns elementos básicos comuns à maioria delas:

- As classes, ou seja, os agrupamentos de conceitos que compõem o domínio com o qual a ontologia está trabalhando. Uma ontologia de culinária poderia ter em si as classes “Receita”, “Utensílio”, “Ingrediente”, etc;
- As instâncias, ou seja, os próprios entes da ontologia, exemplos particulares de cada classe. Em nosso exemplo de uma ontologia de culinária, “Sopa de legumes” poderia ser uma instância da classe “Receita”;
- As relações, que representam as interações entre os conceitos. As relações geralmente são expressas por meio de propriedades, ligando dois elementos da ontologia ou denotando um valor para algum atributo de um elemento. Em nosso exemplo, a instância “Sopa de legumes” poderia ter a propriedade “temIngrediente” conectando-a a “Cenoura”, uma instância da classe “Ingrediente”, o que indicaria que a cenoura é um dos ingredientes dessa receita;
- Os axiomas, mencionados na primeira seção desta dissertação, que são os postulados lógicos verdadeiros que regem a organização da ontologia;

---

<sup>13</sup> <https://www.brapci.inf.br/>



- As funções, que Morais e Ambrósio definem como “Eventos que podem ocorrer no contexto da ontologia”.

Já referente à tipologia que as ontologias podem assumir, tanto Morais e Ambrósio (2007) quanto Almeida e Bax (2003) expõem que a literatura oferece diferentes formas de se classificar uma ontologia. A seguir, são expostos os diferentes critérios pelos quais as ontologias podem ser tipificadas e uma explicação dessa tipificação, de acordo com os autores citados acima:

- De acordo com sua função ou estrutura, Almeida e Bax (2003) oferecem os três primeiros tipos, tratando função e estrutura como critérios distintos, enquanto Morais e Ambrósio (2007) complementam com os dois últimos tipos, ao tratarem função e estrutura como um único critério de tipificação:
  - Ontologias de domínio, que fornecem vocabulário descrevendo os conceitos em um domínio específico do conhecimento;
  - Ontologias de tarefa, que fornecem vocabulário para atividades genéricas, que podem ou não pertencer ao mesmo domínio, mas que devem poder operar entre si;
  - Ontologias gerais ou de alto nível (ALMEIDA; BAX, 2003), chamadas de “genéricas” por Morais e Ambrósio (2007), que oferecem vocabulário para descrever conceitos universais mais amplos, como tempo, espaço, causalidade, etc.;
  - Ontologias de aplicação, que descrevem conceitos pertencentes tanto a um domínio particular quanto a uma tarefa específica. Morais e Ambrósio (2007, p.7) informam que estas devem ser “[...] especializações dos termos das ontologias de domínio e de tarefa correspondentes [...]”;
  - Ontologias de representação, que descrevem e fundamentam os próprios formalismos de processos da representação do conhecimento.
- Quanto ao grau de formalismo (ou seja, quanto ao nível de complexidade da linguagem no qual a ontologia está expressa):
  - Ontologias altamente informais, expressas em linguagem natural, de maneira livre;

- Ontologias semi-informais, expressas em linguagem natural mas de maneira estruturada e restrita;
- Ontologias semi-formais, expressas em linguagem artificial estruturada e definida formalmente;
- Ontologias rigorosamente formais, expressas apenas em termos definidos por semântica formal, pautados por teoremas e provas.
- Quanto à sua aplicação, podem ser:
  - Ontologias de autoria neutra, ou seja, que servem como base para um aplicativo escrito em uma única língua que pode, posteriormente, ser convertido em diferentes sistemas, permitindo seu reuso;
  - Ontologias como especificação, baseadas em ontologias de domínio, para realizar a documentação e manutenção de um *software* relacionada ao domínio;
  - Ontologias de acesso comum à informação, utilizada para tornar compreensível um vocabulário inacessível, permitindo o compartilhamento do conhecimento por ele descrito.
- Quanto a seu conteúdo, podem ser:
  - Ontologias terminológicas, que especificam termos que serão utilizados para representar um domínio específico;
  - Ontologias de informação, que especificam estruturas de bancos de dados;
  - Ontologias de modelagem de conhecimento, que definem especificações próprias do conhecimento, refinadas de acordo com o domínio no qual estão inseridas;
  - Ontologias de aplicação, que trazem as definições necessárias para o funcionamento de uma aplicação;
  - Ontologias de domínio, que expressam as conceituações específicas para o domínio do conhecimento com o qual trabalham;
  - Ontologias genéricas, com conceituações válidas para vários domínios;
  - Ontologias de representação, que, novamente, descrevem e fundamentam os próprios formalismos de processos da representação do conhecimento.

O trabalho de Souza Júnior (2014) indica que as ontologias de domínio são o tipo mais comum de ontologias desenvolvidas em Ciência da Informação, visando a diferentes objetivos, tais como descrever e especificar conceitos ou conhecimento, descrever ou

anotar semanticamente, garantir a integração de vocabulários e a interoperabilidade semântica, etc. Biagetti (2020) nos lembra que as ontologias de domínio especificam os entes de um domínio a partir das categorias maiores fornecidas por uma ontologia fundacional ou de alto nível. A autora ainda destaca que uma diferença significativa entre as ontologias de domínio e as ontologias de alto nível (ou fundacionais) é que no caso das primeiras, o significado dos conceitos dentro do domínio já é acordado pelos especialistas daquela área, enquanto as ontologias fundacionais têm uma tarefa que transcende áreas do conhecimento, buscando negociar o significado de conceitos mais abstrato que fundamentam a realidade, como tempo, espaço, etc (*Ibid.*, 2020).

Considerando o exposto acerca das ontologias e sua natureza, cabe tratar um pouco de uma área relata do conhecimento. Trata-se da Terminologia, e das contribuições que ela pode trazer para o estudo da Ontologia Aplicada à Ciência da Informação.

#### **4.1.6 Terminologia**

A Terminologia é uma ciência interdisciplinar, normalmente definida como o “estudo e tratamento de termos técnicos e/ou-científicos” (KRIEGER, 2008, p.1). Preocupa-se, portanto, com a expressão dos saberes próprios de uma comunidade específica, predominantemente as comunidades acadêmicas e técnicas. Essa definição não é unânime, no entanto: Cabré (2005a) afirma que “Terminologia” é um termo polissêmico, detentor de três faces; por um lado, a Terminologia pode sim ser considerada como uma disciplina científica, que trata dos termos especializados; nesse aspecto, ela é considerada mais ou menos independente enquanto ciência por diferentes teóricos (DIAS, 2000). A Terminologia pode ser entendida também como prática, ou seja, como um conjunto de fazeres que visa ao estabelecimento (ou à descrição) dos termos especializados de dado campo do conhecimento humano; nesse sentido, embora seus fazeres já sejam bem estabelecidos, a Terminologia não se caracterizaria como ciência (*Ibid.*, 2000), mas sim como prática, como conjunto de diretrizes e princípios que regem a coleta e o tratamento de termos (*Ibid.*, 2005a). Por fim, a terminologia pode ser entendida como objeto, isto é, como o léxico especializado de uma área do saber, o conjunto de unidades de cognição que representam o saber especializado. Nesse último aspecto, a terminologia-objeto é produto gerado pela Terminologia-prática (*Ibid.*, 2005a). Ainda quanto à natureza teórica da Terminologia, Budin (2001) explica que alguns autores (como Wüster) defendem sua autonomia enquanto disciplina própria e independente, enquanto outros negam esse caráter de disciplina (como Sager), e ainda há um terceiro

grupo (exemplificado por Cabré) que clama por uma integração transdisciplinar para a Terminologia, neutralizando sua problemática epistemológica.

A norma ISO 1087:2019, que oferece definições para o escopo da Terminologia, a define tanto como um objeto quanto como uma ciência. No primeiro caso, usa o termo “*terminology*”, que define como um “conjunto de designações e conceitos pertencentes a um domínio ou assunto” (ISO 1087:2019, p.8, tradução nossa). Já para a Terminologia enquanto ciência (descrita como “*terminology science*” ou “*terminology studies*”), dá-se a seguinte descrição: “ciência que estuda terminologias, aspectos do trabalho terminológico, os recursos terminológicos resultantes e os dados terminológicos” (*Ibid.*, 2019, p.8, tradução nossa). Sobre o elemento do trabalho terminológico (“*terminology work*”), a norma oferece a seguinte definição: “trabalho que concerne, de maneira sistemática, à coleta, descrição, processamento e apresentação de conceitos e suas designações” (*Ibid.*, 2019, p.8, tradução nossa).

Desta forma, o que se nota é uma consonância entre a norma técnica e a literatura da área: a Terminologia é essencialmente “poliédrica”, multifacetada e detentora de diferentes fundamentos teóricos, enfoques de trabalho e práticas metodológicas de acordo com qual aspecto dela está sendo abordado (DIAS, 2000). Para fins de clareza neste trabalho, “Terminologia”, com a inicial maiúscula, refere-se à ciência terminológica, enquanto “terminologia(s)”, com inicial minúscula e possível pluralidade, refere-se aos objetos manifestados na primeira definição da ISO 1087:2019, similar ao que acontece com “Ontologia” e “ontologias”.

Dentre as correntes que existem dentro da Terminologia, enquanto teoria científica, cabe destacar aquelas de maior relevância para o desenvolvimento da área. Seguindo a ordem cronológica de desenvolvimento das mesmas, inicia-se pela Teoria Geral da Terminologia (TGT), proposta por Eugen Wüster na primeira metade do século XX. A TGT tem como dimensão fundamental da Terminologia o nível conceitual dos termos, em oposição ao nível linguístico ou lexical (SANTOS; COSTA, 2015). Junto a outros teóricos, como Lotte e Drezen, Wüster propunha uma Terminologia focada nos processos cognitivos dos especialistas em determinada área, dos quais viria o conhecimento para a construção do termo; diferente das teorias cognitivas contemporâneas, no entanto, Wüster visava à elaboração de uma teoria prescritiva, que oferecesse fundamentos para o desenvolvimento de linguagens especializadas e precisas (KRIEGER; FINATTO, 2004). Essa abordagem é fruto do processo que levou Wüster à construção de sua teoria: sendo

especialista no tema que tomava por objeto, sendo natural que, em sua visão, o correto fosse dar primazia ao conceito que existia na forma de conhecimento tácito na mente dos especialistas das demais áreas do conhecimento (CABRÉ, 2005b). A essa abordagem, que parte do conceito em direção ao termo, dá-se o nome de “onomasiológica”. Wüster defendia a separação entre o conceito em si e sua designação (o termo), e postulava que a Terminologia deveria dar primazia ao conceito; também defendia que cabia à Terminologia atuar para uma evolução controlada das linguagens especializadas na escrita por meio da normalização, e que a evolução das linguagens naturais (isto é, orais e não especializadas) caberia à Linguística, e não à Terminologia (CABRÉ, 2005b). Essa visão positivista, onde a Terminologia entraria para sanar os problemas causados pela polissemia, se reflete na separação que Wüster faz entre conceito e significado: o primeiro, de caráter paradigmático e atemporal, deveria ser representado por um único termo, dentro do universo de discurso especializado no qual está inserido, uma relação de univocidade; já o segundo, mutável de acordo com o contexto, o recorte temporal, dentre outras variáveis, seria de caráter puramente linguístico, fora da alçada da Terminologia (KRIEGER; FINATTO, 2004).

Outra teoria relevante na Terminologia é a chamada Socioterminologia, desenvolvida inicialmente no Quebec (região francófona do Canadá) e na própria França, se estabelecendo de fato como teoria no início da década de 1980, e tem como principal expoente François Gaudin (CARVALHO; FERREIRA, 2012). Gaudin era forte crítico do caráter prescritivo que Wüster propunha para a Terminologia, e defendia uma abordagem descritiva, preocupada em analisar os contextos nos quais os léxicos especializados se desenvolviam (KRIEGER; FINATTO, 2004). Em suma, a contribuição da Socioterminologia foi o resgate da dimensão linguística do trabalho terminológico; levando em consideração como se dá o uso real dos termos nos contextos especializados, traz-se o estudo da variação para o centro da discussão terminológica, em oposição à busca por um léxico artificial, construído para abarcar conceitos imutáveis, que existiriam de maneira tácita na mente dos especialistas de dado domínio do saber (*Ibid.* 2004).

Uma terceira teoria significativa na Terminologia é a Teoria Comunicativa da Terminologia (TCT), estabelecida ao longo da década de 1990 por Maria Teresa Cabré. A TCT reforçou a abordagem linguística e semasiológica à qual a Terminologia havia se voltado com os postulados da Socioterminologia. Cabré afirma, inicialmente, que a Terminologia não poderia ser vista como uma disciplina isolada da Linguística e das

outras ciências cognitivas; ao contrário, é uma “interdisciplina” (CABRÉ, 2005a), constituída por elementos advindos de outras ciências, como a própria Linguística, a Ontologia e as especialidades, além de estar intrinsecamente ligada à Documentação e, devido aos avanços tecnológicos, às tecnologias da Informação e Comunicação. Posicionar a Terminologia desta forma é significativo para que não ocorram mal-entendidos: embora se aproxime da Linguística, Cabré afirma categoricamente que a Terminologia não é a Linguística, tampouco faz parte dela: o que ocorre é que ambas as disciplinas possuem um objeto de estudo em comum (a linguagem), mas se diferem tanto em teoria quanto em prática (*Ibid.* 2005a). A grande contribuição da TCT foi oferecer uma teoria de bases linguísticas, cognitivas e pragmáticas, que dê conta de explorar não apenas as unidades terminológicas, mas também os aspectos que elas compartilham com as outras unidades lexicais não especializadas (*Ibid.*; 2005a). No lugar da Terminologia prescritiva e positivista de Wüster, o que se tem é uma Terminologia preocupada em observar os elementos do léxico em seu uso real, comunicativo, descrevendo aquelas unidades que se configuram como portadoras de conhecimento e, portanto, como unidades terminológicas, a partir das interações dos atores especializados (KRIEGER; FINATTO, 2004).

Por fim, uma quarta teoria digna de nota é a Teoria Sociocognitiva da Terminologia, desenvolvida por Rita Temmerman, que segue a aproximação que houve entre a Terminologia e as ciências cognitivas, já preconizada por Juan C. Sager, e se volta para o conceito, mas sem abandonar os avanços em direção à Linguística pelos quais a Terminologia vinha passando desde a década de 1980. Nesse sentido, Temmerman oferece uma teoria baseada nos paradigmas hermenêuticos do conhecimento especializado; sua proposta é de que a construção histórica desses paradigmas dentro das áreas do saber afeta e potencialmente altera os conceitos desenvolvidos nestas áreas (KRIEGER; FINATTO, 2004). Dessa forma, Temmerman move o foco dos conceitos estáticos de Wüster para os efeitos da linguagem e da cultura sobre os processos de compreensão e construção dos saberes. As constantes mutações pelas quais o léxico passa e a riqueza de novas unidades de interpretação dos saberes que dali surgem reforçam o papel que a linguagem e as palavras, com seu poder de mobilidade de significados, possuem na construção do conhecimento (*Ibid.* 2004).

Diante das principais teorias que marcam a Terminologia, cabe discorrer sobre como esta área interage com o desenvolvimento de artefatos ontológicos. O

desenvolvimento de ontologias envolve a delimitação do contexto a ser modelado (o domínio) e a conceitualização dos entes ali existentes. No entanto, questionamentos surgem acerca da ocorrência desta conceitualização, isto é, sem que haja consenso entre o que cada termo presente no léxico daquele domínio está denotando. Nesse sentido, o trabalho terminográfico (isto é, de operacionalização da terminologia) se mostra uma etapa significativa nas atividades de Organização do Conhecimento, e em especial nas ontologias (ROCHE, 2015). É neste tipo de SOC que a formalização e a conceitualização adequadas dos conceitos se fazem fundamentais para que tanto os sistemas automatizados quanto os usuários humanos possam obter êxito em seus processos de inferência e navegação interativa, respectivamente (*Ibid.* 2015). Desta forma, é natural que ocorram aproximações entre a Teoria da Terminologia e a Teoria da Ontologia, já que, em diferentes escalas, ambas as teorias possuem objetos de estudo similares, visando à restrição no escopo dos domínios e à representação do conhecimento especializado, seja através da descrição ou da prescrição dos léxicos especializados (AGANETTE; ALMEIDA, 2015). Aquilo que se entende por relação genérica na terminologia será uma relação de subsunção em uma ontologia, por exemplo enquanto o “conceito”, tão almejado na Terminologia, terá outras nomenclaturas de acordo com o contexto no qual se insere na ontologia (um conceito genérico será uma superclasse, ao passo que um conceito individual será entendido como um particular ou instância); essas relações de equivalência demonstram a proximidade teórica que existe entre a Terminologia e a Ontologia aplicada no desenvolvimento de artefatos de *software* (*Ibid.* 2015).

Além disso, lembra Almeida (2020), problemas da computação moderna, como o chamado problema da *Torre de Babel*<sup>14</sup> poderiam ser resolvidos por meio de uma definição clara de cada termo com o qual se está trabalhando, definição que, por vezes, se dá por meio da própria construção de uma ontologia. Embora preocupações da dimensão linguística nominalmente fujam à seara da Ontologia, que se preocupa especificamente com a natureza dos conceitos (ALMEIDA, 2020), é necessário que haja diálogo com especialistas de dado domínio, ou com *corpora* textual daquele campo, para que se faça a representação adequada dos conceitos ali existentes. Dessa forma, contemplar todas as manifestações terminológicas de um conceito permite ao engenheiro do conhecimento, desenvolvedor de ontologias, trazer para o centro de seus esforços o

---

<sup>14</sup>Ou seja, uma eventual “multiplicidade de significados, vocabulários e interpretações” que prejudica o esforço colaborativo necessário ao desenvolvimento de sistemas de informação (ALMEIDA, 2020, p.42).

próprio conceito, ilustrado em termos distintos (ROCHE, 2015). Além disso, as atividades de compartilhamento e reuso do conhecimento são profundamente multi-disciplinares, pois envolvem não apenas a classificação e a organização, mas também sua representação em linguagens compreensíveis por máquinas e de fácil interpretação pelos sistemas de informação, representação esta que não pode ser feita sem recorrer ou aos especialistas do domínio, ou aos *corpora* textuais (SANTOS; COSTA, 2015).

A partir da “virada linguística” sofrida pela Terminologia na década de 1990, graças aos esforços das teorias descritivas, passou-se a considerar as palavras em seu uso ativo como dados de grande valor para a análise terminológica: os termos do léxico denotam conceitos (especialmente quando se trata do léxico construído por comunidades especializadas), e as relações linguísticas, portanto, refletem relações conceituais, que podem ser validadas pelos especialistas do campo para dar origem a uma ontologia informal (ROCHE, 2015). No entanto, autores como Roche (2015) questionam se o enfoque dado ao uso vivo dos termos não acabaria por prejudicar os esforços de representação do conhecimento especializado, já que os sistemas conceituais extraídos do texto nem sempre correspondem às redes conceituais que seriam elaboradas em linguagem formal pelos especialistas naquele tema. Dito de outra forma, a preocupação de autores Roche está no fato de que o *sense-making* daquilo que está posto no texto especializado depende de conhecimento extralinguístico (*Ibid.* 2015). Uma vez que este entendimento acerca do domínio, este ponto de vista conceitual esteja claro e seja aceito por todos os agentes envolvidos na construção do domínio, torna-se viável a construção mesmo de ontologias multilíngues, onde a variação linguística não é um empecilho para a representação do conhecimento especializado (*Ibid.* 2015).

Nesse sentido, Roche (2015) aponta que a Terminologia se aproxima da Ontologia aplicada no fato em que se baseia em uma teoria do conceito, que irá ditar como os entes são compreendidos dentro daquele domínio, e em uma linguagem de expressão, capaz de exprimir os predicados que caracterizam os entes. O entendimento clássico acerca de uma Teoria do Conceito é aquele postulado por Dahlberg (1978a; 1978b) que define o conceito como sendo uma unidade de conhecimento que contém todos os predicados verdadeiros sobre dado objeto referente, expresso de maneira verbal; ou seja, a intensão de todos as características que um objeto possui, expressas em um termo capaz de transmitir esses atributos verdadeiros. Influenciada pela Ontologia kantiana, a autora ainda afirma existirem tanto conceitos individuais, cujos predicados verdadeiros somos



capazes compreender e expressar graças a sua existência no espaço e no tempo, perceptível por nossos sentidos, quanto conceitos gerais, mais abstratos, que não dependem tanto de nossa experiência para serem compreendidos e expressados (*Ibid.*, 1978b). Nesse sentido, o esforço de representação de um domínio do conhecimento depende, tanto da compreensão acerca dos conceitos que compõem aquele domínio, quanto da estruturação e organização dos objetos que compõem aquela realidade, culminando em sua expressão em unidades de representação (os termos) (*Ibid.*, 2015).

De acordo com Roche (2015), a compreensão do domínio depende de algumas estruturas mentais importantes, tais como:

- o conhecimento singular acerca de um único conceito (ou indivíduo dentro do domínio) e o conhecimento plural (ou seja, acerca de uma classe inteira ou um conceito hierarquicamente superior, que serve como superclasse para outros conceitos);
- as características essenciais (que definem e separam os conceitos em hierarquias simples) e as características descritivas (que denotam os atributos que compõem um conceito);
- as relações internas (essenciais à definição do conceito) e as relações externas (que associam os conceitos, em relações partitivas, causais, etc.)

O desenvolvimento dessas estruturas mentais é importante para a organização de um domínio visando sua representação, já que permite fazer inferências acerca do mesmo: características similares em conceitos distintos implicam em uma relação entre estes dois conceitos; termos que expressam conceitos hierarquicamente inferiores trazem consigo a totalidade das características do termo superior, dentre outras (DAHLBERG, 1978b). E, para que essas estruturas sirvam a seu propósito no esforço de representação do domínio, é necessário que seja alcançado consenso acerca do seu funcionamento e implicações por parte dos agentes que atuam no processo de representação. Para tanto, resgata-se a noção de compromisso ontológico dos agentes envolvidos (ALMEIDA, 2020), o já mencionado ponto de vista consensual que estes devem acordar antes de começarem a modelar dado domínio do conhecimento. É nesse sentido que, por depender de uma descrição consensual estável daqueles objetos que está modelando, as

ontologias se impõem como uma modelagem da intersubjetividade dentro de uma comunidade especializada, na visão de Roche (2015).

Santos e Costa (2015) propõem uma metodologia mista para a construção de ferramentas de representação do conhecimento, que mescle as abordagens onomasiológica e semasiológica em uma única metodologia para apreensão e representação do conhecimento. Partindo da perspectiva wüsteriana (já explorada alhures nesta subseção), as autoras tomam o conceito como elemento central do trabalho terminológico, embora façam a ressalva de que apenas à linguagem especializada cabe ser padronizada, já que a padronização envolve justamente questões relativas ao entendimento e ao conhecimento especializado, e não ao discurso ou à linguística. Nesse sentido, os termos “designam conceitos, os quais, dentro de uma prática ou domínio, constituem um sistema ou uma rede conceitual, que compõe parte do conhecimento que um indivíduo precisa dominar para entender [...] aquele campo” (*Ibid.*, 2015, p. 158, tradução nossa).

Dada essa função basilar dos termos para a construção dos domínios do conhecimento, as autoras defendem que a Terminologia deve estar presentes em todas as etapas das atividades de representação do conhecimento, mas especialmente no começo destas atividades, pois é neste momento que ocorre a análise dos *corpora* textuais, que ocorrem as consultas com os especialistas do domínio; em suma, quando estão se formando os sistemas conceituais que vão pautar a representação do domínio, mas antes de ocorrer a formalização da linguagem do artefato ontológico (SANTOS; COSTA, 2015). Naturalmente, é também nesta etapa que se buscará alcançar o compromisso ontológico entre os agentes envolvidos na representação do conhecimento, com o objetivo de se obter conceituações que minimizem as inconsistências em relação ao conhecimento que se está sendo representado (*Ibid.* 2015).

Como mencionado, Santos e Costa lançaram mão de uma abordagem mista, buscando observar se a abordagem onomasiológica ou semasiológica ofereceria melhores resultados em esforços de representação do conhecimento. Aplicando sua abordagem sobre o domínio do tratamento de água potável, as autoras constataram que, embora uma abordagem mista ofereça seus benefícios, é preciso cautela em como ela é executada: o ideal é que se realize primeiro uma análise onomasiológica, centralizada no conhecimento extralinguístico que os especialistas do domínio podem oferecer, diminuindo, já no início das atividades de representação, quaisquer ambiguidades que a

análise do *corpus* textual possa suscitar (SANTOS; COSTA, 2015). Dessa forma, o que se nota é que a interação entre Terminologia e ontologias não pode ser feita sem preparo por parte dos agentes envolvidos: depender de apenas um aspecto da atividade terminográfica pode resultar em inconsistências e falhas na representação do domínio com o qual se está trabalhando.

A seguir, é apresentada a segunda seção do referencial teórico, onde se discorre acerca do movimento pelo acesso aberto e pela abertura dos dados de pesquisa.

#### **4.2 O movimento pelo Acesso Aberto e os Dados Abertos**

Partindo do pressuposto de que é a partir do aprendizado e do uso do conhecimento prévio que os novos saberes são construídos, as antigas concepções de como se deve dar a construção dos saberes científicos, baseadas no sigilo e na confidencialidade técnica não mais servem totalmente à realidade de uma sociedade constantemente conectada, onde a informação é produzida e circula incessantemente por meio de diversas novas formas de comunicação, como por exemplo a *Web* (SILVA; SILVEIRA, 2019).

É nesse sentido que se construiu uma “filosofia aberta” da Ciência, centrada em três pilares: o uso de *softwares* livres para desenvolvimento de aplicações; a opção por arquivos (ou dados) em formato aberto, que garantam a interoperabilidade global; e a promoção do acesso aberto, ou seja, das iniciativas que lutam pela disseminação “ampla e irrestrita” dos resultados da pesquisa científica (COSTA, 2006). Na literatura contemporânea, essa filosofia aberta da Ciência é chamada simplesmente de Ciência Aberta, termo que, de acordo com Borges e Sanz Casado (2021), abrange discussões referentes a diversos processos decorrentes da ampla abertura da comunicação e da produção científicas. Dentre as discussões que a Ciência Aberta engloba, estão os Processos Abertos (ou seja, a Pesquisa Aberta Reprodutível, a Avaliação Científica Aberta, as Políticas de Ciência Aberta e as Ferramentas de Ciência Aberta), o Acesso Aberto e os Dados Abertos (SILVA; SILVEIRA, 2019). Esta seção do referencial aborda estes dois últimos itens: o Acesso Aberto e os Dados Abertos.

### 4.2.1 As iniciativas de Acesso Aberto

Inicialmente, é importante contextualizar que a necessidade por mais abertura no acesso aos frutos da Ciência advém da maneira como o financiamento da mesma se construiu historicamente. Conforme os avanços técnico-científicos passaram a ser considerados itens-chave no desenvolvimento das nações, o financiamento científico, que antes era feito pela nobreza ou pelas classes mais abastadas, passa por um processo de institucionalização, tornando-se tanto uma responsabilidade quanto um ativo de interesse para os Estados e as corporações, que assumem um papel de financiadores da Ciência (SILVA; SILVEIRA, 2019). A partir da década de 1980, no entanto, surgem problemas com essa dinâmica: o monopólio detido pelas editoras de periódicos científicos sobre o acesso aos artigos acadêmicos acabava por elevar os custos de assinaturas a preços exorbitantes, fora da realidade orçamentária das bibliotecas universitárias que desejavam acessar estes trabalhos (BARBOSA; PAVAN, 2017).

A situação foi mais ou menos contornada pela década seguinte, com a criação de repositórios de *preprints*, que reuniam (e ainda reúnem) artigos científicos que não passaram pela revisão dos pares. Dentre estes repositórios, destaca-se o estadunidense *Arxiv*<sup>15</sup>, repositório pioneiro criado por pesquisadores da área da Física no ano de 1990 (e consolidado com o nome *Arxiv* em 1998), cujo principal diferencial à época era seu funcionamento baseado no autoarquivamento pelos próprios pesquisadores que desejassem compartilhar seus estudos na plataforma, além de um sistema de gestão voluntário, não dependente de vínculos financeiros com editoras científicas (AUTRAN; BORGES, 2014; COSTA, 2006; DAMASIO, 2018). No entanto, permanecia a incongruência: embora a maior parte do financiamento para a Ciência viesse a partir do orçamento público, seus resultados ainda eram mantidos fora do alcance do cidadão comum, que não dispunha de condições para assinar um periódico científico. Além disso, a situação era desfavorável mesmo para as bibliotecas institucionais acadêmicas, cujo orçamento já não conseguia cobrir todos os custos de acesso dos periódicos tidos como mais prestigiosos.

É esta situação de inconformidade que leva diferentes grupos de acadêmicos ao redor do globo a se articularem e a proporem iniciativas com o objetivo de ampliar o acesso à sua produção científica. Proeminente dentre elas foi a *Budapest Open Access*

---

<sup>15</sup> <https://arxiv.org/about>

*Initiative* (BOAI), divulgada em 2002, que propunha duas formas de se facilitar o acesso do público em geral aos artigos científicos: o autoarquivamento e os periódicos de acesso aberto. Para concretizar a primeira proposta, a BOAI (2002) apela para que sejam criados instrumentos que auxiliem a comunidade acadêmica no depósito dos seus trabalhos em repositórios eletrônicos abertos, desenvolvidos e mantidos por instituições acadêmicas e de ensino, cujo acesso deveria ser facilitado para os usuários finais destes trabalhos. Já sobre a segunda proposta, os periódicos de acesso aberto, a Declaração apela tanto para os esforços para a criação de novos periódicos nessa modalidade, quanto para a transição planejada dos periódicos tradicionais para o formato de acesso aberto; nesse sentido, o *copyright* detido pelas grandes editoras deixaria de ser um impeditivo para o acesso à produção científica, sem que houvesse prejuízo aos responsáveis pela editoração destes periódicos (BUDAPEST..., 2002).

A influência da Declaração de Budapeste e das outras declarações que a seguiram foram sentidas pelos atores envolvidos na produção científica, e as duas propostas apresentadas na BOAI se consolidaram na prática, sendo comumente conhecidas como as “vias” para o acesso aberto: via dourada, no caso dos periódicos de acesso aberto, e via verde, no caso dos repositórios (KURAMOTO, 2014). Na maioria dos países, a construção dos repositórios abertos se deu por meio das próprias instituições superiores de ensino, e o próprio pesquisador que tenha escrito um trabalho faz o depósito do mesmo. Assim, vai se criando uma cultura voltada para o autoarquivamento e para o compartilhamento dos trabalhos já em sua origem, fortalecendo a via verde (*Ibid.* 2014). Quanto aos artigos em acesso aberto, alguns obstáculos inerentes à realidade capitalista em que a ciência é construída se fizeram presentes na consolidação da via dourada. Por exemplo, embora existam muitos periódicos científicos gratuitos, ainda há uma quantidade de revistas que cobra taxas de processamento de artigos (ou *article processing charge* (APCs), em inglês); diferente das assinaturas, essas taxas são cobradas dos pesquisadores que estão submetendo seu trabalho ao periódico, e foram o compromisso que o movimento AA e as editoras comerciais conseguiram alcançar para que houvesse uma abertura dos periódicos sem uma grande perda financeira para quem os publica (BARBOSA; PAVAN, 2017 ).

Apesar de imperfeita, a adoção do modelo AA levou a efeitos positivos na comunidade acadêmica como um todo. Um exemplo disso é a Vantagem de Citação de AA (*Open Access Citation Advantage* (OACA), em inglês), um conceito que demonstra como o Acesso Aberto favorece o aumento no número de citações dos trabalhos acadêmicos e, portanto, um acréscimo no fator de impacto e no renome dos autores

daqueles trabalhos (FURNIVAL; HUBBARD, 2011). Também a adoção da Via Verde pelas instituições de ensino levou a um aumento na visibilidade de suas pesquisas, e o próprio desenvolvimento de repositórios institucionais se tornou motivo de reconhecimento para as equipes responsáveis nas instituições (*Ibid.*, 2011; KURAMOTO, 2014).

#### 4.2.2 Dados abertos de pesquisa

Se na subseção anterior abordamos o Acesso Aberto aos artigos científicos, ou seja, ao produto-final da atividade dos cientistas, nesta subseção abordaremos o movimento pela abertura dos dados de pesquisa, os produtos-meio gerados pelo fazer científico. O movimento pelos dados abertos científicos (também chamados “dados de pesquisa” ou ainda, em inglês, *Open Data*) preconiza que determinados dados sejam disponibilizados publicamente de forma gratuita, sem restrições de *copyright*, patentes ou outros mecanismos de controle (RODRIGUES et al, 2010). Nesse sentido, é importante distinguir que o movimento trata especificamente da disponibilização dos dados científicos primários, resultados diretos da investigação científica; a abertura de dados governamentais também está relacionada ao movimento, embora envolva também outras questões à parte, já que a maior parte dos esforços de pesquisa científica são feitos com recursos governamentais (LA REFERENCIA, 2018).

Dessa forma, Krebs e Caregnato (2017) afirmam que a temática dos dados abertos de pesquisa envolve tanto os dados antes da publicação do artigo científico em periódico quanto depois da mesma; no primeiro caso, estuda-se o uso dos dados, enquanto no segundo trata-se do compartilhamento propriamente dito. Trabalhar com dados abertos de pesquisa envolve tratar questões de ética na pesquisa, já que o compartilhamento dos dados poderia levar a atitudes antiéticas, como o uso dos dados de outro pesquisador para fins de plágio ou mesmo a divulgação de dados confidenciais coletados durante um estudo (KREBS; CAREGNATO, 2017). Além disso, a abertura dos dados científicos primários envolve também questões técnicas (como o uso de padrões, a preservação e manutenção dos recursos *Web*, dentre outras) e de privacidade (como a garantia do anonimato e da randomização em estudos envolvendo sujeitos humanos) (LA REFERÊNCIA, 2018). Em última instância, a abertura destes dados envolve também questões de caráter financeiro e jurídico: Murray-Rust (2008) relata a dificuldade de se conseguir manter a propriedade dos dados complementares produzidos em sua pesquisa

frente ao interesse da editora onde a pesquisa foi publicada a fim de manter o controle sobre os mesmos.

Se durante muito tempo os dados de pesquisa foram tratados como subproduto finalizado, sem considerar seu ciclo de vida, suas iterações e linhagens, essa perspectiva passa a mudar com o avanço da Internet e a conectividade cada vez maior entre os pesquisadores de diferentes partes do globo, no contexto da chamada *e-Science* (SAYÃO; SALES, 2014). Uma declaração redigida em 2006 pela STM<sup>16</sup> e pela ALPSP<sup>17</sup>, duas eminentes associações científicas estadunidenses, expunha que os dados, em suas diversas formas, eram o verdadeiro resultado-chave dos esforços científicos, com importância como recurso comunitário para os cientistas (ALPSP; STM, 2006). Além disso, a declaração afirma que os conjuntos de dados (*datasets*) que são fruto dos esforços de pesquisa deveriam, se não permanecerem sob posse dos cientistas que os produziram, ao menos serem disponibilizados para acesso fácil para outros acadêmicos (*Ibid.* 2006). Outro fator a ser considerado em se tratando da disponibilização de dados de pesquisa é a questão do *Big Data*, ou seja, dos enormes volumes de dados que são produzidos constantemente graças aos avanços tecnológicos na *Web*, e que necessitam de serem tratados, preservados e disponibilizados para que deles se possa extrair o máximo de valor (SAYÃO; SALES, 2014). Diversos argumentos para essa disponibilização e abertura dos dados foram elencados por Murray-Rust (2008), e são apresentados a seguir:

- O fato de que os dados pertencem à raça humana como um todo, ou seja, podem ser de interesse geral da população (como os dados referentes ao genoma humano, por exemplo);
- Os dados comumente são gerados com recursos públicos e instituições governamentais, e portanto deveriam ser disponibilizados para a sociedade em geral;
- Os dados demonstram fatos, e não há como deter posse de um fato, mesmo que juridicamente;
- Os patrocinadores da pesquisa não podem tirar todo o valor dos dados se estes não forem disponibilizados publicamente;

---

<sup>16</sup> *International Association of Scientific, Technical and Medical Publishers*

<sup>17</sup> *Association of Learned and Professional Society Publishers*

- Restringir o reuso de dados cria práticas que vão contra a comunalidade da Ciência;
- Os dados são necessários para o bom-andamento de diversas atividades humanas;
- A pesquisa científica em si ocorre de maneira mais rápida havendo acesso adequado aos dados.

Além destes argumentos, a literatura entende que é a partir do reuso dos dados que seu valor aumenta exponencialmente, tanto em atividades científicas como não-acadêmicas (DOWNS, 2021). Isto se dá porque os benefícios da disponibilização vão além da simples possibilidade de se repetir experimentos para fins de verificação dos resultados; através da normalização do reuso, o fluxo da produção científica passa por transformações, dando ênfase à cooperação independente de barreiras geográficas ou disciplinares (SAYÃO; SALES, 2014).

Além disso, a disponibilização dos dados pode levar a novos esforços de expansão da pesquisa original, ao estudo de fenômenos que fugiam ao escopo da pesquisa original, se forem conduzidos por cientistas de outra área, e mesmo ao desenvolvimento de produtos e recursos que façam uso destes dados, e que não poderiam existir se não houvesse acesso a eles (DOWNS, 2021). Dessa forma, as práticas de Dados Abertos afetam positivamente todo o conjunto das atividades de pesquisa, pois o valor do dado de pesquisa passa a ser vinculado ao seu potencial de (re)uso e ao seu potencial reinterpretação em outras áreas além daquela onde ele foi gerado (SAYÃO; SALES, 2014).

Na literatura, é comum se basear na divisão proposta pela *National Science Board* para classificar os diferentes tipos de dados científicos, conforme relembram Sayão e Salles (2016): dados experimentais seriam aqueles que resultam de experimentos em ambientes controlados, como em laboratórios; dados computacionais seriam aqueles que resultam da execução de modelos computacionais que simulam condições de teste predeterminadas; finalmente, dados observacionais seriam aqueles que resultam da observação de fenômenos específicos. Os meios para a disponibilização dos pacotes de dados, sejam eles científicos ou não, se dão via a criação e manutenção de repositórios de dados, os quais necessitam de proteção jurídica e intelectual quanto à propriedade dos dados ali depositados, conforme afirma a declaração conjunta da STM com a ALSPS



(2006). Os princípios que regem a dinâmica da disponibilização são os chamados Princípios FAIR, que são apresentados a seguir, conforme expostos por Abadal (2021):

- a. FINDABLE (Encontráveis): este princípio estipula que os dados de pesquisa e os metadados que os descrevem possam ser encontrados pela comunidade após sua publicação, através de ferramentas de busca;
- b. ACCESSIBLE (Acessíveis): os dados e metadados devem ser acessíveis e podem ser baixados por outros pesquisadores utilizando seus identificadores;
- c. INTEROPERABLE (Interoperáveis): Tanto os dados como os metadados devem ser descritos seguindo as normas da comunidade acadêmica, utilizando padrões abertos, permitindo sua troca e reuso;
- d. REUSABLE (Reutilizáveis): os dados e metadados devem poder ser reutilizados por outros pesquisadores, deixando clara a sua procedência e as suas condições de utilização.

Como pode-se perceber, os princípios FAIR dialogam harmoniosamente com os quatro Princípios do *Linked Data*, apresentados na seção 4.1.2 deste trabalho. E de fato, outro conceito desenvolvido por Tim Berners-Lee em prol da *Web Semântica* se aplica no caso do compartilhamento dos dados abertos de pesquisa. Trata-se do esquema de implementação das “5 estrelas” para Dados Abertos. O esquema divide os Dados Abertos em cinco níveis, de acordo com o formato em que estão disponibilizados; ou seja, de acordo com o grau de acessibilidade e conectividade que possuem. Os cinco níveis são os seguintes, de acordo com a página do *5 Star Data* (2012):

- a. 1 estrela: os dados são disponibilizados na *Web* em qualquer formato, sob uma licença aberta. É a maneira mais simples de se publicar os dados e permite que os usuários tenham acesso a eles, mas os dados estão “presos” ao documento, sendo inacessíveis por aplicações que desejem acessá-los;
- b. 2 estrelas: os dados são disponibilizados de maneira estruturada, como num formato de planilha *Excel*, por exemplo. Aqui, os dados continuam presos ao documento, mas se apresentam de maneira estruturada, o que permite sua leitura por aplicações que consigam acessá-los;
- c. 3 estrelas: os dados são disponibilizados em formato estruturado e não-proprietário, como em formato *Comma-separated values* (CSV), por exemplo.

Esse nível é um pouco mais complexo que os anteriores, pois requer conversores para traduzir os dados do formato proprietário para o novo formato, mas em compensação não deixa o acesso aos dados condicionado ao uso de um *software* específico;

- d. 4 estrelas: utilização de URIs para identificação dos recursos. Nesse formato, os dados em si se tornam nodos no grande esquema da *Web*, mas isso requer tempo e recursos para a representação adequada dos dados e aplicação de padrões sobre os mesmos. O exemplo mais utilizado de padrão para uso neste nível são as triplas em RDF;
- e. 5 estrelas: os dados são conectados com os dados de outros usuários da *Web*, contextualizando-os na rede. Aqui há alguns riscos, como a falta de manutenção em *links* de acesso, mas é neste nível que a ligação com outros dados e a descrição por meio de metadados padronizados agregam valor adicional aos dados que estão sendo disponibilizados.

Desta forma, percebe-se como a questão dos Dados Abertos, bem como a temática da Ciência Aberta num todo, relacionam-se com a problemática maior dos dados abertos e ligados que povoam o modelo da *Web Semântica*.

#### 4.2.3 Repositórios de dados abertos no Brasil e no mundo

Como exposto anteriormente no trabalho, a disponibilização de dados de pesquisa encorajada pela comunidade científica, bem como os benefícios do reuso e da reprodutividade são largamente reconhecidos. Como vimos também, a disponibilização dos conjuntos de dados (ou *datasets*) se dá por meio de seu depósito e manutenção em repositórios.

O interesse da comunidade acadêmica pelos dados enquanto objeto-fim entrou em voga a partir do começo da década de 2000 quando alguns países preconizaram iniciativas que visavam a facilitar o acesso a esse tipo de conteúdo, bem como a criação da *Open Knowledge Foundation* (OKF)<sup>18</sup>, uma organização sem fins lucrativos que promove projetos para o fomento de boas práticas de gestão de dados, inclusive propondo soluções jurídicas para a administração deste tipo de recurso (CORRÊA, 2016). Apesar destes e outros esforços (especialmente aqueles levados a cabo por blocos

---

<sup>18</sup> <https://ok.org.br/>

plurinacionais como a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) ou a União Europeia), a comunidade científica ainda apontava uma que a infraestrutura de sistemas existente na época não satisfazia as necessidades relativas à manutenção e curadoria dos dados (RODRIGUES *et al.*, 2010; *Ibid.*, 2016). Rodrigues e seus colegas apontaram a existência de cinco cenários que correspondiam a diferentes maneiras de tratar a curadoria de dados naquele período: no primeiro cenário, mais precário, os próprios cientistas e técnicos que faziam uso dos dados eram responsáveis por sua curadoria; no segundo cenário, as próprias organizações científicas produtoras de dados arcavam com este esforço; no terceiro, as ações de curadoria partem das universidades, e por isso os repositórios guardavam uma diversidade de dados de várias áreas; no quarto cenário, a curadoria é feita por organismos oficiais, responsáveis pela gestão das atividades científicas no país; no último cenário, a curadoria era feita por comunidades *online*, incluindo repositórios informais e *wikis* (*Ibid.* 2010).

Esses repositórios tratam-se de ambientes informatizados que contribuem no processo da comunicação científica por meio da preservação e disponibilização de acesso de dados, contribuindo também para o fortalecimento da memória e da identidade das instituições ou comunidades que os administram (MONTEIRO *et al.*, 2019). Para tanto, é necessário que esses repositórios se façam conhecidos, especialmente aqueles que armazenam dados que não vêm das *Big Sciences*, ou seja, daquelas áreas que produzem um grande volume de dados graças à sua dependência do orçamento governamental (como as áreas da Saúde, da Física Aplicada, dentre outras) (SAYÃO; SALES, 2014). As conexões entre diferentes repositórios também auxiliam no aproveitamento e na validação das práticas de gestão e compartilhamento de dados, de forma que a criação de redes federadas para o compartilhamento de dados entre diferentes repositórios vem se tornando prática comum, e mesmo repositórios de áreas distintas aproximam-se cada vez mais, conforme diferentes disciplinas percebem potencialidades no uso dos mesmos conjuntos de dados (CORRÊA, 2016). Nesse sentido, desde o final da década de 2000, diferentes iniciativas surgiram para aumentar a visibilidade dos repositórios de dados, visando tanto fortalecer as relações entre eles quanto validar sua existência perante a comunidade acadêmica e a sociedade num todo. Dentre elas, se destaca a *Registry of Research Data Repositories Initiative*<sup>19</sup>, que

---

<sup>19</sup> <https://www.re3data.org>

desenvolveu e opera um diretório que reúne as descrições de repositórios digitais de dados de pesquisa no formato de uma base de dados (SAYÃO; SALES, 2016).

Os repositórios de dados abertos podem ser categorizados em cinco grupos amplos, de acordo com Correa (2016):

- a. Repositórios institucionais: são aqueles desenvolvidos por instituições de ensino e pesquisa com o propósito de armazenar, gerir e preservar os dados produzidos durante pesquisas conduzidas naquela instituição. Geralmente são hospedados e mantidos em plataformas que facilitam tanto o autoarquivamento dos conjuntos de dados (tendo sido projetados inicialmente no contexto das iniciativas de Acesso Aberto através da Via Verde) quanto sua descoberta pelos usuários. No entanto, este tipo de repositório é frequentemente visto como pouco flexível e de baixa integração a iniciativas que visam a interoperabilidade;
- b. Repositórios temáticos: são aqueles que reúnem dados de pesquisas de uma área ou disciplina específica, como repositórios das já mencionadas áreas que compõem a chamada *Big Science*. As plataformas nas quais estes repositórios são desenvolvidos frequentemente oferecem ferramentas que facilitam o reuso dos conjuntos ali armazenados, prática facilitada pelo foco disciplinar que caracteriza esta categoria; no entanto isto também dificulta a adoção de padrões e práticas que favoreçam iniciativas de interoperabilidade supradisciplinares;
- c. Repositórios editoriais: são similares aos institucionais, pois buscam preservar e aumentar a visibilidade da produção científica de fontes específicas (periódicos), frequentemente incrementando o tratamento bibliográfico feito pelos periódicos antes da publicação dos artigos. Também é comum que este tipo de repositórios reúna produção tanto de periódicos gratuitos quanto pagos;
- d. Repositórios de propósito geral: são repositórios de uso genérico, nos quais pesquisadores de quaisquer áreas podem depositar sua produção. Frequentemente, integram iniciativas de interoperabilidade em seu *design*, adequando seus arquivos a licenças abertas e os vinculando a identificadores persistentes;
- e. Repositórios próprios: são repositórios elaborados independentemente por pesquisadores individuais ou pequenos grupos, utilizando de serviços e *websites* que oferecem serviços úteis ao compartilhamento de dados (como hospedagem de arquivos, imagens, etc.). Embora deem muita flexibilidade aos pesquisadores, seu

caráter quase *ad hoc* é pouco recomendado para a preservação formal de pacotes de dados e outros produtos do esforço de pesquisa científica.

Sayão e Salles (2016) defendem uma divisão diferente para os repositórios de dados: além dos repositórios institucionais trazidos por Corrêa, os autores fazem uma separação entre repositórios disciplinares e multidisciplinares, similares aos repositórios temáticos de Corrêa, além de trazer a categoria de repositórios orientados por projetos, que seriam aqueles que reúnem a produção resultante de um projeto ou empreendimento científico específico; os autores afirmam que, dependendo do grau de tratamento, até mesmo dados não necessariamente acadêmicos, como aqueles oriundos de projetos governamentais, poderiam ser considerados matéria-prima para nova produção científica se depositados de maneira apropriada em repositórios adequados.

Dessa forma, pode-se compreender como os repositórios atuam, em diferentes contextos, para a preservação e divulgação da produção científica, incluindo dos pacotes de dados, validando a visão de que os dados podem atuar também como objeto-fim da investigação científica. Côrrea (2016) reforça que o avanço desta visão centrada também nos dados e na importância dos repositórios para seu compartilhamento e reuso vem se consolidando mesmo frente às editoras de periódicos científicos, que cada vez mais adequam suas políticas editoriais e exigem que seus autores depositem seus dados em repositórios abertos de destaque, quer sejam temáticos, editoriais ou institucionais. As vantagens que demonstram quão valiosos são os repositórios de dados foram compiladas em uma lista de benefícios e funções oferecidas pelos repositórios, desenvolvida por Sayão e Salles (2016), e que inclui itens como:

- a. o supracitado aumento na visibilidade dos dados, já que a consulta a eles é facilitada;
- b. um aumento no compartilhamento dos dados, tanto pelo papel técnico de preservação dos dados quanto pelo esforço de integração entre pesquisadores que os repositórios exercem;
- c. a creditação aos autores dos dados, que são explicitamente identificados nos metadados do *dataset*;
- d. a questão da preservação digital, que engloba questões tecnológicas, gerenciais e de padronização referentes à custódia dos pacotes de dados;

- e. o fortalecimento da memória e da transparência científicas, que podem ser validadas pela conferência dos dados armazenados no repositório;
- f. a segurança dos dados, que também está relacionada às questões técnicas das atividades de preservação;
- g. as próprias ações de disponibilização e curadoria dos pacotes de dados, que são normatizadas e se tornam padrão no tratamento dos *datasets*;
- h. o amplo leque de serviços inovadores que se abre para os pesquisadores quando os dados passam a ser gerenciados como um recurso institucional passível de tratamento, preservação e divulgação;
- i. o alto grau de reuso de dados que os repositórios permitem possibilitam a realização de novas pesquisas;
- j. por meio do uso de protocolos de interoperabilidade, ocorre a construção de redes de repositórios, guiados pelos princípios dos Dados Abertos;
- k. finalmente, as coleções de dados se tornam indicadores da qualidade da produção científica feita naquela instituição, atestando sua produtividade e seu valor acadêmico.

No Brasil, a implementação de repositórios de dados ainda é incipiente. Em seu relatório de 2018, Pavão *et al* (2018) identificaram 15 repositórios de dados nacionais; destes, quatro são de escopo internacional, mas contam com a participação de instituições brasileiras, enquanto os outros 11 atuam inteiramente no âmbito nacional. Dentre estes repositórios, nota-se uma prevalência de repositórios de dados governamentais em prol daqueles de dados científicos, embora algumas instituições de ensino superior, em sua maioria públicas, possuam repositórios próprios (PAVÃO *et al.*, 2018).

Uma rede de repositórios bastante significativa no cenário nacional e que é de particular importância para esta pesquisa é a Rede de Repositórios de Dados Científicos do Estado de São Paulo, lançada pela FAPESP. Criada em 2019, o repositório contava com dados das seis universidades públicas do estado de São Paulo, mais dados do Instituto Tecnológico de Aeronáutica (ITA) e da Embrapa Informática Agropecuária (CNPTIA/Embrapa) (ZIEGLER; 2019). A rede consiste na integração dos repositórios de cada instituição, que são geridos internamente, e conectados por um portal que permite a busca por metadados, desenvolvido pela Universidade de São Paulo (USP), que permite ao usuário a recuperação de *datasets* produzidos em pesquisas apoiadas pela FAPESP

nas instituições contempladas (ZIEGLER; 2019). Em particular, o conjunto de dados de pesquisa que servirá como objeto de estudo deste trabalho se localiza no repositório da USP, cujas normas são estabelecidas pela resolução Nº 7900/2019 da Universidade. De acordo com a resolução, a criação do repositório de dados da USP ocorreu considerando os princípios da Ciência Aberta e levando em conta tanto a realidade contemporânea referente às ações de disponibilização de dados (ou seja, a exigência tanto de agências de fomento quanto de periódicos para que os dados das pesquisas vinculadas a eles sejam disponibilizados publicamente), quanto a necessidade de preservação e aumento da visibilidade da produção científica da USP junto à população num geral (USP, 2019). Nesse contexto, a Universidade condiciona o armazenamento dos dados a dois critérios: o preenchimento adequado dos metadados necessários e a conformidade às questões éticas e legais relacionadas ao uso dos dados (USP, 2019). A página da USP<sup>20</sup> relativa às ações de gestão de dados de pesquisa também apresenta os Princípios *FAIR* para os pesquisadores e alunos interessados, além de oferecer um modelo de Plano de Gestão de Dados.

#### 4.2.4 O padrão CIDOC-CRM

No contexto deste estudo, o modelo ontológico selecionado para a representação dos dados com os quais trabalharemos (e dos quais falaremos mais na seção 5) foi o *CIDOC Conceptual Reference Model* (CIDOC-CRM), um modelo de integração de informação para o campo da herança cultural (ou seja, Arqueologia, Arte, História, dentre outras áreas). Desenvolvido desde a década de 1990 e reconhecido como um padrão ISO desde o ano de 2006, o CIDOC-CRM é fruto do trabalho do Comitê Internacional de Documentação, órgão do Conselho Internacional de Museus (ICOM), e é, na prática, uma ontologia formal que visa a facilitar a integração, o compartilhamento e a mediação de dados e informações do domínio da herança cultural (CIDOC, 2021). Elaborado por meio de uma perspectiva *bottom-up*, o CIDOC-CRM parte de conceitos bem-estabelecidos e baseados em documentação científica para representações mais genéricas e abstratas, pautado pelas necessidades das instituições de resguardo e preservação cultural; desta forma, embora ofereça classes e propriedades que podem ser consideradas *top-level*, o escopo do CIDOC-CRM é tido como sendo o conhecimento factual sobre o passado a partir da perspectiva humana documentada (*Ibid.* 2021). Tendo em vista o escopo em que

---

<sup>20</sup> <http://prp.usp.br/gestao-de-dados-cientificos/>

se o CIDOC-CRM busca operar, não se pode deixar de ressaltar que uma de suas forças é justamente o estabelecimento destes conceitos *top-level*, que permitem a integração de outros domínios no contexto que o sistema busca representar, se necessário (BIAGETTI, 2020). Essa integração facilita o diálogo entre a informação advinda de museus e instituições similares, que fazem uso do CIDOC-CRM, e a informação oferecida por instituições culturais de outras naturezas, tais como bibliotecas ou arquivos (*Ibid.* 2020)

Parte desta facilitação provém do fato que o CIDOC-CRM tem um foco maior em estabelecer estruturas semânticas de base que permitam a interoperabilidade do que em determinar o que as instituições culturais deveriam estar representando ou na terminologia que deveriam usar para fazer essa representação (CIDOC, 2021). Essas estruturas semânticas, pautadas nas necessidades da comunidade que integra o ICOM e baseadas no estabelecimento de classes de topo, permitem a representação não apenas do objeto mantido e estudado pelas instituições culturais, mas também de conceitos abstratos, eventos e outras entidades relacionadas a este objeto, considerando sua existência no tempo presente ou passado (*Ibid.* 2021).

Do ponto de vista da Organização do Conhecimento, o CIDOC-CRM tem algumas características bastante importantes. Quanto à quantidade de conceitos no modelo, a versão 7.1.1, utilizada neste trabalho, oferece quase 100 classes (chamadas genericamente de entidades) e quase 200 propriedades; a classe primordial é a *E1 CRM Entity*, que, a partir de duas propriedades-base: *P1 is identified by*, que liga uma entidade a seu nome ou identificador, e *P2 has type*, que permite representar tipologias e especificações dentro das classes (CIDOC, 2021). Em relação ao compromisso ontológico, o CIDOC-CRM adota um olhar baseado em eventos, ou seja, uma perspectiva focada não no objeto cultural em si, mas sim no evento que conecta aquele objeto ao contexto em que está inserido; a outros objetos, atores, ocorrências no tempo, etc (BIAGETTI, 2020). Nesse sentido, o CIDOC-CRM se baseia na diferenciação filosófica clássica entre endurantes (ou seja, entes persistentes no tempo, que preservam sua identidade ao longo dos eventos pelos quais passam) e perdurantes (entes que ocorrem e tem fim durante uma extensão limitada no tempo), para subdividir *E1 CRM Entity* em suas duas subclasses mais importantes: *E77 Persistent Item* e *E2 Temporal Entity*. A partir destas duas subclasses de *E1 CRM Entity*, o CIDOC-CRM consegue representar tanto aquelas entidades que se mantém estáveis ao longo dos eventos pelos quais passam (no caso de *E77 Persistent Item* e suas subclasses) quanto aquelas que são



limitadas no tempo ou que evoluem com a passagem do tempo até alcançarem seu fim (no caso de *E2 Temporal Entity* e suas subclasses) (*Ibid.* 2021). E, se tratando de um modelo de ontologia, as classes modeladas pelo CIDOC-CRM evidentemente são complementadas por uma série de propriedades que permitem destacar as diferentes interações entre “agentes, atividades e locais que se encontram em um único evento” (BIAGETTI, 2020, p.21, tradução nossa), considerando o viés adotado pelo modelo, em que eventos são vistos como extensões de tempo onde os entes se cruzam.

Estes são os principais elementos a serem considerados acerca do modelo CIDOC-CRM. Perspectivas mais gerais quanto ao uso de ontologias para representação de dados da área da Arqueologia serão apresentados na seção 5 deste documento. A seguir, trataremos brevemente dos repositórios de dados abertos, numa perspectiva tanto nacional quanto internacional.

## **5 DESCRIÇÃO DA METODOLOGIA ADOTADA E CONTEXTUALIZAÇÃO DOS DADOS DO ESTUDO**

Este estudo trata-se de uma pesquisa aplicada, de caráter experimental, visando ao desenvolvimento de um artefato de representação do conhecimento, uma ontologia de domínio. Moresi (2003) define uma pesquisa aplicada como sendo aquela que objetiva gerar conhecimentos de aplicação prática, voltada à resolução de problemas específicos, definição que contempla este trabalho, visando à elaboração de uma ontologia capaz de representar de maneira adequada os dados de pesquisa selecionados. O autor também define pesquisa experimental como sendo uma investigação empírica que se dá a partir da observação e manipulação de variáveis para fins de observação de um fenômeno (*Ibid.* 2003).

A seguir, descreveremos o conjunto de dados selecionado para a realização deste estudo, bem como o estudo que os originou.

### **5.1 Contextualização dos dados utilizados**

O pacote de dados escolhido para a realização deste trabalho acadêmico é fruto da pesquisa intitulada *Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos*, de autoria da pesquisadora paulista Victoria Arroyo Adaime, do Museu de

Arqueologia e Etnologia da USP. Trata-se da dissertação de mestrado da pesquisadora, do ano de 2021, que lhe garantiu o título de Mestre em Arqueologia pela referida instituição de ensino. O trabalho foi orientado pelos professores Vagner Carvalheiro Porto e Cintia Alfieri Gama-Rolland, e foi financiado com apoio da FAPESP e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ).

O trabalho de Adaime se baseou em uma análise contextual, própria dos estudos da Arqueologia, para explorar como as práticas funerárias evoluíram nos períodos do Reino Novo (1550-1069 AEC<sup>21</sup>) e do Terceiro Período Intermediário (1069-656 AEC), considerando as alterações político-sociais que marcaram estes momentos da historiografia do Egito Antigo. Esta análise se baseou no estudo de múltiplos relatórios de três escavações, realizadas ao longo do século XX, em três sítios arqueológicos, a saber: Gurob e Lahun, no Médio Egito e Matmar, no Alto Egito; Gurob é datado do Reino Novo, enquanto Lahun e Matmar ambos datam do Terceiro Período Intermediário (ADAIME, 2021). Particularmente, a autora analisou os amuletos funerários utilizados nos rituais de mumificação e sepultamento típicos da cultura egípcia antiga, traçando um panorama das mudanças que ocorreram na transição destes períodos históricos a partir de elementos como a tipologia, os materiais e o posicionamento de amuletos funerários encontrados em enterramentos nos sítios mencionados anteriormente.

Durante a elaboração de sua pesquisa, Adaime compilou seus dados em um banco de dados relacional<sup>22</sup>, elaborado utilizando o *software FileMaker*. Este *software* permite relacionar os amuletos ao enterramento onde foram encontrados, bem como ao sítio arqueológico onde este enterramento estava localizado. Além disso, a autora também reuniu outras características importantes dos amuletos, enterramentos e sítios que analisava, tais como morfologia e material no caso dos primeiros, tipologia e área no caso dos segundos e período no caso dos terceiros, dentre outras (*Ibid.* 2021). Durante seu estudo, a autora designou um número para controle interno para os amuletos, enterramentos e sítios; no caso dos enterramentos, ainda há um número adicional, atribuído àquele enterramento no relatório de escavação no qual ele consta.

---

<sup>21</sup> Antes da Era Comum.

<sup>22</sup> Bancos de dados relacionais são aqueles que se baseiam em matrizes de relação que ligam conceitos (chamados “tuplas”) a atributos, fazendo uso de operações matemáticas complexas, mas que ocorrem nos “bastidores” do banco de dados, para conectar tabelas distintas que compartilhem atributos em comum; este tipo de banco de dados existe desde a década de 1970, e trouxe avanços significativos para o campo da modelagem de dados (CORONEL; MORRIS; ROB, 2011).

Este banco de dados relacional deu origem a três séries de tabelas, contendo os formulários com as informações compiladas de três sítios arqueológicos, 211 enterramentos com amuletos e 757 amuletos funerários. Também há uma quarta série, trazendo as correlações mais importantes extraídas das outras três séries de tabelas, servindo como uma versão resumida ou compilada do restante das tabelas. Nossos esforços de modelagem ontológica se baseiam majoritariamente nas tabelas desta quarta série.

Adaime (2021) faz extensivo uso dos dados ao longo de sua dissertação, inclusive com esforços de classificação dos mesmos. Foi a própria autora que classificou os amuletos de acordo com sua morfologia, separando inicialmente por tipo de imagem retratada (divindades, animais, etc) e a seguir pela forma representada na imagem (sapo, cabeça humana, flor de lótus, etc). A autora também elaborou uma classificação para os tipos de enterramentos nos quais se localizavam os amuletos, criando categorias como “fosso”, “nicho” e outras, baseadas em descrições dos espaços físicos que compunham os enterramentos.

Como um todo, esses dados foram disponibilizados em formato PDF, em um documento de mais de 1350 páginas, no Repositório de Dados Abertos de Pesquisa da FAPESP<sup>23</sup>. Embora a autora tenha disponibilizado seus dados para consulta e reuso, o formato PDF não é um bom exemplo de formato de arquivo para o compartilhamento de dados visando a interoperabilidade. De fato, o formato é citado como exemplo do nível mais baixo de disponibilização em acesso aberto no modelo das “5 Estrelas” já que, embora seja fácil de armazenar e compartilhar, acaba por “prender” os dados dentro do arquivo, exigindo que sejam desenvolvidas aplicações específicas do tipo *scraping bots*<sup>24</sup> para extrair o conteúdo do mesmo (5 STAR OPEN DATA, 2012).

Nesse sentido, este conjunto de dados se mostrou propício ao esforço científico pretendido por esta pesquisa. Primeiramente, isso se dá por se tratarem de dados oriundos da pesquisa aberta, realizados por uma pesquisadora brasileira com o auxílio de instituições brasileiras (dentre outras), para fins de avaliação de uma eminente instituição de ensino superior do país. Além disso, os dados reunidos, por serem nativos da Arqueologia, se mostram um desafio prático para esforços de representação formal ou

---

<sup>23</sup> O arquivo contendo o banco de dados elaborado por Adaime pode ser encontrado no link: <http://repositorio.uspdigital.usp.br/handle/item/304>

<sup>24</sup> *Web scraping* é o nome dado à técnica de se extrair dados brutos da *Web* e armazená-los em um formato desejado para uso futuro. *Scraping bots* seriam aqueles *bots* (programas ou ferramentas capazes de serem programados para executar tarefas simples) utilizados para essa atividade (ZHAO, 2017).

semiformal, considerando as várias facetas inerentes às pesquisas nas áreas das Humanidades. Dessa forma, foi feita a escolha por este conjunto de dados para o desenvolvimento da presente dissertação de mestrado, considerando que os resultados de tal experimentação com a modelagem de dados pode vir a contribuir com o desenvolvimento da Representação do Conhecimento para a área das Humanidades e Ciências afins.

## 5.2 Descrição da metodologia utilizada

Levando em conta a grande variedade de metodologias de construção de ontologias disponíveis para uso, optou-se por buscar dentre aquelas que fossem de fácil entendimento por parte do pesquisador responsável pela elaboração da ontologia em questão. Dessa forma, opta-se pelo uso da metodologia *OntoForInfoScience*, metodologia desenvolvida a partir de uma perspectiva voltada para Ciência da Informação e que se baseia, portanto, mais nos princípios da Organização e Representação do Conhecimento que em perspectivas mais técnicas, próprias da Computação e da Engenharia de *Softwares*, como ocorre com outras metodologias (MENDONÇA, 2015). Apesar deste caráter mais amigável a pesquisadores e profissionais da CI, a *OntoForInfoScience* não é menos rigorosa que outras metodologias já consagradas e desenvolvidas por pesquisadores da Computação, e de fato as tem como referências em sua elaboração (MENDONÇA, 2015). Nesse sentido, a *OntoForInfoScience* apresenta várias características destas metodologias clássicas em suas etapas, como o emprego de ontologias de fundamentação, a definição de classes e relações e procedimentos de documentação e monitoramento de eficiência do artefato ontológico desenvolvido (MENDONÇA; SOARES, 2017). A *OntoForInfoScience* é composta por uma pré-etapa e oito etapas propriamente ditas, que são resumidas a seguir com base em Mendonça e Soares (2017):

- a. a pré-etapa de avaliação prévia, que consiste em questionar a real necessidade de se construir um novo artefato ontológico para o propósito almejado;
- b. a Etapa 1, de especificação, onde se elabora o *template* com as informações gerais do artefato a ser desenvolvido;
- c. a Etapa 2, de aquisição e extração do conhecimento, onde são selecionadas e analisadas as fontes que fornecerão os conceitos do domínio, resultando em diferentes artefatos de representação;

- d. a Etapa 3, de conceitualização, onde os conceitos do domínio são identificados a partir dos artefatos elaborados na etapa 2 e o domínio começa a ser estruturado;
- e. a Etapa 4, de fundamentação, onde seleciona-se a ontologia de fundamentação cujo compromisso ontológico mais satisfaz as necessidades do artefato sendo desenvolvido;
- f. a Etapa 5, de representação formal da ontologia, ou seja, da conversão da estrutura conceitual do domínio modelado para uma linguagem lógica formal;
- g. a Etapa 6, de avaliação, na qual ocorrem os processos de validação e verificação ontológicas do artefato;
- h. a Etapa 7, de documentação, onde se formaliza a produção do documento da ontologia, isto é, do registro definitivo de todas as etapas e processos envolvidos em sua construção;
- i. a Etapa 8, de disponibilização, na qual a ontologia desenvolvida é finalmente publicada e seu acesso é concedido ao público interessado.

A partir da construção desta ontologia e da descrição dos dados de pesquisa selecionados em forma de Dados Abertos e Ligados, espera-se aumentar a visibilidade da produção científica brasileira, bem como contribuir para o fomento de um ambiente onde a disponibilização dos dados nesse formato se torne cada vez mais usual. Sendo assim, diante do contexto apresentado, seguimos apresentando acerca das considerações obtidas a partir da aplicação da metodologia apresentada acima e dos resultados deste trabalho.

## 6 DESCRIÇÃO DOS PROCEDIMENTOS METODOLÓGICOS E APRESENTAÇÃO DOS RESULTADOS

A execução dos procedimentos metodológicos se deu por etapas, conforme elencados anteriormente. Seguindo as diretrizes propostas pela *OntoForInfoScience*, apresentamos, etapa a etapa, como se desenvolveu nossa pesquisa.

### 6.1 Etapa 0: avaliação da necessidade da ontologia

Esta pré-etapa da construção metodológica da ontologia consiste em uma avaliação acerca da real necessidade de se construir o artefato, levando em conta pontos tais como as vantagens trazidas pelo uso de ontologias e as particularidades do domínio em relação à criação deste artefato. De acordo com Mendonça, a questão básica que norteia esta etapa é a seguinte:

O projeto a ser desenvolvido e seu contexto necessitam ou demandam a construção de uma ontologia? Ou a criação de outro instrumento de representação, tal como um tesauro, seria suficiente? (MENDONÇA, 2015)

Nesse sentido, acreditamos que a construção do artefato ontológico se mostra necessária por alguns motivos. Inicialmente, consideremos as diferenças entre as ontologias e outros SOCs.

As ontologias são capazes de demonstrar as relações semânticas dentro de um domínio em um grau maior que aquele alcançado por outros SOCs, já que suas relações não são pré-definidas, como as de um tesauro, por exemplo (BIAGETTI, 2020). De fato, estudos como os de Lima e Maculan demonstram que as ontologias permitem não apenas a exposição das relações semânticas do domínio em sua totalidade, como também permitem distinguir conceitos concretos de conceitos abstratos dentro do dito domínio. Essa diferenciação, que não é requerida nem tampouco fornecida por outros SOCs tradicionais, permite não apenas a capacidade de inferência acerca dos conceitos no domínio, como também uma descrição clara dos mesmos e das suas propriedades (LIMA; MACULAN, 2017). Naturalmente, outras alternativas à construção de uma ontologia ainda seriam viáveis: a norma ISO 25964:2013, por exemplo, aumentou significativamente o escopo do que pode ser representado em um tesauro sem descaracterizá-lo como tal (ISO:2013). As novas relações normatizadas pela norma poderiam ser utilizadas para representar a relação entre os dados de pesquisa e a

pesquisa que os originou, por exemplo. No entanto, as diferenças de finalidade entre os tesouros (voltados para o controle terminológico e a recuperação da informação pelo sujeito humano) e as ontologias (voltadas para a definição consensual dos conceitos e sua representação em linguagem legível por máquinas) nos fez optar pelas segundas, considerando o objetivo da pesquisa para a representação efetiva do conhecimento contido em dados de pesquisa passíveis de tratamento especializado.

Também na própria área da Arqueologia encontramos apoio teórico no empreendimento do esforço de desenvolvimento de uma ontologia. A análise de dados arqueológicos envolve diversos elementos multidisciplinares e aspectos relacionais que poderiam ser representados em sua plenitude em outros SOCs, mas com maiores dificuldades devido à baixa flexibilidade oferecida por estes sistemas, ao menos quando comparados com as ontologias. E a representação apropriada destas facetas é necessária por dois motivos quando se trata de dados oriundos de pesquisas arqueológicas: por um lado, representar aspectos interdisciplinares apropriadamente, seja em linguagem formal ou natural, permite que outros indivíduos além dos especialistas que possuem aquele conhecimento tácito trabalhem com os dados à sua disposição (*Ibid.* 2009).

Nesse sentido, é interessante observar como diferentes olhares ontológicos podem servir para uma organização do conhecimento arqueológico mais ou menos efetiva. Signore (2009) relata que, durante a formação do Catálogo Italiano de herança cultural, o foco inicial dos esforços de organização era o próprio objeto e suas propriedades, através do uso de cartões com dados sobre aquele objeto, a ser interpretado por outro especialista da área. Este olhar marcou os primeiros esforços de computadorização da área, levando a representações dos objetos em linguagem formal, com entidades e propriedades bem definidas e uma ênfase nas relações partitivas entre os objetos (*Ibid.* 2009). Esta centralidade do objeto no compromisso ontológico foi complementada pelo uso de tesouros e taxonomias, que permitiam expressar essas relações partitivas e relacionar, ainda que de maneira simples, entidades de naturezas diferentes, como uma obra e seu autor ou local de origem, denotando relações associativas e propriedades dos objetos no catálogo, segundo Signore (2009).

No entanto, o autor explica que o uso destes SOCs não permitia a criação de árvores conceituais mais complexas, que trouxessem conceitos tão distintos a ponto de serem capazes de expressar a interdisciplinaridade inerente ao estudo da Arqueologia.

Nesse momento, ocorre uma alteração no olhar ontológico que pauta os esforços de organização do conhecimento dentro do Catálogo Italiano, levando à adoção do padrão semântico CIDOC-CRM, que tem como elemento central o *evento* e não o objeto, permitindo a representação e a inferência de informações contextuais, tais como agentes, atividades e locais que se inter cruzam naquele evento que está sendo especificado na ontologia (BIAGETTI, 2020; SIGNORE, 2009;). Desta forma, a adoção do CIDOC-CRM para a representação de artefatos em linguagem formal demonstra como as ontologias possuem flexibilidade quanto à sua capacidade de expor diferentes facetas do domínio, relacionando-as e possibilitando a inferência e a produção de novos saberes. Signore (2009) inclusive aponta que mesmo ontologias semi-formais, que não dispõem de informações completas acerca de um domínio e por isso não se dedicam a descrever esse domínio em linguagem completamente formal, ainda oferecem grande valor em iniciativas de representação do conhecimento e de integração da informação, ainda que tenham de recorrer muitas vezes a incompletudes e inconsistências que não seriam aceitas em uma ontologia formal típica.

A natureza difusa e por vezes incerta dos dados arqueológicos também é apontada como um elemento que favorece a criação de ontologias na área. Há de se considerar tanto a natureza extremamente diversa dos dados brutos das pesquisas em Arqueologia quanto o grande fator de incerteza que surge quando se faz inferências a partir destes dados, o que exige o máximo de delimitação e especificação dos conceitos dentro do domínio. Autores como Nicolucci (2020) e Cardinal (2019) apontam que há uma imprecisão intrínseca ao estudo dos dados arqueológicos, já que estes dados podem vir em diversos formatos e as fontes primárias da Arqueologia (ou seja, os sítios e artefatos) comumente sofrem alterações ou simplesmente não se apresentam em sua integralidade na maioria dos casos. Desta forma, processos de organização do conhecimento arqueológico que visem à inferência e à produção de novas conclusões e hipóteses na área devem levar em conta essa imprecisão intrínseca, fator que dificulta o desenvolvimento de outros SOC's com finalidades distintas que a das ontologias. Biagetti (2020) também aponta que a natureza potencialmente iterativa das ontologias favorece seu uso como SOC bibliográfico, servindo para a descrição de recursos como pacotes de dados, que podem sofrer alterações com o tempo, o que vai ao encontro das preocupações de Cardinal (2019) quanto às alterações nos sítios arqueológicos e às inferências por vezes imprecisas que se extraem deles.



Com isso, concluímos a avaliação prévia quanto a necessidade desta pesquisa e da construção de nosso artefato ontológico. Tendo exposto nossos argumentos a favor deste esforço acadêmico, passamos à descrição das outras etapas de nossa metodologia.

## 6.2 Etapa 1: especificação da ontologia

A etapa seguinte da *OntoForInfoScience* consiste na elaboração de um formulário básico que responda a algumas questões referentes ao propósito geral e os objetivos do modelo ontológico que se visa construir. A apresentação destas características se faz a partir do *template* (quadro 1) a seguir, sugerido por Mendonça (2015).

**Quadro 1:** especificações do modelo ontológico proposto

<b>Domínio/Esopo Geral</b>
O modelo ontológico desenvolvido nesta pesquisa visa a representação do conhecimento expresso nos dados de pesquisa coletados pela pesquisadora Victoria Arroyo Adaime em sua dissertação de mestrado, <i>Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos</i> . Mais especificamente, ela abrange as representações criadas por Adaime para os amuletos, enterramentos e sítios arqueológicos explorados em seu trabalho, descrevendo-os e classificando-os a partir da exploração de trabalhos prévios relativos a eles, que serviram como referência para a autora e que também são descritos no modelo ontológico.
<b>Propósito Geral</b>
O modelo ontológico desenvolvido nesta pesquisa tem por objetivo explorar as perspectivas para a representação de dados de pesquisa oriundos das Ciências Humanas. Em particular, é de nosso interesse avaliar se as ontologias se mostram eficientes na representação deste tipo de conhecimento, considerando a necessidade de se representar o domínio selecionado em sua totalidade, bem como de recuperar o conhecimento exposto no modelo.
<b>Classes de usuário</b>
A ontologia elaborada nesta pesquisa é pensada para a apreciação de pesquisadores tanto da Ciência da Informação quanto para interessados na temática dos dados representados. No caso dos primeiros, ela deverá ser de particular interesse aos

estudiosos que trabalham com temas alinhados ao dessa pesquisa, como as ontologias enquanto SOC e sua eficiência na recuperação do conhecimento representado.
<b>Uso pretendido</b>
O uso pretendido para este artefato ontológico é a representação adequada das facetas mais importantes do domínio com o qual se está trabalhando, considerando o objetivo geral desta pesquisa de se avaliar as ontologias como meio para a representação e recuperação efetiva do conhecimento que envolve dados abertos de pesquisa.
<b>Tipo da ontologia</b>
Segundo seu nível de conhecimento, o artefato ontológico desenvolvido se classifica como uma ontologia de domínio, considerando sua especificidade. Embora o modelo CIDOC-CRM, no qual ela se baseia, possua algumas classes e propriedades de alto nível, semelhante a uma ontologia de fundamentação, ele é comumente classificado como um artefato intermediário, algo entre as ontologias de domínio e as ontologias <i>top-level</i> fundacionais (BRUSEKER, 2019).
<b>Grau de formalidade</b>
Em relação ao seu grau de formalidade, a ontologia em desenvolvimento qualifica-se como uma ontologia semi-formal, baseada nos princípios estipulados pelo modelo CIDOC-CRM, que restringe a modelagem semântica de suas classes e propriedades com o uso de Lógica de Primeira Ordem <sup>25</sup> , no momento em que apresenta a declaração das entidades.
<b>Delimitação do escopo de cobertura</b>
- <b>Ponto de partida da ontologia:</b> entidades do mundo real, em particular entidades temporais, e suas representações informacionais. O ponto de partida é a classe <i>E1 CRM Entity</i> e algumas de suas subclasses: <i>E2 Temporal Entity</i> , <i>E52 Time-Span</i> , <i>E53 Place</i> e <i>E77 Persistent Item</i> . A partir destas quatro subclasses, novas entidades são descritas. Outras subclasses de <i>E1 CRM Entity</i> também figuram, mas com menos destaque. Ao todo, o modelo permite a descrição de artefatos culturais, os eventos que levaram a sua criação e seu estudo, bem como propriedades que os caracterizam e aos estudos desenvolvidos acerca deles.

<sup>25</sup> A Lógica de Primeira Ordem (também chamada Lógica de Quantificadores ou Lógica de Predicados) é uma linguagem lógica baseada no uso de constantes individuais (que representam sujeitos ou agentes em uma sentença) e constantes de propriedades (que representam os predicados dos sujeitos); o uso também de variáveis individuais para suas sentenças permite que LPO seja usada para exprimir um conjunto maior de situações ou circunstâncias, enquanto a existência de quantificadores a diferencia de outras linguagens lógicas (ALMEIDA, 2021).

- **Limite do domínio coberto:** o objetivo é descrever os dados coletados durante a pesquisa de Adaime (2021), considerando a multiplicidade de atores e eventos que levaram a sua criação em primeiro lugar. Desta forma, parte-se de uma perspectiva *bottom-up*, descrevendo inicialmente o amuleto, bem como algumas de suas características tipológicas, geográficas e temporais. A seguir, adota-se a perspectiva centrada em eventos, própria do CIDOC-CRM: descreve-se o evento de criação daquele amuleto, o evento da atividade da escavação que o descobriu, o evento de criação da obra que relata esta escavação, o evento da criação da pesquisa de Adaime, dentre outros. Apensos a todos estes eventos, representam-se os agentes responsáveis por sua execução. Finalmente, descrevemos as representações informacionais criadas por Adaime para sua dissertação, bem como as principais características elencadas pela autora como elementos significativos para a descrição destes dados.

- **Questões de competência:**

QC1. Em que ano ocorreu a escavação onde foi descoberto dado amuleto representado na pesquisa de Adaime (2021)?

QC2. Quem foi o pesquisador ou grupo de pesquisadores responsáveis por essa escavação?

QC3. Que forma possui dado amuleto, de acordo com a classificação criada por Adaime?

QC4. Que tipos de enterramentos contendo amuletos funerários havia nos sítios arqueológicos estudados por Adaime, de acordo com a classificação criada pela autora?

Fonte: elaborado pelo autor (2022) a partir do modelo fornecido por Mendonça (2015)

### 6.3 Etapa 2: Aquisição e Extração de conhecimento

De acordo com Mendonça (2015), é nesta etapa que o responsável pela criação do artefato ontológico retira, das fontes de referências do domínio que está representando, o conhecimento necessário para a descrição adequada daquele tema. No entanto, Mendonça reconhece que, por se tratar de uma etapa um tanto independente do restante do desenvolvimento ontológico, esta seção é por vezes ignorada pela maioria das metodologias para a construção de artefatos ontológicos, e mesmo na *OntoFoInfoScience* ela é tratada de maneira menos exaustiva, apresentando esta seção mais como uma série de recomendações do que restrições de cumprimento obrigatório.

Nesse sentido, e considerando que o artefato ontológico construído nesta pesquisa remete a um domínio extremamente específico (o conjunto de dados oriundos da pesquisa de Adaime (2021)), explorado exaustivamente em uma única fonte (a

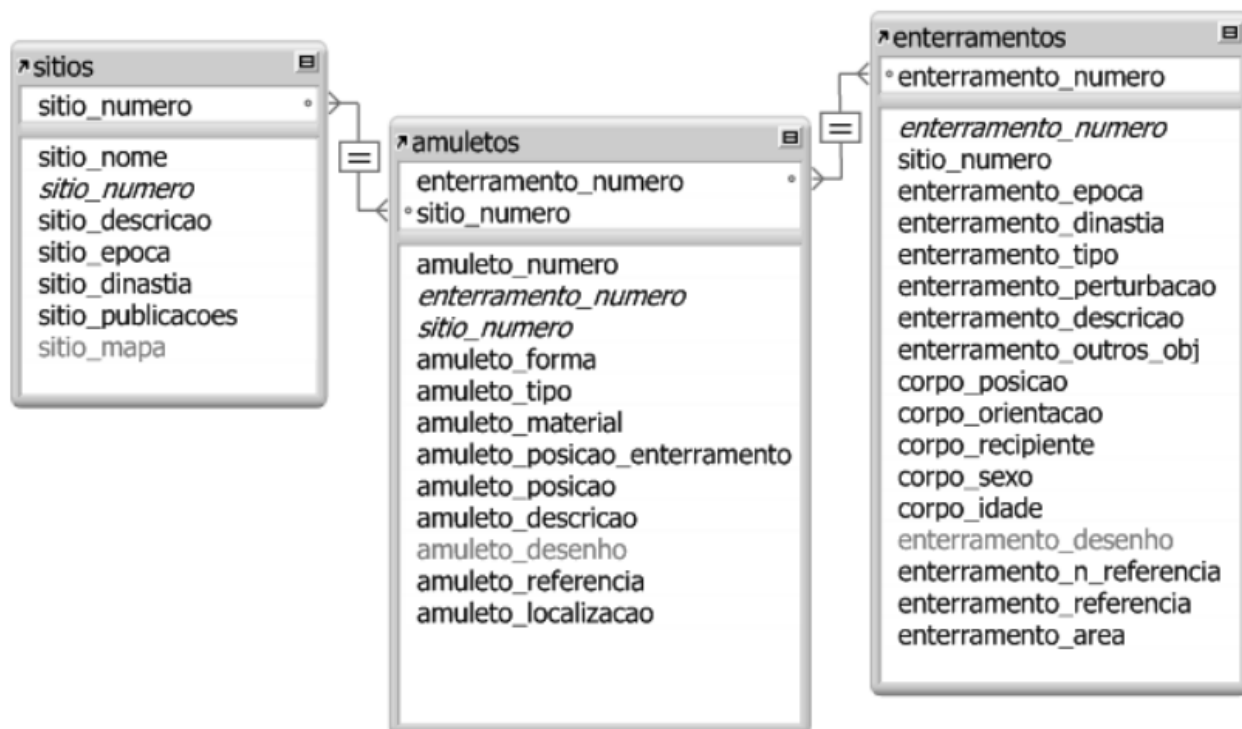
dissertação de Adaime), optamos por passar por esta etapa de maneira menos rigorosa. Ainda assim, cabe apontar algumas das recomendações de Mendonça (2015) às quais observamos durante esta pesquisa.

Quanto ao método para aquisição e extração de conhecimento selecionado, optamos pela análise informal do material de referência do domínio; que, neste caso, trata-se da própria pesquisa de Adaime (2021). A análise informal foi escolhida pois consideramos que, já havendo um tratamento dos principais elementos que compunham o domínio por parte da autora, a partir da criação das tabelas presentes no banco de dados (ADAIME, 2021), a análise formal, utilizando *softwares*, fazia-se desnecessária. A seleção do material de referência do domínio também ocorreu sem maiores complicações, considerando o escopo do artefato construído; o trabalho de Adaime (2021) serviu como base. Já os relatórios de pesquisa que serviram de base para a dissertação de Adaime não chegaram a figurar como referencial nesta pesquisa, já que não consideramos haver necessidade de pormenorizar a natureza das escavações ou dos elementos relacionados a elas e representados por Adaime (sítios, enterramentos, amuletos). Finalmente, Mendonça (2015) aponta que os produtos desta etapa seriam três, a saber: o glossário de conceitos, o glossário de verbos e o glossário de relações, que ajudariam a estipular que elementos textuais encontrados em obras de referência do domínio se tornariam classes e propriedades da ontologia. Em nosso caso, considerando que tomamos a própria dissertação de Adaime (2021) como obra central de referência, o conjunto de relações estipulados pela autora durante a criação de seu banco de dados serviu à guisa daqueles três glossários propostos pela *OntoForInfoScience*. Dessa forma, os elementos que compõem o banco de dados relacional de Adaime (2021) são reproduzidos a seguir<sup>26</sup>:

---

<sup>26</sup> Esse tipo de diagrama é comumente utilizado para auxiliar na expressão de bancos de dados relacionais, permitindo observar as entidades dentro do banco de dados, seus atributos e as relações entre elas; o modelo também é particularmente útil pois facilita visualizar que, em um banco de dados relacional, as tabelas são independentes entre si, mas conectam-se através dos relacionamentos entre seus atributos (CORONEL; MORRIS; ROB, 2011).

Figura 1: diagrama relacional do banco de dados de Adaime (2021)



Fonte: Adaime, 2021.

A representação diagramática produzida por Adaime (2021), somada à leitura de seu trabalho, ajudou a contextualizar que aspectos do conhecimento com o qual trabalhamos eram interessantes de serem modelados em nossa ontologia. Em particular, um aspecto importante que incluímos em nossa representação, e que figura no diagrama de Adaime, é a representação de que relatório de escavação lido pela autora durante a elaboração de sua dissertação continha as informações de dado enterramento e, portanto, de dado amuleto. Além disso, também buscamos representar os cientistas responsáveis pela criação de dado relatório de escavação, bem como a escavação arqueológica em si. A razão para estas representações ficará mais clara na terceira e quarta etapas da metodologia, descritas a seguir.

#### 6.4 Etapas 3 e 4: Conceituação e Fundamentação Ontológica

Considerando que a obra de referência que está sendo utilizada para a construção desta ontologia já traz consigo escolhas acerca de quais entidades e propriedades devem constar no domínio, partimos diretamente em busca de uma ontologia que nos permitisse fundamentar o desenvolvimento deste trabalho, adequando os conceitos apresentados

por Adaime ao que esta ontologia de fundamentação nos possibilitasse. Devido a isso, as Etapas 3 e 4 da *OntoForInfoScience* de Conceituação e Fundamentação Ontológica, respectivamente, se desenvolveram de forma consideravelmente aplicada, razão pela qual são apresentadas juntas nesta seção.

Inicialmente, cabe apontar a ressalva que Mendonça (2015) faz ao afirmar que, em um projeto de desenvolvimento ontológico de menor porte, a Etapa 4 poderia não ser necessária, já que implica no comprometimento ontológico com dada perspectiva filosófica de entendimento do mundo e dos domínios. Essa afirmação foi ao encontro de nossa experiência, visto que nosso domínio de modelagem, além de pequeno, se deriva de esforços científicos da Arqueologia. Dito de outra forma, trata-se da modelagem de elementos “secundários” (os dados compilados e tratados por Adaime (2021)), além da representação menos exaustivas dos elementos primários da pesquisa em Arqueologia (os próprios artefatos, sítios, etc). Dado este cenário, escolhemos fazer uso do modelo CIDOC-CRM como ontologia de fundamentação para nossa pesquisa.

O CIDOC-CRM tem por escopo teórico todo o “conhecimento factual acerca do passado, em uma escala humana” (CIDOC, 2021, p. 10, tradução nossa). Nesse sentido, e considerando seu objetivo de servir como modelo para a integração e intercâmbio tanto de documentação científica quanto de evidências materiais acerca do passado, o CIDOC-CRM utiliza de duas definições claras para definir sua função e suas propriedades, a saber: a noção de “realidade material” e de “unidades de descrição”. A realidade material consiste, dentro da proposta do CIDOC-CRM, em tudo aquilo que, possuindo substância, é perceptível com o uso dos sentidos ou o uso de instrumentos, incluindo quaisquer representações que a mente humana faça dos elementos desta realidade que é capaz de perceber (*Ibid.* 2021). A fim de discriminar as diferentes partes que compõem a realidade perceptível (ou seja, o universo possível dentro do modelo CIDOC-CRM), os seres humanos fazem uso do que o CIDOC-CRM chama de “unidades de descrição” ou, utilizando do jargão filosófico, “particulares”, ou seja, entidades concretas que exemplificam partes da realidade (*Ibid.* 2021; ALMEIDA, 2020).

Desta forma, o que o CIDOC-CRM faz é propor classes e propriedades que permitem categorizar particulares, partes da realidade, em instâncias concisas e inambíguas, que permitam a transmissão do conhecimento factual e documentado, (CIDOC, 2021). No entanto, o CIDOC-CRM não adota necessariamente uma perspectiva materialista, ignorando a existência de uma possível existência sem substância física ou a

ocorrência de vieses cognitivos; o que ocorre é que, a fim de cumprir seu objetivo, o modelo se atém àquela realidade material, conforme definida no parágrafo anterior, assegurando a representação e visando à integração da informação independente das crenças de seu observador ou de especulações sem fundamento material (*Ibid.* 2021). Cabe ressaltar que há debates em andamento, entre os responsáveis pela organização do CIDOC-CRM, acerca de qual escola filosófica o modelo se encaixaria melhor: embora seja um debate corrente, a ênfase que o modelo dá à “realidade material”, documentada segundo o método científico, o CIDOC-CRM se qualificaria como um modelo pautado no Realismo Científico, embora subscrever a essa escola filosófica apresente alguns problemas, visto que ela é embasada no conhecimento categórico oferecido pelas Ciências da Natureza e por isso menos receptiva às evidências produzidas pelas ciências que lidam com o passado humano como um todo (DOERR, 2021).

Outro elemento importante a se considerar quando se fala do compromisso ontológico do CIDOC-CRM é o de que seu entendimento é o de que o conteúdo de uma base de conhecimento que descreva dado domínio não é o domínio em si, mas sim declarações que representam aquele domínio e que podem ser interpretadas por sujeitos com capacidade de compreender a relação que existe entre as declarações e os entes declarados. Para que esta relação possa vir a ocorrer, o modelo presta especial atenção à entidade que liga a representação do elemento da realidade dentro da base de conhecimento ao elemento em si: seu nome, representado pela classe *E41 Appellation* (CIDOC, 2021). O CIDOC-CRM entende que os nomes equivalem a representações discretas dos particulares que compõem a realidade, podendo identificá-los independente do contexto em que estão inseridos; no entanto, isso não quer dizer que os nomes equivalham aos particulares em si, apenas permitindo que sejam feitas declarações acerca deles, como explicado anteriormente (CIDOC, 2021).

O último ponto a ser considerado acerca do modelo CIDOC-CRM enquanto ontologia já dá pistas de como foi conduzida a Etapa 3 da metodologia, de Conceitualização. A principal distinção que conduz o processo de representação no modelo (seção 4.2.3) é a diferenciação entre endurantes, representados pela classe *E77 Persistent Item*, e perdurantes, representados pela classe *E2 Temporal Entity*. A classe *E2*, particularmente, serve como eixo para a perspectiva *event-based* que o CIDOC-CRM possui: através dela é possível representar entidades que evoluem ao longo do tempo até atingirem seu fim, enquanto suas subclasses permitem descrever a relação entre objetos,

lugares e pessoas (instâncias indiretas da *E77 Persistent Item*) com eventos e períodos no tempo, incluindo os mais diversos tipos de atividades. É nessa dinâmica, relacionando diversas instâncias de perdurantes com diversas instâncias de endurantes, que o CIDOC-CRM representa a realidade, e foi a partir dela que se deu a conceitualização de nossa ontologia.

É nessa etapa, como descreve Mendonça (2015), que o conhecimento a ser representado pela ontologia é organizado e estruturado, e na qual os conceitos pertinentes ao domínio são tratados para poderem figurar de maneira lógica no artefato desenvolvido. A metodologia *OntoForInfoScience* leva em conta dois aspectos ao longo do decorrer desta etapa, a saber: “i) a utilização dos recursos da etapa anterior de aquisição e extração do conhecimento; e (ii) a importância do envolvimento dos especialistas na conceitualização do domínio” (MENDONÇA, 2015, p.198). Nesse sentido, acreditamos que nossa pesquisa se atém consideravelmente a estes dois pontos, visto que utilizamos aquilo que servia como recurso produzido na Etapa 2, que se tratava da própria pesquisa elaborada por Adaime (2021); ou seja, os recursos à nossa disposição eram aqueles produzidos por um especialista do domínio com qual estávamos trabalhando.

Antes de começarmos a descrever as decisões e processos de desenvolvimento metodológico desta etapa, é importante já apresentar o produto final de sua execução: o mapa conceitual de nossa ontologia, elaborado utilizando a ferramenta *CMapTools*<sup>27</sup>. Uma versão completa do mapa consta no Apêndice A deste trabalho, tendo em vista as dificuldades de reproduzi-lo em sua íntegra junto ao corpo do texto. Recortes do mapa figurarão ao longo desta seção; nesse sentido, cabe esclarecer um pouco como eles devem ser lidos.

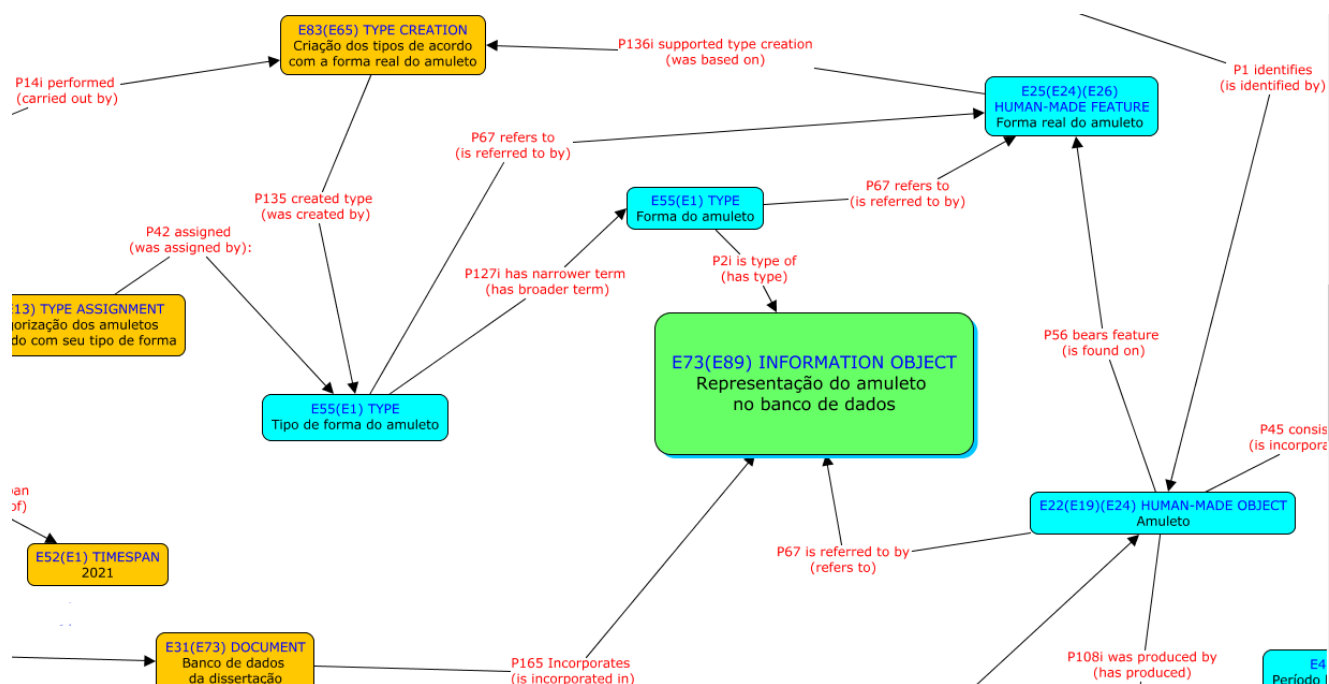
Inicialmente, é importante destacar qual o objeto central do mapa conceitual, ou seja, o referente ao redor do qual giram as relações de maior relevância para o objetivo a ser alcançado pela elaboração do mapa. Em nosso caso, o objetivo era representar adequadamente os processos de construção de conhecimento que levaram à representação dos amuletos, enterramentos e sítios arqueológicos no banco de dados elaborado por Adaime (2021). Considerando a relevância dos amuletos dentro do trabalho da autora, foi tomada a decisão de se ter como referente a representação dos amuletos no banco de dados, representada no mapa conforme a Figura 2.

---

<sup>27</sup> Disponível em <https://cmap.ihmc.us/cmaptools/>.



**Figura 2:** recorte do mapa conceitual destacando o nodo “Representação do amuleto no banco de dados”



Fonte: elaborado pelo autor (2023)

O recorte acima ilustra algumas das decisões de visualização que tomamos durante a elaboração do mapa. Nós em azul representam classes do CIDOC-CRM que possuem mais de um valor possível; ou seja, apresentam mais de uma instância dentro do modelo proposto e que, portanto, deram origem a novas subclasses dentro de nosso modelo. Já os nós em amarelo representam as instâncias em si das classes CIDOC-CRM que possuem um único valor dentro do modelo; dito de outra forma, os nós em amarelo representam o próprio indivíduo, enquanto os nós em azul representam as classes que podem ter mais uma instância e que, a grosso modo<sup>28</sup>, deram origem a novas subclasses não-nativas do modelo CIDOC-CRM<sup>29</sup>. As classes oriundas do CIDOC-CRM possuem, entre parênteses, as superclasses das quais se derivam diretamente. Entre os nós, as relações que os ligam estão representadas em vermelho, conectando o domínio das propriedades com seu objeto (*range*) com o uso de uma seta.

<sup>28</sup> A exceção é a classe *E57(E55) MATERIAL* (que não aparece no recorte acima apresentado), cuja declaração junto ao CIDOC-CRM desencoraja a criação de instâncias específicas e subclasses. Neste caso, os materiais dos amuletos constam como instâncias próprias da classe.

<sup>29</sup> Por exemplo, o nodo que representa a “Forma real do amuleto”, na Figura 2, se refere à classe *MP9 Forma real do amuleto*, uma das classes elaboradas durante a construção desta pesquisa, que é subclasse de *E25(E24)(E26) HUMAN-MADE FEATURE*, uma classe nativa do modelo CIDOC-CRM.

Em muitos casos, a instância representada nos nodos azuis se trata de uma subclasse da classe maior ali representada com o título em azul-escuro. A relação entre as classes originárias do CIDOC-CRM (designadas pelo prefixo *Ex*) e as classes criadas para o modelo proposto nesta dissertação (designadas pelo prefixo *MPx*, abreviação para “Modelo Proposto”) pode ser observada nos apêndices B e C deste trabalho, que contêm tanto a descrição das classes utilizadas quanto de suas instâncias, sejam estas indivíduos em si (valores únicos) ou subclasses, bem como sua definição<sup>30</sup>. No caso das classes originárias do CIDOC-CRM, as definições são apresentadas conforme constam no documento de Declaração de Classes e Propriedades do CIDOC-CRM versão 7.1.1., enquanto as classes criadas para o modelo proposto apresentam definições elaboradas nesta pesquisa. A maioria dos elementos representados no mapa (e, portanto, no arquivo da ontologia) possui uma definição<sup>31</sup> junto aos Apêndices B e C. As exceções incluem aquelas instâncias auto-descritivas, ou cuja classe à qual pertencem já contém em sua descrição informação suficiente para a compreensão das instâncias.

Finalmente, o Apêndice D deste documento corresponde ao Dicionário de verbos, elemento componente da Etapa 3 da metodologia *OntoForInfoScience*, que descreve as propriedades utilizadas para relacionar os indivíduos representados. O referido apêndice contém uma descrição das propriedades, as classes que podem servir de domínio e objeto, sua quantificação dentro do modelo CIDOC-CRM e suas propriedades (transitividade, reflexividade, etc.), bem como se foi utilizada a propriedade em sua configuração direta conforme o modelo CIDOC-CRM ou sua configuração inversa<sup>32</sup>. Ainda, deve-se mencionar que não foram criadas novas propriedades além daquelas que constam no modelo CIDOC-CRM.

Apresentados os elementos que compõem o mapa conceitual, podemos descrever os passos que compuseram a elaboração desta Etapa 3 da metodologia. O primeiro passo para a execução desta etapa, de acordo com Mendonça (2015), é a construção de um dicionário de conceitos, elaborado a partir dos recursos da Etapa 2, e que faz a

---

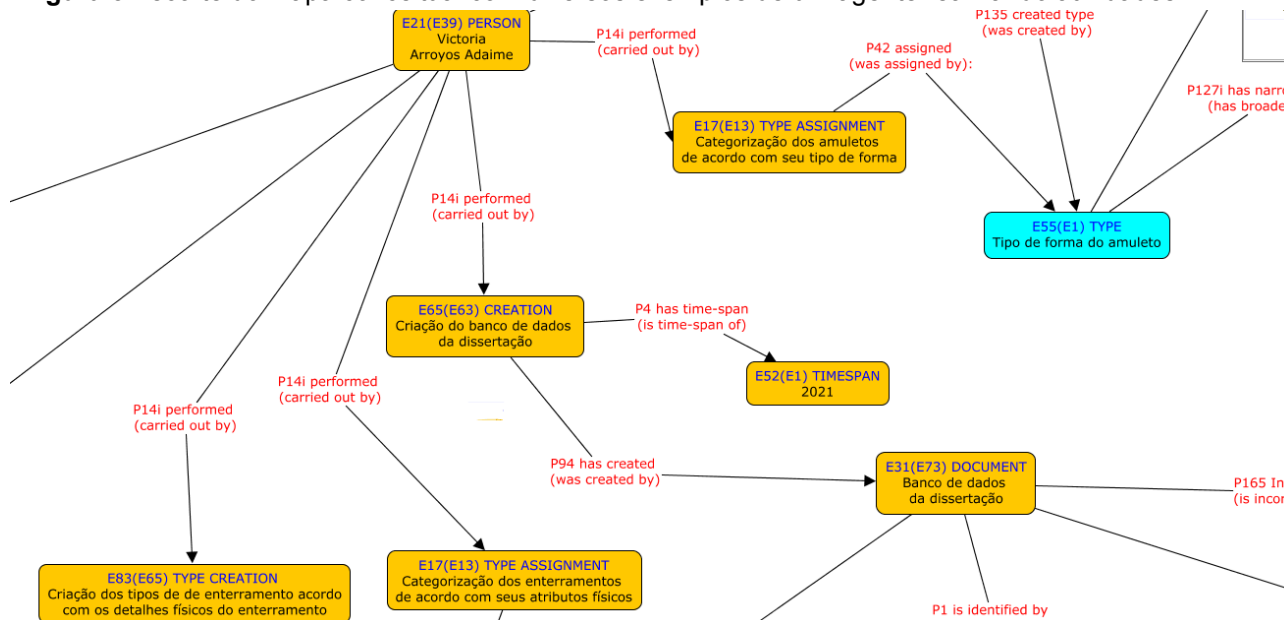
<sup>30</sup> Desta forma, os referidos apêndices fazem as vezes dos Dicionários de conceitos e da Tabela de valores possíveis, elementos que compõem a Etapa 3 da metodologia *OntoForInfoScience*.

<sup>31</sup> Estas definições foram elaboradas tendo como fontes principais de referência tanto o trabalho de Adaime (2021) quanto o *The British Museum Dictionary of Ancient Egypt* (SHAW; NICHOLSON, 2003).

<sup>32</sup> A representação das propriedades na ordem inversa pode ocorrer porque o CIDOC-CRM, diferente de outros modelos ontológicos, não considera o inverso de suas propriedades declaradas como sendo novas propriedades em si, mas apenas como configurações possíveis, interpretações implícitas a qualquer relação entre dois entes representados (CIDOC, 2021). No mapa conceitual, bem como no arquivo da ontologia em si, as propriedades em sua configuração inversa se apresentam no formato PXXi, conforme exposto a seguir com a propriedade *P14i performed*.

relação entre o conceito do domínio e sua definição textual. Devido à natureza centrada em eventos do CIDOC-CRM, muitos dos conceitos que figuram em nosso mapa conceitual são representações textuais de eventos que fizeram parte da representação dos objetos estudados por Adaime (2021) durante sua dissertação, e que figuram no banco de dados da monografia da autora. Na Figura 3, abaixo, é possível observar que uma instância da classe *E21(E39) PERSON*, neste caso a representação da pesquisadora Victoria Arroyos Adaime, foi responsável por realizar diversos eventos que, vistos em conjunto, fizeram parte do processo de construção de conhecimento que culminou na sua dissertação de mestrado e no banco de dados vinculado a ela. A realização destas ações é representada pela propriedade *P14i performed*, inversa da propriedade direta *P14 carried out by*, que liga Adaime aos eventos das atividades realizadas por ela.

**Figura 3:** recorte do mapa conceitual com diversos exemplos de um agente realizando atividades



Fonte: elaborado pelo autor (2023)

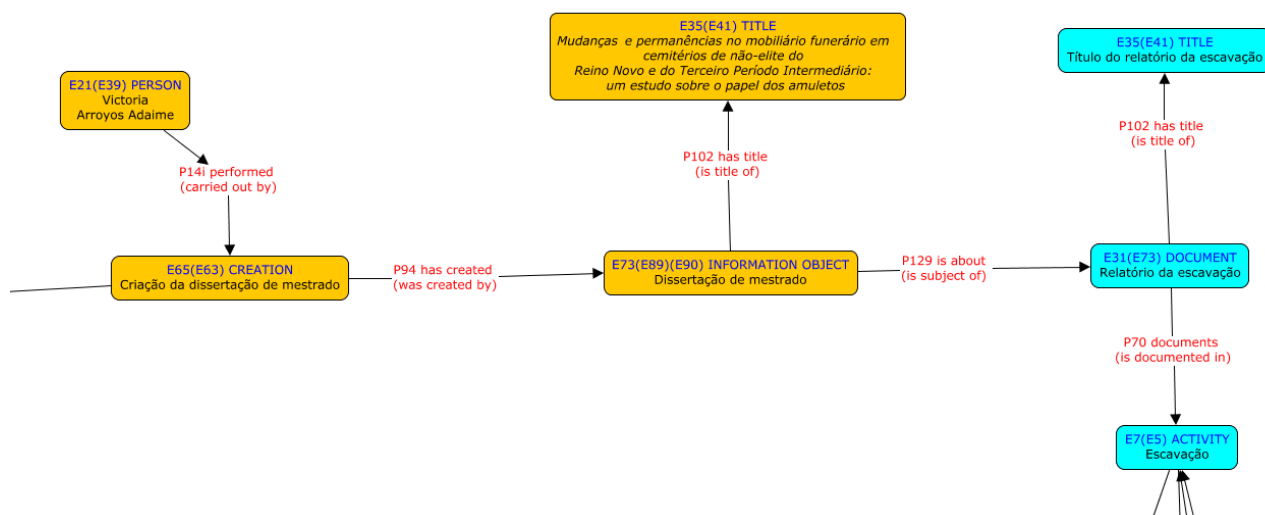
Essa necessidade de se representar por meio de classes aquilo que são eventos é o cerne de sua abordagem *event-based* do CIDOC-CRM, e por isso nosso modelo de representação possui diversos indivíduos que representam não entidades concisas e permanentes do mundo (ou *endurantes*), mas sim períodos, atividades e outros eventos que se desenrolam ao longo do tempo (os *perdurantes*). No modelo CIDOC-CRM, conforme já explicado, os endurantes são instâncias e subclasses da classe *E77 Persistent Item*, enquanto os perdurantes são as instâncias e subclasses da classe *E2 Temporal Entity*. Tanto *E77 Persistent Item* quanto *E2 Temporal Entity* são classes

fundamentais do modelo CIDOC-CRM e são, junto de algumas outras, instâncias da classe primordial *E1 CRM Entity*.

Dando continuidade à descrição das decisões tomadas na etapa de Conceitualização, é importante destacar que o representado corresponde apenas a um recorte do todo que compõe a dissertação de mestrado de Adaime e seu banco de dados. Embora este fator influencie de maneira mais perceptível a Etapa 5, de formalização do modelo proposto no arquivo da ontologia, ele também deve ser destacado enquanto apresentamos o mapa conceitual e a representação do modelo presente nele.

Essa opção se reflete na relação explicitada na Figura 4, abaixo, que expõe a relação entre os relatórios de escavação que serviram para o desenvolvimento do trabalho de Adaime:

**Figura 4:** recorte adaptado do mapa conceitual, com a representação do processo de criação de referencial na dissertação de mestrado de Adaime (2021)



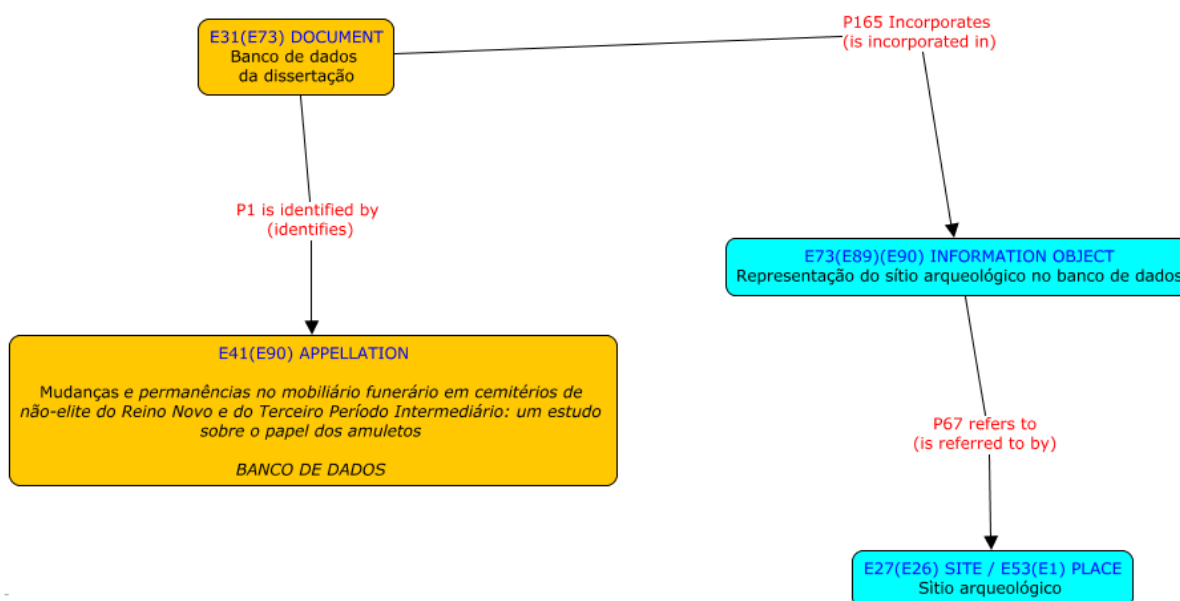
Fonte: elaborado pelo autor (2023)

Aqui, é possível observar que há um indivíduo humano (“Victoria Arroyos Adaime”) envolvido na realização de uma atividade de criação (“Criação da dissertação de mestrado”); esta atividade criou um objeto informacional (“Dissertação de mestrado”) com um identificador próprio (seu título) que tem por assunto um tipo de documento (“Relatório da escavação”, que corresponde à subclasse *MP16 Relatório de escavação*). Cada documento deste tipo possui um título próprio (que corresponde à subclasse *MP23 Título do relatório da escavação*) e documenta uma atividade específica; neste caso, as atividades de escavação (que correspondem às instâncias da subclasse *MP6 Escavação*). A opção pelo uso da propriedade *P129 is about* e não *P67 refers to* se deve

ao fato de *P129* se referir à relação entre um objeto proposicional e seu assunto, enquanto *P67* se refere à relação entre um objeto proposicional e o objeto específico ao qual ele representa em dado contexto.

A diferença pode ser vista em maiores detalhes na Figura 5, abaixo, que apresenta um exemplo de como *P67 refers to* foi utilizada no modelo proposto nesta pesquisa:

**Figura 5:** recorte adaptado do mapa conceitual, com a modelagem dos os sítios arqueológicos estudados por Adaime (2021) e sua representação no banco de dados criados pela autora

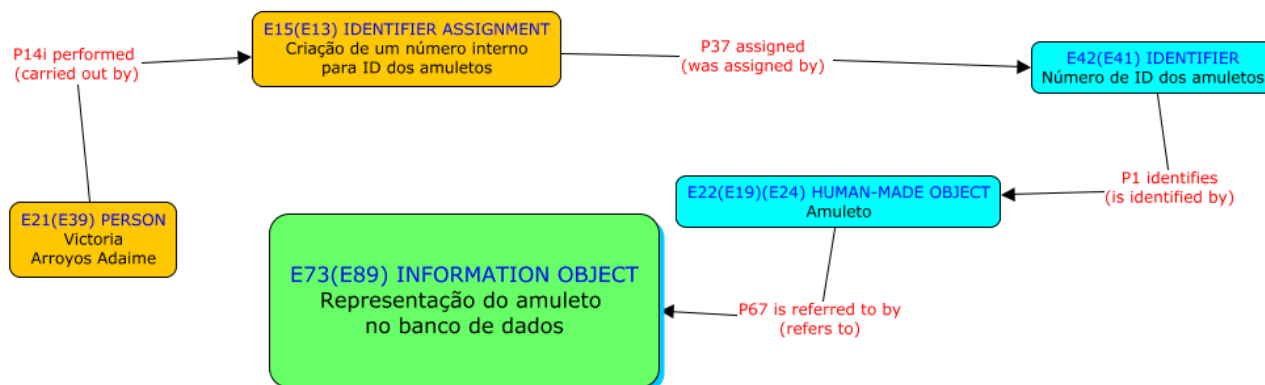


Fonte: elaborado pelo autor (2023)

No recorte, é possível observar que há um documento (“Banco de dados da dissertação”) que incorpora em si objetos informacionais específicos (as representações dos sítios arqueológicos, correspondentes à subclasse *MP19 Representação do sítio arqueológico no banco de dados*). Cada instância desta classe de objetos informacionais, por sua vez, tem como ponto de referência um objeto real: neste caso, os sítios arqueológicos (correspondentes à subclasse *MP20 Sítio Arqueológico*, que foi concebida como uma subclasse tanto da classe *E27(E26) SITE* quanto de *E53(E1) PLACE*). A mesma linha de raciocínio foi aplicada tanto à relação entre os enterramentos (subclasse *MP5 Enterramento*) e suas representações no banco de dados (subclasse *MP18 Representação do enterramento no banco de dados*) quanto aos amuletos funerários (subclasse *MP1 Amuleto*) e suas representações, que, como já exposto, servem de objeto referente de nosso mapa conceitual (subclasse *MP17 Representação do amuleto no banco de dados*).

Durante a elaboração de seu trabalho, Adaime (2021) atribuiu um número de identificação a cada elemento que constaria em seu banco de dados (a saber: os amuletos, enterramentos e sítios arqueológicos). Estes números de identificação foram representados como instâncias das subclasses *MP11*, *MP12* e *MP13*, conforme exemplificado na Figura 6:

**Figura 6:** recorte adaptado do mapa conceitual, com a relação entre os números de identificação dos elementos estudados (amuletos) por Adaime (2021), os elementos em si e suas representações.

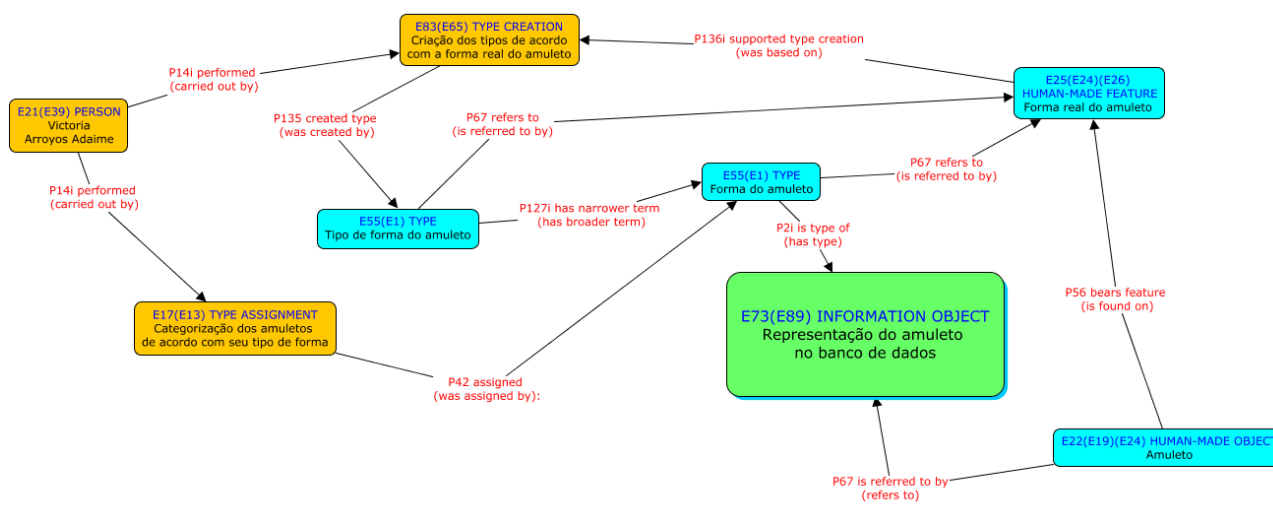


Fonte: elaborado pelo autor (2023)

Aqui, observa-se que é uma instância de *E15(E13) IDENTIFIER ASSIGNMENT* (ou seja, de uma atividade de atribuição de identificadores realizada por um sujeito humano, que é “Victoria Arroyos Adaime”) que atribui um número específico de identificação (instância da subclasse *MP11 Número de ID de amuletos*) a dado objeto criado por humanos; neste caso, um amuleto (instância da subclasse *MP1 Amuleto*). A mesma lógica se aplica para os enterramentos e sítios arqueológicos. Embora a classe *E15(E13) IDENTIFIER ASSIGNMENT* se refira especificamente a uma ação discreta de atribuição de identificador (CIDOC, 2021), optamos por utilizá-la de maneira mais genérica, como uma atividade que pode dar origem a diversos números distintos; tal decisão, embora contraintuitiva frente ao uso declarado da classe, nos pareceu mais aprofundada por permitir que os próprios números de identificação possam ser tratados como entidades em si mesmos, se necessário.

Outro caso em que o olhar centrado em eventos proposto pelo CIDOC-CRM afetou a modelagem foi na seção em que buscamos representar como se deu a classificação dos amuletos estudados por Adaime de acordo com sua tipologia, como é possível visualizar na Figura 7:

Figura 7: recorte adaptado do mapa conceitual, com a cadeia de relações entre os amuletos, sua forma real, sua representação e a representação das tipologias criadas por Adaime para as formas presentes nos relatórios de escavação.



Fonte: elaborado pelo autor (2023)

Na figura, observa-se que um agente humano (“Victoria Arroyos Adaime”) realiza duas atividades distintas: uma atividade de criação de tipo (“Criação dos tipos de acordo com a forma real do amuleto”), instância de *E83(E65) TYPE CREATION*, e uma atividade de atribuição de tipos (“Categorização dos amuletos de acordo com seu tipo de forma”), instância de *E17(E13) TYPE ASSIGNMENT*.

A atividade de criação de tipos teve por base a forma real (instâncias de *MP9 Forma real do amuleto*, subclasse de *E25(E24)(E26) HUMAN-MADE FEATURE*) que os amuletos possuíam, conforme documentado nos relatórios de escavação dos quais a dissertação de Adaime (2021) trata; isto é representado pela propriedade *P136i supported type creation*. Essa atividade deu origem aos tipos de forma, representados como instâncias de *MP22 Tipo de forma do amuleto*. No trabalho de Adaime (2021), os tipos de forma são definidos tanto por sua extensão de acordo com as formas reais apresentadas pelos amuletos quanto pelo significado que cada forma real pode conter; por exemplo, é a partir da observação de que há vários amuletos com formato de gato, vinculado à deusa Bastet, que surge a tipologia “Animais” na categorização de Adaime, dentro da qual constam a forma de gato bem como de outros animais. Para representar este processo de construção de conhecimento empreendido pela autora, optamos por representar as grandes categorias de acordo com a subclasse *MP22 Tipo de forma do amuleto*, enquanto as formas específicas dentro das tipologias foram representadas como

instâncias da subclasse *MP8 Forma do amuleto*. Essa relação de especificação dos conceitos foi possibilitada por meio do uso da propriedade *P127i has narrower term*, que simula relações hierárquicas dentro de um tesouro. Devido a essa especificação, apenas as instâncias de *MP8 Forma do amuleto* são representadas como sendo tipos de *MP8 Forma do amuleto*, via a propriedade *P2i is type of*. Finalmente, observemos que tanto as instâncias de *MP22 Tipo de forma do amuleto* quanto de *MP8 Forma do amuleto* relacionam-se às instâncias de *MP9 Forma real do amuleto*, da qual são derivadas a partir da cadeia apresentada. Como se tratam de construtos mentais criados por Adaime a partir da observação do que era apresentado nos relatórios de escavação, tanto *MP22 Tipo de forma do amuleto* quanto *MP8 Forma do amuleto* são subclasses diretas de *E55(E1) TYPE*).

A atividade de atribuição de tipos (“Categorização dos amuletos de acordo com seu tipo de forma”, instância de *E17(E13) TYPE ASSIGNMENT*) liga-se às instâncias de *MP8 Forma do amuleto*, representando, indiretamente, que as mesmas foram atribuídas às instâncias de *MP17 Representação do amuleto no banco de dados*. As instâncias de *E17(E13) TYPE ASSIGNMENT* no CIDOC-CRM se referem especificamente a uma ação discreta de atribuição de identificador (CIDOC, 2021); contudo, ressaltamos que, em nossa modelagem, optamos por utilizá-las de maneira genérica, referindo-se a diversas instâncias possíveis de *MP8 Forma do amuleto*.

A representação do processo de categorização e atribuição dos tipos de enterramento<sup>33</sup> se deu sob a mesma lógica dos amuletos apresentada acima, com duas diferenças: a primeira é que, no caso dos enterramentos, a subclasse que serviu como base para a criação dos tipos é *MP3 Aspectos físicos do enterramento*<sup>34</sup>, que representa os detalhes físicos e de aparência e funcionalidade que compunham os enterramentos em si. A segunda diferença é que não há especificação da classe *MP21 Tipo de enterramento* como ocorre na relação *MP22 Tipo de forma do amuleto P127i has narrower term MP8 Forma do amuleto* no caso dos amuletos.

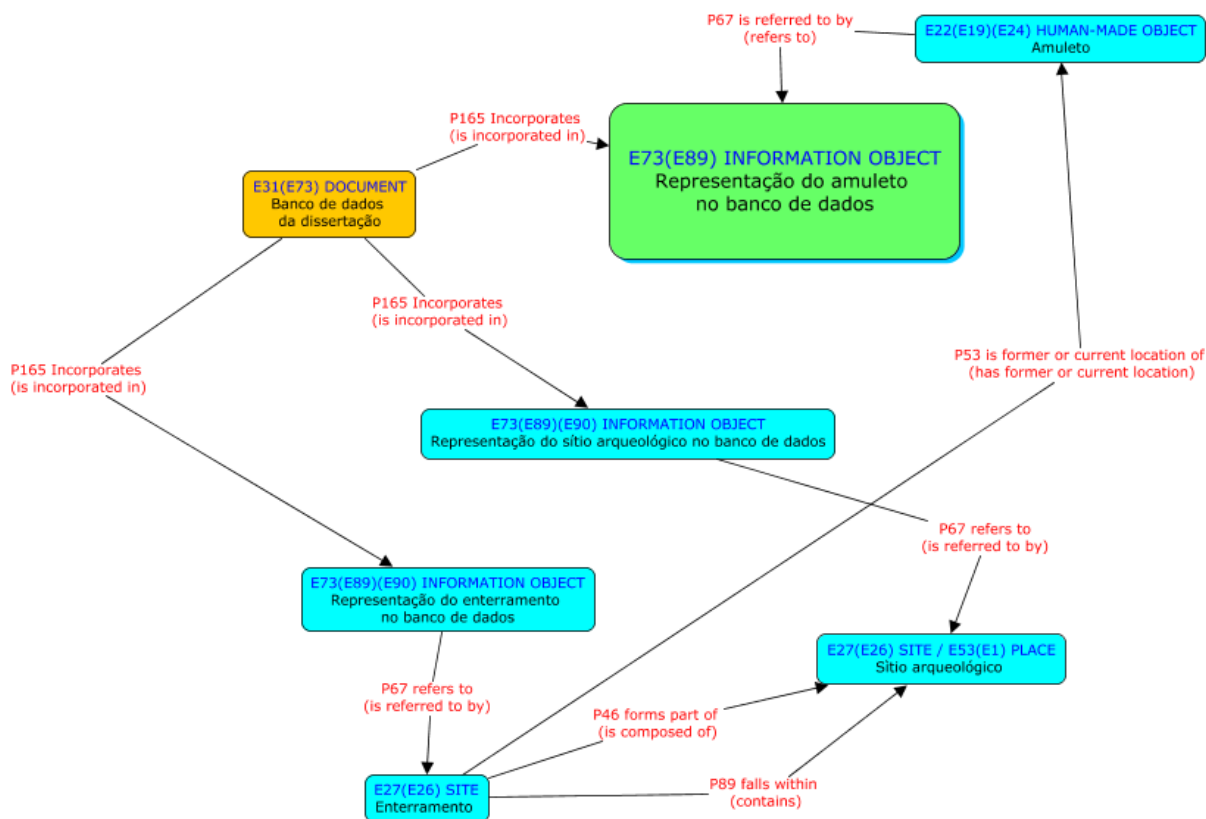
Outra relação importante para o modelo representado é aquela entre os amuletos, os enterramentos e os sítios arqueológicos. Ela é exposta a seguir, na Figura 8:

<sup>33</sup> Que pode ser vista junto da íntegra do mapa conceitual no Apêndice A deste documento.

<sup>34</sup> Há uma particularidade acerca da relação entre *MP3 Aspectos físicos do enterramento* e *MP5 Enterramento*: a propriedade *P46 is composed of* teve de ser utilizada para relacionar as duas classes, embora a Declaração do CIDOC-CRM a destine para relacionar partes específicas de uma coisa física a outra, e não aspectos físicos gerais. Seu uso teve de ser adaptado a fim de que não recorrêssemos ao uso de *P56 bears feature*, cujo uso implicaria na quebra de consistência no momento da formalização da ontologia.



**Figura 8:** recorte adaptado do mapa conceitual, expondo a teia de relações entre os amuletos, enterramentos e sítios arqueológicos, bem como suas representações dentro do banco de dados criado por Adaime (2021).



Fonte: elaborado pelo autor (2023)

Aqui, pode-se observar que os amuletos funerários reais são representados como instâncias de *E22(E19)(E24) HUMAN-MADE OBJECT*, através da subclasse *MP1 Amuleto*; eles se relacionam com os enterramentos onde foram encontrados por meio da propriedade *P53 is former or current location of*, que conecta um lugar (domínio) a dado objeto que nele esteve ou está (*range*). Os enterramentos são representados pela subclasse *MP5 Enterramento*, que é uma subclasse tanto de *E27(E26) SITE* quanto de *E53(E1) PLACE*<sup>35</sup>. Por serem derivadas destas duas superclasses, as instâncias de *MP5 Enterramento* se conectam às instâncias de *MP20 Sítio Arqueológico* via as propriedades *P46 forms part of*, que liga uma parte específica de um objeto físico (domínio) ao objeto em si (*range*) e *P89 falls within*, que liga um lugar específico a outro maior. A subclasse *MP20 Sítio Arqueológico* representa os 3 sítios arqueológicos estudados por Adaime segundo os relatórios de escavação, durante a elaboração de sua dissertação de mestrado. *MP20 Sítio Arqueológico* também é subclasse tanto de *E27(E26) SITE* quanto

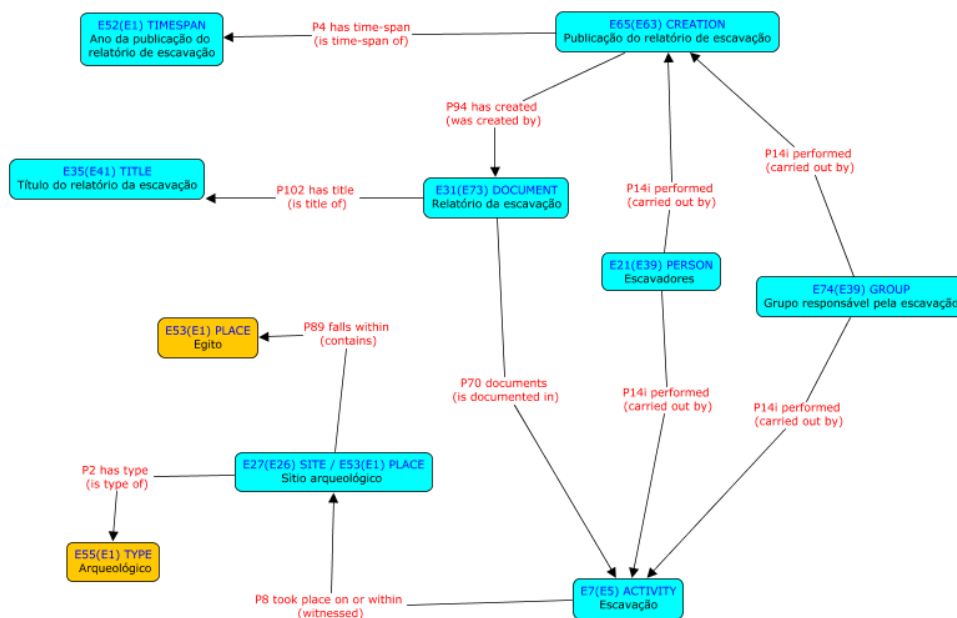
<sup>35</sup> Essa relação de subclasse com *E53(E1) PLACE* deliberadamente não consta no mapa conceitual, mas será ilustrada na Etapa 5 de nossa metodologia.

de *E53(E1) PLACE*, considerando as definições<sup>36</sup> das duas propriedades dentro do documento de declaração do CIDOC-CRM.

As classes *MP1 Amuleto*, *MP5 Enterramento* e *MP20 Sítio Arqueológico* se referem aos elementos no mundo real, e não às suas representações no banco de dados da dissertação de Adaime (2021). Como já explicitado em recortes anteriores, estas representações aparecem no modelo proposto nesta pesquisa via as classes *MP17 Representação do amuleto*, *MP18 Representação do enterramento no banco de dados* e *MP19 Representação do sítio arqueológico no banco de dados*, que se conectam aos objetos reais segundo a propriedade *P67 refers to*. O banco de dados foi representado como uma instância de *E31(E73) DOCUMENT*, e se relaciona com as representações via a propriedade *P165 Incorporates*, que relaciona uma instância de um objeto informacional (domínio) a outra instância de um objeto simbólico (*range*) que esteja contida dentro dele, tal como ocorre com as representações que integram o banco de dados da dissertação.

A relação entre os sítios arqueológicos, as escavações neles ocorridas e sua documentação nos relatórios de escavação é complexa e será exposta, inicialmente, na Figura 9:

**Figura 9:** recorte adaptado do mapa conceitual, expondo a teia de relações entre os sítios arqueológicos, as escavações e os relatórios de escavação, conforme apresentado por Adaime (2021).



Fonte: elaborado pelo autor (2023)

Aqui, observamos que *MP20 Sítio Arqueológico* possui o tipo “Arqueológico”, modelado como instância de *E55(E1) TYPE*, especificando que se trata de um tipo

<sup>36</sup> Reproduzidas, junto a todas as outras definições de classes utilizadas no trabalho, nos Apêndices B e C deste documento.

específico de sítio (dada a definição que o CIDOC-CRM apresenta para a classe *E27(E26) SITE*). As instâncias de *MP20 Sítio Arqueológico* se conectam a “Egito” (instância de *E53(E1) PLACE*) através da propriedade *P89 falls within*, denotando que os sítios estudados estão localizados no Egito. Cada instância de *MP20 Sítio Arqueológico* se conecta a uma ou mais instâncias de *MP6 Escavação* (que é subclasse de *E7(E5) ACTIVITY*) por meio da propriedade *P8 took place on or within*, que liga um evento (neste caso, a escavação) a um lugar (neste caso, o sítio arqueológico). Como os sítios foram palco de mais uma escavação ao longo de sua existência, as instâncias de *MP20 Sítio Arqueológico* não ficam restritas a estarem relacionadas a apenas uma instância de *MP6 Escavação*.

As escavações em si se relacionam a três subclasses distintas, duas das quais são bastante similares: a subclasse *MP10 Grupo responsável pela escavação*, que representa organizações responsáveis pelo desenvolvimento de alguma escavação (e é subclasse de *E74(E39) GROUP*, que representa grupos em geral) e a subclasse *MP7 Escavador*, que representa aqueles indivíduos que, sozinhos ou coletivamente, coordenaram as escavações ocorridas nos sítios arqueológicos (sendo subclasse de *E21(E39) PERSON*, que representa pessoas em geral. Tanto *MP7* quanto *MP10* são, em último grau, herdeiras de *E39(E77) ACTOR*, o que permite que sejam ligadas aos eventos representados em *MP6 Escavação* via a propriedade *P14i performed*. Como mais de uma pessoa ou grupo pode ser responsável por uma escavação, cada instância de *MP6 Escavação* pode estar conectada a mais de uma instância de *MP7 Escavador* ou *MP10 Grupo responsável pela escavação*.

Os agentes representados nas instâncias de *MP7 Escavador* ou *MP10 Grupo responsável pela escavação* foram responsáveis por criar os relatórios que documentam cada evento de escavação, representados pela subclasse *MP16 Relatório de escavação*. Isto é representado por meio da cadeia de relações que conecta as instâncias de *MP7* e *MP10* a alguma instância de *MP15 Publicação do relatório de escavação*, via a propriedade *P14i performed*. *MP15 Publicação do relatório de escavação* reflete a perspectiva centrada em eventos do CIDOC-CRM, e tem suas instâncias relacionadas às de *MP2 Ano da publicação do relatório de escavação* (que representa as extensões de tempo na qual foram publicados os relatórios finalizados) por meio da propriedade *P4 has time-span*. Cada publicação de relatório deve resultar na criação de um relatório de pesquisa, motivo pelo qual cada instância de *MP15 Publicação do relatório de escavação*

se relaciona a uma instância de *MP16 Relatório de escavação* (que é subclasse de *E31(E73) DOCUMENT*). Por sua vez, cada instância de *MP16 Relatório de escavação* possui um título, representado nas instâncias da subclasse *MP23 Título do relatório da escavação*. Finalmente, e fechando o ciclo de representação das escavações, cada relatório documenta ao menos uma escavação, e portanto as instâncias de *MP16 Relatório de escavação* se relacionam às instâncias de *MP6 Escavação* via a propriedade *P70 documents*.

Tendo apresentado as relações mais importantes estabelecidas em nosso mapa conceitual e explicado o raciocínio por trás delas, encerramos o relatório das Etapas 3 e 4 de nossa pesquisa. A seguir, iniciamos a descrição dos desenvolvimentos que ocorreram durante a Etapa 5.

### **6.5 Etapa 5: Formalização da ontologia**

A etapa de formalização da ontologia, seguindo o que é proposto pela metodologia *OntoForInfoScience*, consiste na descrição formal do domínio com o qual se está trabalhando em linguagem legível por máquinas (MENDONÇA, 2015). Desta forma, cabe identificar primeiramente quais os recursos utilizados para o desenvolvimento deste segmento da pesquisa.

Considerando que nossa ontologia de fundamentação (o modelo CIDOC-CRM) é disponibilizada em formato de arquivo em seu *website*<sup>37</sup>, iniciamos os procedimentos com o *download* do arquivo-base contendo a taxonomia do CIDOC-CRM, para que o mesmo pudesse ser editado conforme as necessidades da pesquisa. O uso do arquivo já existente permitiu que os esforços de representação da pesquisa ocorressem de maneira mais eficiente, evitando o retrabalho resultante de ter que refazer as representações que já constavam no arquivo consolidado. Dentre os arquivos referentes à versão 7.1.1. do CIDOC-CRM, utilizada nesta pesquisa, havia um modelo em formato RDFS; foi este o arquivo escolhido para edição, convertido para o formato OWL após o *download*. Em seguida fizemos o *download* do *software* Protegé, um programa específico para a criação e edição de ontologias, amplamente utilizado no desenvolvimento deste tipo de artefato representacional (SACHS, 2006). Tendo em mãos o arquivo-matriz do CIDOC-CRM 7.1.1. e conhecendo aos poucos as funcionalidades oferecidas pelo Protegé, foi dado início à

---

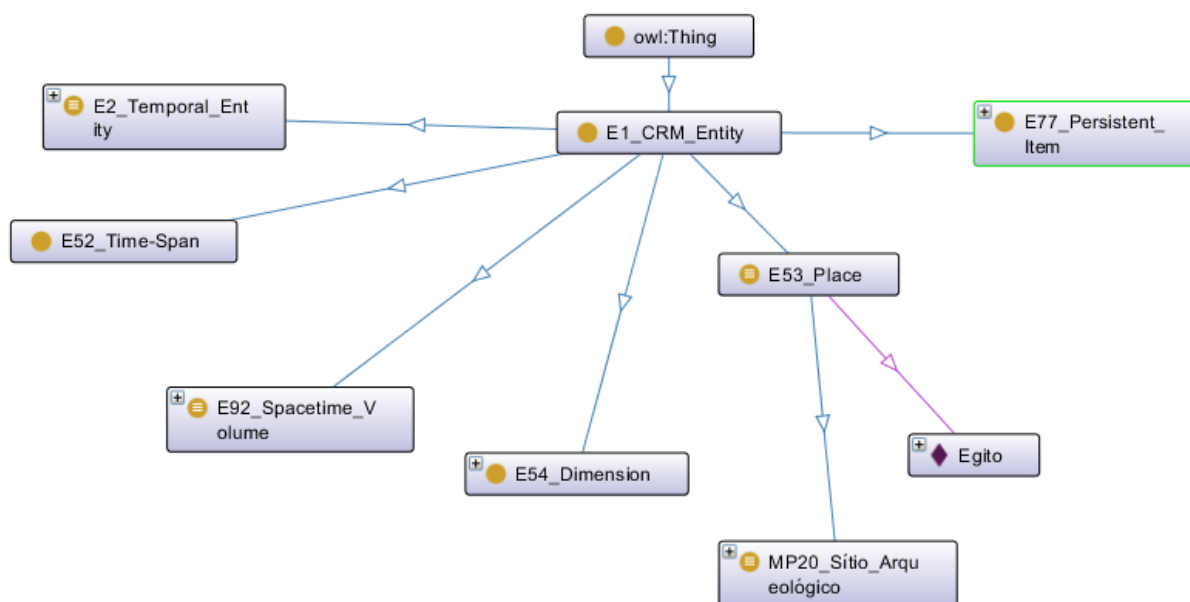
<sup>37</sup> <https://cidoc-crm.org/Version/version-7.1.1> e <https://cidoc-crm.org/versions-of-the-cidoc-crm>.

formalização do modelo proposto, seguindo, sempre que possível, os passos recomendados por Mendonça (2015).

### 6.5.1 Primeiro passo: criação da taxonomia geral da ontologia

O primeiro passo consiste na criação da taxonomia geral da ontologia. De acordo com Mendonça (2015), este passo consiste na estruturação de uma taxonomia do domínio analisado, na qual se categorizam as entidades mais gerais até as mais específicas. Conforme explicado, as novas classes criadas nesta pesquisa consistem em subclasses das categorias já existentes no CIDOC-CRM, enquanto as instâncias individuais podem pertencer tanto às classes nativas do CIDOC-CRM quanto às classes do modelo proposto na pesquisa. Considerando o tamanho do modelo CIDOC-CRM somado às adições realizadas durante a pesquisa, a exibição de todas as suas entidades fica inviável; nesse sentido, a Figura 10, gerada com o módulo OntoGraf do Protegé, expõe apenas as classes primordiais do CIDOC-CRM, e já inclui a classe *MP20 Sítio Arqueológico*, não-nativa do CIDOC-CRM e criada durante a etapa de Conceitualização desta pesquisa.

Figura 10: taxonomia básica do modelo CIDOC-CRM 7.1.1



Fonte: elaborado pelo autor (2023)

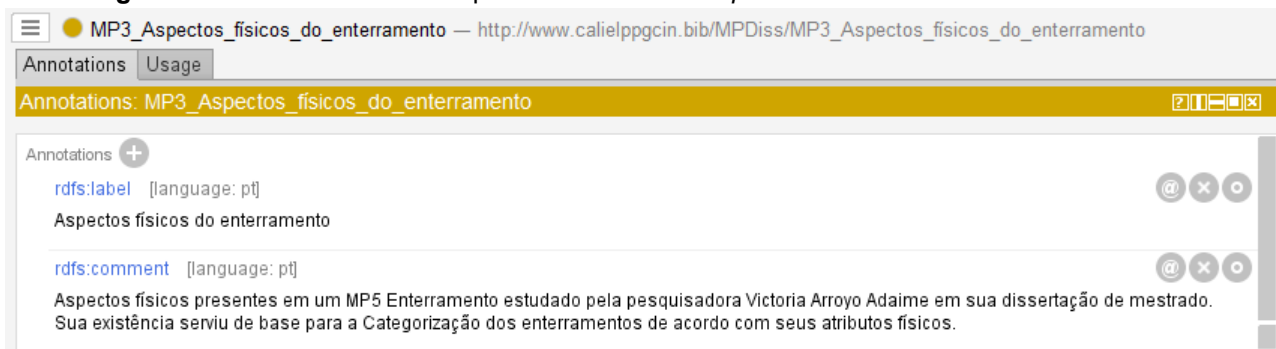
A Figura 10 explicita as relações entre a classe-raíz *E1 CRM-Entity* e suas subclasses diretas *E2 TEMPORAL ENTITY*, *E52 TIME-SPAN*, *E53 PLACE*, *E54 DIMENSION*, *E77 PERSISTENT ITEM* e *E92 SPACETIME VOLUME*. Assim, destaca-se que toda a formalização da ontologia se deu, portanto, a partir da correspondência entre

as instâncias desenvolvidas durante a pesquisa (sejam indivíduos em si ou subclasses novas) e as subclasses que surgem da taxonomia primordial apresentada acima. Esta dinâmica pode ser observada na Figura 10: como instâncias da classe *E53 PLACE*, existem tanto o indivíduo “Egito” quanto a subclasse *MP20 Sítio Arqueológico*.

#### 6.5.2. Segundo passo: Definição das propriedades descritivas da ontologia

O segundo passo para a concretização da etapa de Formalização da ontologia consiste na definição das propriedades descritivas das entidades presentes na ontologia; ou seja, na declaração dos atributos informativos que garantem a riqueza semântica dos entes e relacionamentos representados na ontologia. Estes atributos são representados dentro do Protegé como *Annotations*, atribuíveis a quaisquer entidades dentro do arquivo com o qual se está trabalhando. Mendonça (2015) propõe alguns atributos que considera fundamentais, considerando as *Annotations* disponíveis na versão 4.3 do Protegé, com a qual o autor trabalhava. Em nosso caso, optamos por não incluir *Annotations* além das que constam na versão 5.5 do Protegé, com a qual trabalhamos. Desta forma, as *Annotations* selecionadas para inclusão no modelo foram *rdfs : label*, que descreve um “nome” para a entidade à qual é atribuído; e *rdfs : comment*, de preenchimento livre; ambos já estavam sendo usados no arquivo original do CIDOC-CRM, ao que demos continuidade. O atributo *rdfs : label* foi aplicado a todas as entidades presentes no arquivo; as classes e propriedades já presentes no arquivo original disponível no *website* do CIDOC-CRM possuem diversos usos de *rdfs : label*, em diferentes idiomas, ao passo que as entidades criadas durante a pesquisa possuem apenas um uso de *rdfs : label*, especificamente na língua portuguesa. Para o atributo *rdfs : comment*, seguimos o padrão que vinha sendo usado no arquivo original, que utilizava o campo para incluir as definições das classes e propriedades; desta forma, as definições presentes nos apêndices B e C deste documento foram incluídas no arquivo da ontologia através do atributo *rdfs : comment*. A Figura 11, traz um exemplo do uso das *Annotations* na representação da subclasse *MP3 Aspectos físicos do enterramento*.

**Figura 11:** uso das *Annotations* para a classe *MP3 Aspectos físicos do enterramento*



Fonte: elaborado pelo autor (2023)

### 6.5.3. Terceiro passo: Definição formal para as classes da Ontologia

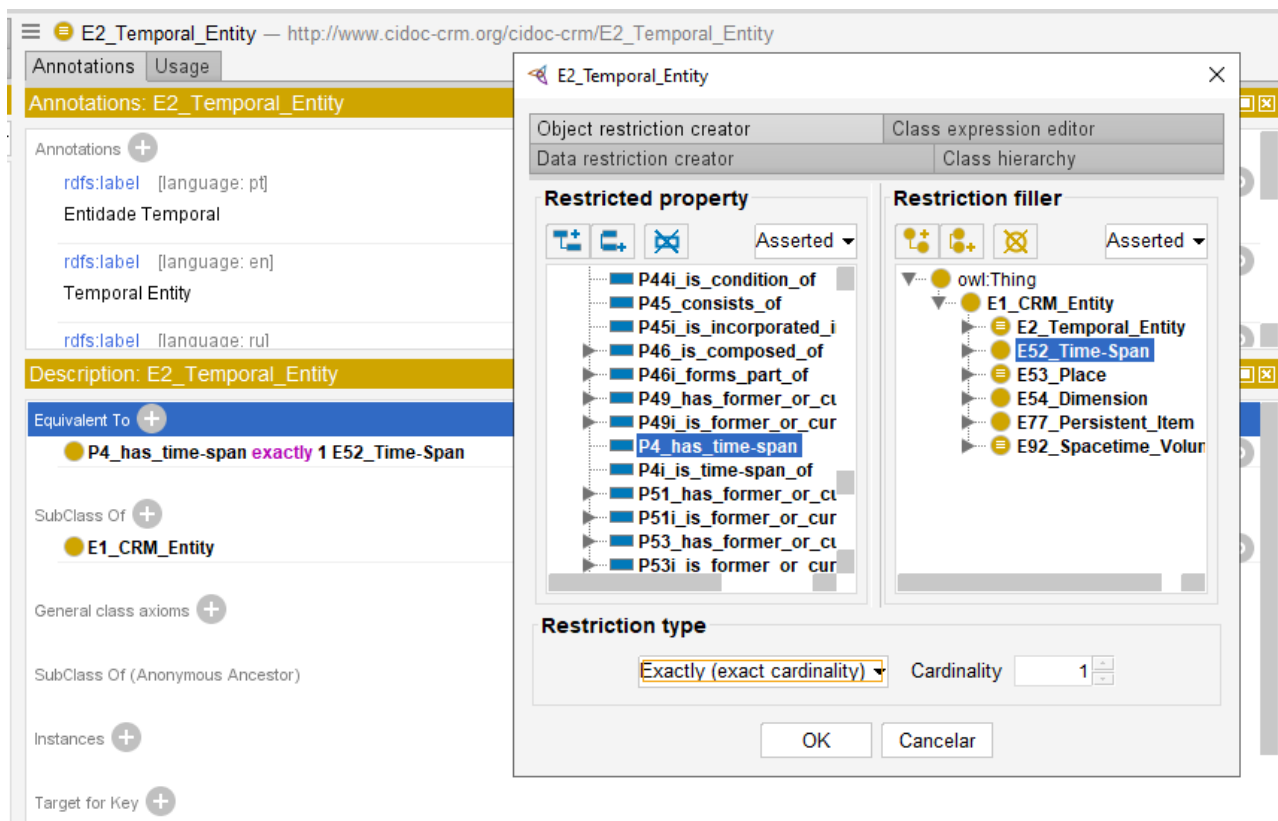
O terceiro passo da Etapa 5 consiste na definição formal das classes dentro do arquivo da ontologia. Segundo Mendonça (2015), é esta etapa que garante que o potencial de inferência, que é o diferencial das ontologias, possa ser utilizado por aplicações capazes de entender a linguagem na qual o arquivo da ontologia está construído. Neste caso, a definição formal das classes ocorre a partir das restrições de tipo e cardinalidade<sup>38</sup> que se declara acerca delas, considerando sua modelagem conceitual. Em nosso caso, tivemos dois pontos de partida para esta etapa: as classes e as propriedades nativas do modelo CIDOC-CRM. Ao todo, definimos formalmente as 22 subclasses criadas durante nossa pesquisa, bem como cerca de 20 classes nativas do CIDOC-CRM que serviram de uma maneira ou outra ao nosso esforço de modelagem<sup>39</sup>.

A formalização das classes nativas do CIDOC-CRM levou em consideração as restrições de cardinalidade apresentadas pela Declaração de Classes e Propriedades do CIDOC-CRM versão 7.1.1., e foi feita utilizando a ferramenta de restrição cardinalidade para propriedades de objetos que o Protegé oferece. A Figura 12 oferece um exemplo do procedimento de declaração, realizado através da ferramenta de restrição, expondo a declaração da classe *E2 TEMPORAL ENTITY*.

<sup>38</sup> A Cardinalidade é uma “propriedade matemática que identifica o número de elementos em um conjunto” (ALMEIDA, 2020, p.239). Nesse contexto, indica a quantidade de indivíduos que podem estar presentes no conjunto que é a intersecção que se forma quando duas classes se ligam via uma propriedade.

<sup>39</sup> A taxonomia completa do CIDOC-CRM 7.1.1, bem como todas suas propriedades, foi mantida no arquivo para fins de completude, mas podem ser removidas a fim de evitar o excesso de informações desnecessárias (*code-bloat*) no arquivo da ontologia.

**Figura 12:** uso da ferramenta de restrição do Protégé para definição formal da classe *E2 TEMPORAL ENTITY*



Fonte: elaborado pelo autor (2023)

As declarações foram feitas a partir das restrições inerentes a cada propriedade de objeto<sup>40</sup>, conforme descrito na Declaração do CIDOC-CRM 7.1.1. No caso acima, por exemplo, a Declaração define, primeiramente, que o objeto da propriedade *P4 has time-span* deve ser uma instância de *E2 TEMPORAL ENTITY*, enquanto seu *range* deve ser uma instância da classe *E52 TIME-SPAN*. A seguir, é estipulado que a propriedade se quantifica por meio da relação *many to one, necessary* (1,1:0,n). Esta quantificação é definida no início da Declaração da seguinte maneira:

***many to one, necessary (1,1:0,n):*** *An individual domain instance of this property must have exactly one instance of this property, but an individual range instance can be referenced by zero, one or more instances of this property. In other words, this property is necessary and not repeatable for its domain, and optional and repeatable for its range. In some contexts, this situation is called a “fan-in” (CIDOC, 2021, grifo nosso).*

<sup>40</sup> Propriedades de objeto (*object properties*) são, resumidamente, aquelas que conectam dois indivíduos em um modelo, expondo relacionamentos entre eles. Contrastam com as propriedades de dados (*data properties*), que conectam indivíduos com conjuntos de dados literais em formato pré-definido (W3C, 2012)



Considerando que essa propriedade é necessária e não-repetível para seu objeto, mas opcional e repetível para seu *range*, compreende-se ser uma propriedade funcional<sup>41</sup>, denotando que entidades temporais (como períodos, atividades e eventos) só podem ter uma única extensão de tempo dentro da qual ocorrem. Uma distinção importante do modelo CIDOC-CRM é que ele se refere a quantificações para suas propriedades, evitando o uso do termo “Restrição de cardinalidade”; isso é devido ao fato de o CIDOC-CRM se propor mais como um modelo de implementação que como um guia de modelagem que deve ser seguido à risca (CIDOC, 2021).

Assim, a modelagem das classes no Protegé se deu com base no que declara o modelo CIDOC-CRM acerca de suas propriedades. A Figura 13, a seguir, demonstra como foi feita a declaração formal da classe *MP5 Enterramento*, uma subclasse desenvolvida durante nossa pesquisa que, conforme explicado na Etapa 3, representa os enterramentos propriamente ditos; e os objetos reais estudados, através dos relatórios de escavação, por Victoria Adaime para a construção de sua dissertação de mestrado. A representação formal da classe *MP5 Enterramento* na Figura 13 pode ser comparada com sua representação no mapa conceitual, apresentada anteriormente na Figura 8.

Figura 13: definição formal da classe *MP5 Enterramento*

The screenshot displays the Protegé ontology editor interface. On the left, the 'Class hierarchy' pane shows a tree structure starting from 'owl:Thing'. Under 'E18\_Physical\_Thing', 'E19\_Physical\_Object' is listed, and under it, 'E27\_Site' is shown. 'E27\_Site' has two subclasses: 'MP20\_Sito\_Arqueológico' and 'MP5\_Enterramento'. The 'MP5\_Enterramento' class is highlighted in blue. On the right, the 'Description: MP5\_Enterramento' pane shows the following information:

- Equivalent To:** A logical definition consisting of several axioms:
  - (P46\_is\_composed\_of some MP3\_Aspectos\_fisicos\_do\_enterramento)
  - and (P53L\_is\_former\_or\_current\_location\_of min 1 MP1\_Amuleto)
  - and (P1\_is\_identified\_by exactly 1 MP12\_Número\_de\_ID\_de\_enterramentos)
  - and (P46\_forms\_part\_of exactly 1 MP20\_Sito\_Arqueológico)
  - and (P67L\_is\_referred\_to\_by exactly 1 MP18\_Representação\_do\_enterramento\_no\_banco\_de\_dados)
  - and (P89\_falls\_within exactly 1 MP20\_Sito\_Arqueológico)
- SubClass Of:** A list of parent classes: 'E27\_Site' and 'E53\_Place'.
- General class axioms:**
  - P56L\_is\_found\_on exactly 1 E19\_Physical\_Object
  - P53\_has\_former\_or\_current\_location min 1 E53\_Place
  - P45\_consists\_of min 1 E57\_Material
  - P89\_falls\_within min 1 E53\_Place
- Instances:** 'Enterramento\_115' and 'Enterramento\_125'.

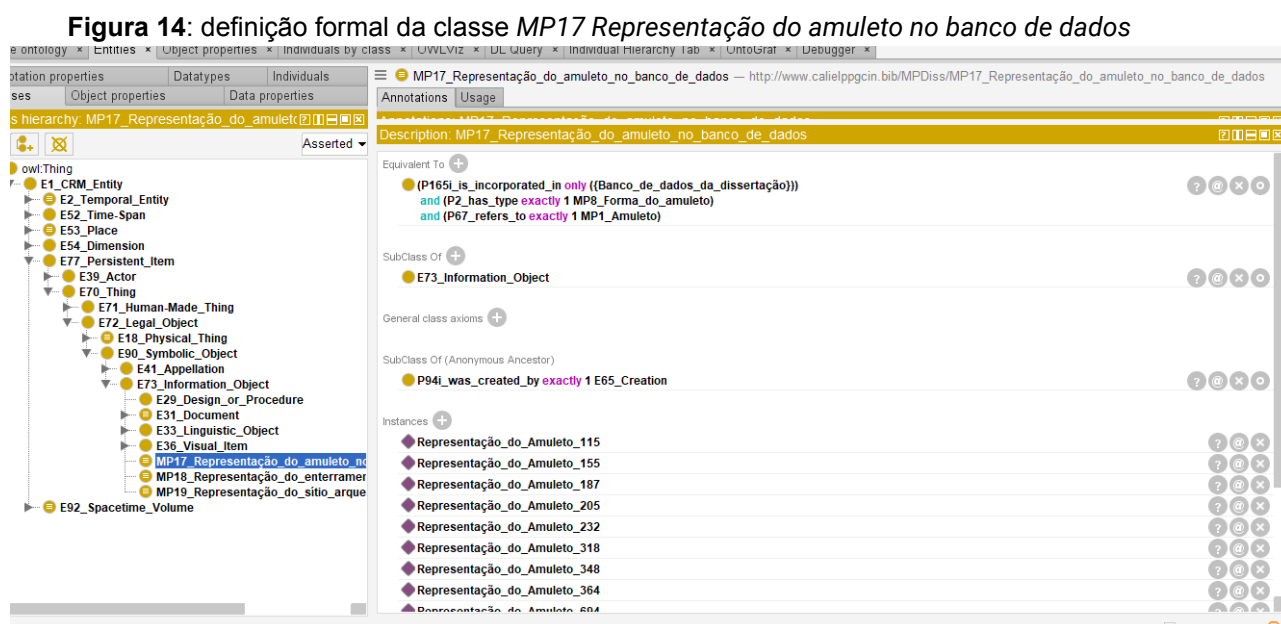
Fonte: elaborado pelo autor (2023)

Na Figura 13, pode-se observar factualmente como ocorreu a formalização das classes. À esquerda, onde consta a taxonomia do modelo, podemos observar que *MP5 Enterramento* aparece como uma subclasse da classe CIDOC-CRM *E27 Site*; esta é a única definição de hierarquia declarada explicitamente sobre a taxonomia na qual *MP5*

<sup>41</sup> Propriedades funcionais são aquelas propriedades de objeto em que, para cada indivíduo X, pode haver apenas um indivíduo Y se que conecta a ele através daquela propriedade (W3C, 2012).

*Enterramento* está inserida. No entanto, à direita, no campo “SubClass of”, podemos observar que, através de inferências, o Protégé reconhece *MP5 Enterramento* também como subclasse de *E53 Place*. Esta relação taxonômica foi intencionalmente deixada de fora do mapa conceitual apresentado na Etapa 3 da metodologia, a fim de expor aqui como se dá o processo de inferência. Neste caso, a inferência ocorre devido ao uso das propriedades *P53i has former or current location* e *P89 falls within*, que possuem como domínio instâncias de *E53 Place*. Ao declarar formalmente *MP5 Enterramento* como domínio destas propriedades, o motor de inferência do Protégé o reconhece como instância de *E53 Place*.

Outra definição formal importante foi a que se refere ao nosso objeto referente no mapa conceitual: a representação do amuleto no banco de dados, expressa através da classe *MP17*. A Figura 14 expressa a definição geral da classe, e pode ser comparada com sua representação conceitual na Figura 2, apresentada anteriormente:



Fonte: elaborado pelo autor (2023)

Na Figura 14, está declarado que as instâncias da classe *MP17 Representação do amuleto no banco de dados* estão incorporadas no “Banco de dados da dissertação”, modelado como uma instância da classe *E31 Document*. Além disso, são declaradas as restrições necessárias à classe para que esta faça sentido. Primeiramente, é declarado que cada instância da classe é ligado (pela propriedade *P2 has type*) a apenas uma instância de *MP8 Forma do Amuleto*; isto é, a formalização da relação conceitual de que cada amuleto possuía uma dada forma real, que serviu de base para a criação de

tipologias e tipos mais específicos, como será explicado na Figura 16, mais à frente. Ainda sobre a Figura 14, cada instância de *MP17* liga-se (através da propriedade *P67* *refers to*) a apenas uma instância de *MP1 Amuleto*. Desta forma, o que se está declarando em linguagem formal neste caso é que cada representação de amuleto representa um único amuleto e possui um único tipo, que é o conceito extraído da forma que o amuleto real detinha.

Uma cadeia importante de relações apresentadas na Etapa 3, de Conceitualização, é aquela que representa a relação entre as formas reais dos amuletos e as tipologias criadas a partir destas formas. As Figuras 15 e 16, a seguir, expõem como se deu a formalização desta cadeia de relações a partir da declaração formal das classes *MP8 Forma do amuleto* e *MP22 Tipo de Forma do amuleto*. O recorte do mapa conceitual que traz a representação informal destas relações está presente na Figura 7, apresentada anteriormente.

**Figura 15:** definição formal da classe *MP8 Forma do Amuleto*

The screenshot displays the OWL editor interface for the class *MP8 Forma do Amuleto*. The left pane shows a class hierarchy starting from *owl:Thing* down to *MP8 Forma do Amuleto*. The central pane shows the class description:

```

Equivalent To
  (P42i_was_assigned_by_only ((Categorização_dos_amuletos_de_acordo_com_seu_tipo_de_forma)))
  and (P2i_is_type_of_min 1 MP17_Representação_do_amuleto_no_banco_de_dados)
  and (P127_has_broader_term_exactly 1 MP22_Tipo_de_forma_do_amuleto)
  and (P67_refers_to_exactly 1 MP9_Forma_real_do_amuleto)

SubClass Of
  E55_Type
  E89_Propositional_Object

General class axioms
  P135i_was_created_by_exactly 1 E83_Type_Creation
  P94i_was_created_by_exactly 1 E65_Creation

Instances
  Coração
  Falcão
  Gato
  Mesa
  Sekhmet
  Situla
  
```

Fonte: elaborado pelo autor (2023)

Figura 16: definição formal da classe *MP22 Tipo de Forma do Amuleto*

The screenshot shows the Protegé ontology editor with the following details:

- Class Hierarchy:** A tree view on the left showing the class hierarchy. The class *MP22 Tipo de Forma do Amuleto* is highlighted in blue. Its parent is *MP8 Forma do amuleto*, which is a subclass of *E71 Human-Made Thing*, *E70 Thing*, *E39 Actor*, and *E77 Persistent Item*. *MP8* is also a subclass of *E89 Propositional Object* and *E90 Symbolic Object*.
- Description:** The description field for *MP22 Tipo de Forma do Amuleto* is empty.
- Equivalent To:** A logical axiom defining the class as equivalent to:
 
$$(P67\_refers\_to\ some\ MP9\_Forma\_real\_do\_amuleto) \wedge (P135i\_was\_created\_by\ only\ ((Criacao\_dos\_tipos\_de\_amuleto\_com\_base\_na\_forma\_real\_do\_amuleto))) \wedge (P127i\_has\_narrower\_term\ min\ 1\ MP8\_Forma\_do\_amuleto)$$
- SubClass Of:** The class is a subclass of *E55 Type* and *E89 Propositional Object*.
- General class axioms:** None are listed.
- SubClass Of (Anonymous Ancestor):** Two axioms are listed:
  - $P135i\_was\_created\_by\ exactly\ 1\ E83\_Type\_Creation$
  - $P94i\_was\_created\_by\ exactly\ 1\ E65\_Creation$
- Instances:** A list of instances is shown: *Animais*, *Divindade*, *Formas naturais*, *Outras formas*, *Partes do corpo*, and *Simbolos divinos*.

Fonte: elaborado pelo autor (2023)

Na Figura 15, podemos observar que está declarado que as instâncias da classe *MP8 Forma do Amuleto*, num todo, foram designadas em um evento de “Categorização dos amuletos de acordo com seu tipo de forma”, uma instância de *E17 Type Assignment*. Além disso, cada instância pode servir como tipo para uma ou mais instâncias de *MP17 Representação do amuleto no banco de dados*, relação já representada de maneira inversa quando da formalização de *MP17*, apresentada na Figura 14. Também é declarado que cada instância de *MP8* possui, como termo mais geral, exatamente uma instância de *MP22 Tipo de Forma do Amuleto* e representa apenas uma instância de *MP9 Forma real do amuleto*; essas relações são feitas via as propriedades *P127 has broader term* e *P67 refers to*. Finalmente, a Figura 16 apresenta uma inferência significativa para o processo de representação: devido ao uso da propriedade *P67 refers to*, que tem como domínio instâncias de *E89 Propositional Object*, o Protegé compreende que *MP8* também é uma subclasse de *E89*. Desta forma, infere-se formalmente a noção de que *MP8* representa a uma categoria de objetos conceituais, que, por sua vez, se refere a uma classe de características reais de objetos materiais (as formas físicas do amuletos reais, representadas pela classe *MP9 Forma real do amuleto*).

A Figura 16, por sua vez, traz a representação formal de *MP22 Tipo de Forma do amuleto*. Aqui, declara-se, via a propriedade *P67 refers to* que as instâncias de *MP22* se referem a alguma instância de *MP9 Forma real do amuleto*, o que permite a inferência de que *MP22* é também é uma subclasse de *E89 Propositional Object*; ou seja, similar ao que ocorreu com *MP8 Forma do amuleto*, essa inferência garante a representação

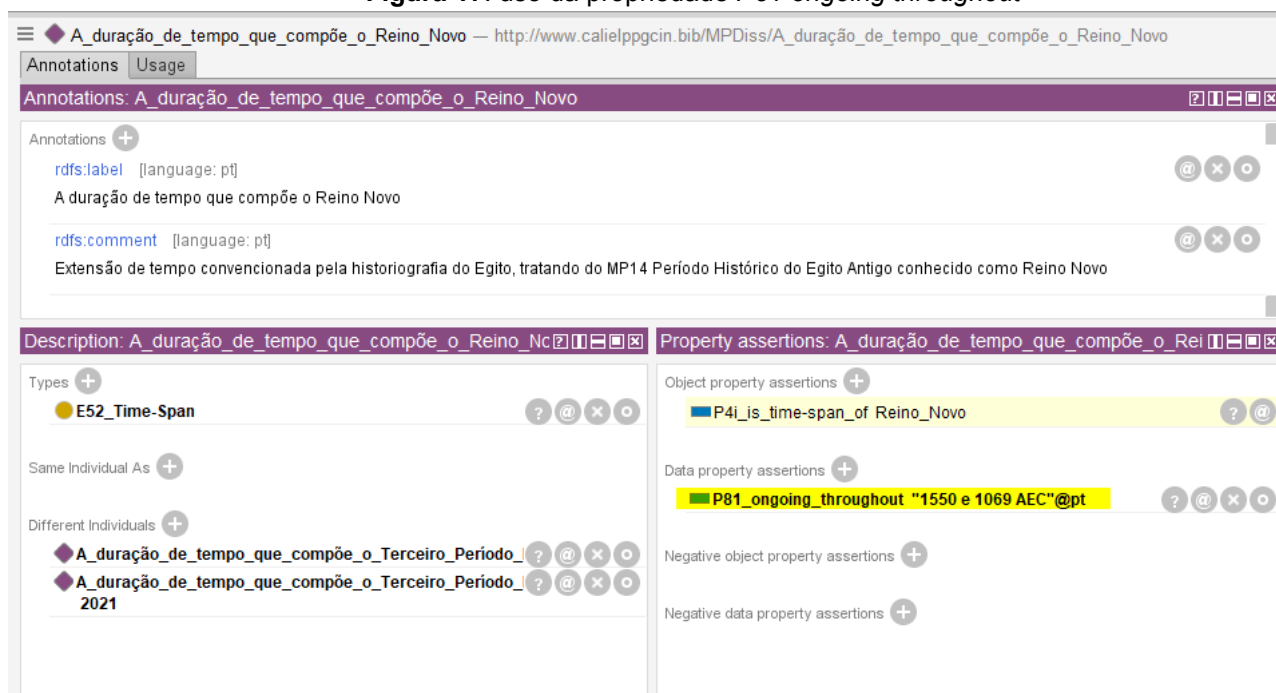
adequada dos conceitos que deram origem a *MP22* e de sua relação com as características reais que deram origem a *MP9*. Além disso, é declarado que *MP22* surgiu a partir de um evento de categorização representado no indivíduo “Criação dos tipos de acordo com a forma real do amuleto”, modelado como instância de *E83 Type Creation*. Finalmente, declara-se formalmente que cada instância de *MP22* possui, como “termo” mais específico, uma instância de *MP8 Forma do amuleto*. Aqui, as restrições de cardinalidade *P67\_refers\_to some MP9\_Forma\_real\_do\_amuleto* e *P127i\_has\_narrower\_term min 1 MP8\_Forma\_do\_amuleto* são importantes pois permitem a representação formal da ideia de que foi a existência das formas reais dos amuletos (instâncias de *MP9*) que levou à criação das tipologias maiores (instâncias de *MP22*), as quais, portanto, se mantêm conceptualmente estáveis na medida em que se referem às formas reais, enquanto elementos de qualificação (as instâncias de *MP8*) dos amuletos (ou, em nosso caso, de suas representações no banco de dados, conforme explicado no parágrafo anterior).

#### 6.5.4. Quarto passo: definição das propriedades de dados das classes

O quarto passo da Etapa 5 consiste na construção das propriedades de dados que serão utilizadas no modelo que se está representando, bem como restringir o tipo de dados que pode ser aceito por cada uma destas propriedades. Considerando que não construímos propriedades além daquelas que o CIDOC-CRM já trazia, este passo consistiu apenas em selecionarmos quais destas propriedades seriam utilizadas em nosso modelo.

Desta forma, fizemos uso da propriedade *P81 ongoing throughout*, que relaciona uma instância de *E52 Time-span* a uma instância de *E61 Time Primitive*; ou seja, a um valor literal que represente unidades de tempo. Em nosso caso, *P81* foi utilizada para dar valor à extensão de tempo que compõe os períodos históricos envolvidos na pesquisa de Adaime (2021), o Reino Novo e o Terceiro Período Intermediário. Um exemplo desta relação consta na Figura 17, que demonstra o uso da propriedade para declarar o valor da extensão de tempo do Reino Novo.

**Figura 17:** uso da propriedade *P81 ongoing throughout*



Fonte: elaborado pelo autor (2023)

### 6.5.5. Quinto Passo e Sexto Passo: criação das instâncias das classes e especificação das relações ontológicas

O quinto passo da Etapa de Formalização consiste na criação das instâncias individuais das classes declaradas anteriormente no modelo. Mendonça (2015) afirma que uma das maiores dificuldades neste passo é o de escolher se dado conceito será representado como uma classe nova ou apenas como um indivíduo no modelo; dificuldade que pode ser sanada caso os procedimentos da Etapa de Conceitualização tenham sido realizados com êxito. Em nosso caso, como as Etapas 3 e 4 deram origem ao conteúdo dos Apêndices B e C, que discriminam quais conceitos se apresentam como novas classes e quais se apresentam como indivíduos das classes (quer sejam as classes nativas do CIDOC-CRM ou as classes elaboradas durante a pesquisa), este quinto passo da Etapa de Formalização consiste na formalização dos indivíduos que integram nosso modelo. Ao todo, foram modelados formalmente 200 indivíduos. Como mencionado na etapa de Conceitualização, a opção por se trabalhar com um recorte da dissertação de Adaime fez com que o esforço de modelagem formal fosse limitado àquelas entidades que buscávamos de fato representar dentro do escopo desta pesquisa. Desta forma, foram modelados apenas os indivíduos que possuem algum uso no modelo;

dito de outra forma, não modelamos nenhuma instância de qualquer classe que não fosse ser utilizada na ontologia.

Já o sexto passo consiste na especificação das propriedades que existiam em nível conceitual para a linguagem formal do arquivo da ontologia. Aqui, formalizamos aquelas relações apresentadas nas Etapas 3 e 4 e descritas no Apêndice D, passando para a linguagem formal aquelas relações construídas a partir das propriedades oferecidas pela Declaração do CIDOC-CRM 7.1.1. Embora tenhamos mantido no arquivo da ontologia todas as propriedades nativas do CIDOC-CRM 7.1.1., fizemos uso apenas das 25 propriedades que constam no Apêndice D deste documento.

A seguir, apresentamos alguns exemplos de instâncias individuais representadas no modelo, bem como as relações que eles possuem com outros indivíduos. Inicialmente, a Figura 18 traz a representação de uma instância da classe *MP17 representação do amuleto no banco de dados*, o indivíduo “Representação do amuleto 115”.

**Figura 18:** representação do indivíduo "Representação do amuleto 115"

The screenshot displays the Protege ontology editor interface. The top menu bar includes options like 'Active ontology', 'Entities', 'Object properties', 'OWL Viz', 'Individual Hierarchy Tab', 'DL Query', 'OntoGraf', 'Debugger', and 'Individuals by class'. The main window is divided into several panes:

- Class hierarchy:** Shows a tree of classes, with 'MP17\_Representação\_do\_amuleto\_no\_banco\_de\_dados' selected.
- Annotations:** Shows the annotation 'rdfs:label [language:pt] Representação do amuleto 115'.
- Description:** Shows the description 'Representação do Amuleto\_115'.
- Property assertions:** Lists several assertions for the instance, including 'P1061\_forms\_part\_of Banco\_de\_dados\_da\_dissertação', 'P1651\_is\_incorporated\_in Banco\_de\_dados\_da\_dissertação', 'P2\_has\_type Gato', and 'P67\_refers\_to Amuleto\_115'.
- Direct instances:** Lists other instances of the class, such as 'Representação do Amuleto\_155', 'Representação do Amuleto\_187', etc.

The bottom status bar indicates 'Reasoner active' and 'Show Inferences'.

Fonte: elaborado pelo autor (2023)

Na Figura 18, é importante notar que todas as relações que existem entre este indivíduo e outras instâncias dentro do modelo se dão de maneira inferida, e não declarada; em outras palavras, embora nada tenha sido dito diretamente acerca deste indivíduo, ele é o objeto (*range*) em declarações formais de relacionamentos que envolvem outros indivíduos, a saber: o indivíduo que representa, no modelo, o “Amuleto 115”, o indivíduo que representa as formas do tipo “Gato”, e o indivíduo que representa o

“Banco de dados da dissertação”. Esta capacidade de inferência, conforme explorado ao longo desta pesquisa, é o grande diferencial que separa as ontologias de outros artefatos de representação, validando o raciocínio humano da etapa de Conceitualização.

Outra representação de instância individual consta na Figura 19, que traz a representação formal do indivíduo “Gato”, instância de *MP8* que representa a forma de gato que alguns amuletos funerários presentes na pesquisa de Adaime (2021) apresentavam.

Figura 19: representação do indivíduo “Gato”

The screenshot displays the OWL2Viz interface for the ontology 'MPDiss1'. The left pane shows a class hierarchy where 'MP8\_Forma\_do\_amuleto' is a subclass of 'E28\_Conceptual\_Object'. The middle pane shows the description of the individual 'Gato', which is an instance of 'MP8\_Forma\_do\_amuleto'. The right pane shows property assertions for 'Gato', including 'P21\_is\_type\_of Representação\_do\_Amuleto\_155', 'P67\_refers\_to Forma\_de\_Gato', and 'P21\_is\_type\_of Representação\_do\_Amuleto\_115'. The bottom right pane shows data property assertions, including 'P141i\_was\_assigned\_by Categorização\_dos\_amuletos\_de\_acordo\_com\_seu\_tipo\_de\_forma' and 'P42i\_was\_assigned\_by Categorização\_dos\_amuletos\_de\_acordo\_com\_seu\_tipo\_de\_forma'.

Fonte: elaborado pelo autor (2023)

Na Figura 19, podemos ver, do lado esquerdo, tanto a taxonomia de classes na qual o *MP8* Forma do amuleto se localiza quanto, abaixo, os indivíduos que são instâncias de *MP8*, tais como “Gato”. Quanto aos relacionamentos do indivíduo “Gato”, podemos observar que houve 3 declarações, duas das quais envolvem a propriedade *P2i is type of*. Estas declarações denotam que “Gato” serve como tipo para dois indivíduos, “Representação do amuleto 115” e “Representação do amuleto 155”, obedecendo à restrição de cardinalidade apresentada na Figura 15 que permite que instâncias de *MP8 Forma do amuleto* sirvam como tipo a uma ou mais instâncias de *MP17 Representação do amuleto no banco de dados*. A terceira declaração formal na Figura 19 é a que faz uso da propriedade *P67 refers to* para representar que o conceito representado pelo indivíduo “Gato” se refere a outro indivíduo “Forma de gato”, que é instância de *MP9 Forma real do amuleto*. Além destas três declarações formais, “Gato” ainda está presente em dois relacionamentos inferidos via as declarações formais de outros indivíduos. A partir do que



é declarado acerca do indivíduo “Animais”, instância de *MP22 Tipo de Forma do Amuleto*, infere-se que o conceito representado por “Gato” tem “Animais” como conceito mais geral, via a propriedade *P127 has broader term*. Finalmente, a relação marcada pela propriedade *P42i was assigned by* (e sua superior, *P141i was assigned by*) infere que o tipo “Gato” foi atribuído por um evento de “Categorização dos amuletos de acordo com seu tipo de forma”, que é instância de *E17 Type assignment*.

Outro exemplo de modelagem de indivíduo é apresentada na Figura 20, que traz a representação formal do indivíduo “Amuleto 115”, instância de *MP1 Amuleto*, que foi citada no parágrafo anterior, e que representa o amuleto real, o objeto material ao qual foi atribuído o número 115, conforme será exposto a seguir.

**Figura 20:** representação do indivíduo “Amuleto 115”

The screenshot shows the OntoGraf interface for the individual 'Amuleto\_115'. The top navigation bar includes tabs for 'Class hierarchy: MP1\_Amuleto', 'Annotations: Amuleto\_115', 'Description: Amuleto\_115', and 'Property assertions: Amuleto\_115'. The left sidebar shows a class hierarchy starting with 'E77\_Persistent\_Item' and 'E71\_Human-Made\_Thing', leading to 'MP1\_Amuleto'. Below this, 'Direct instances: Amuleto\_115' lists several instances, with 'Amuleto\_115' selected. The main area displays 'Annotations: Amuleto\_115' with the label 'Amuleto 115'. The 'Description: Amuleto\_115' section shows the type 'MP1\_Amuleto'. The 'Property assertions: Amuleto\_115' section lists several object property assertions, including 'P56\_bears\_feature Forma\_de\_Gato', 'P45\_consists\_of Vidro', 'P108i\_was\_produced\_by Criação\_do\_Amuleto\_115', 'P67i\_is\_referred\_to\_by Representação\_do\_Amuleto\_115', 'P12i\_was\_present\_at Criação\_do\_Amuleto\_115', 'P1\_is\_identified\_by 115', 'P31i\_was\_modified\_by Criação\_do\_Amuleto\_115', 'P46\_is\_composed\_of Forma\_de\_Gato', 'P53\_has\_former\_or\_current\_location Enterramento\_38', and 'P92i\_was\_brought\_into\_existence\_by Criação\_do\_Amuleto\_115'.

Fonte: elaborado pelo autor (2023)

Na Figura 20, além das outras instâncias de *MP1 Amuleto*, podemos observar os relacionamentos dos quais “Amuleto 115” faz parte. Inicialmente, observa-se que o indivíduo está inserido em quatro declarações formais de relacionamentos. Pela propriedade *P56 bears feature*, que liga um objeto físico a uma característica física que ele possua, declara-se que “Amuleto 115” possui uma “Forma real de gato”, instância de *MP9 Forma real do amuleto*. Em seguida, a propriedade *P45 consists of* conecta o “Amuleto 115” ao material do qual ele é feito, representado pelo indivíduo “Vidro” (instância de *E57 Material*). A propriedade *P108i was produced by* liga o indivíduo “Amuleto 115” ao evento de produção do objeto material que ele representa; este evento é representado pelo indivíduo “Criação do amuleto 115”, instância de *MP4 Criação do*

*amuleto*. Finalmente, a propriedade *P67i is referred to by* informa que “Amuleto 115” está relacionado a uma instância de *MP17 Representação do amuleto no banco de dados*, o indivíduo “Representação do amuleto 115”, que representa o construto informacional criado por Adaime (2021) para a construção de seu banco de dados e que tem por referente o amuleto real. O restante dos relacionamentos de “Amuleto 115” é manifestado por inferências; a maioria destas inferências surge como reflexo dos relacionamentos declarados, e se trata da aplicação de super-propriedades a “Amuleto 115” e ao outro componente dos relacionamentos declarados. As duas inferências restantes surgem daquilo que foi declarado sobre outros indivíduos: a propriedade *P1 Is identified by liga* “Amuleto 115” ao número que identifica o amuleto real, o indivíduo “115”, instância de *E42 Identifier*; enquanto a propriedade *P52 has former or current location* ligas “Amuleto 115” ao seu local de origem, representado pelo indivíduo “Enterramento 38”, instância de *MP5 Enterramento*.

#### 6.5.6. Sétimo passo: definição das propriedades das relações ontológicas

O sétimo e último passo da Etapa de Formalização da ontologia consiste na definição das propriedades descritivas e lógicas de cada relação presente no modelo. As propriedades descritivas incluem aquelas apresentadas no segundo passo desta etapa da metodologia (Subseção 6.5.2.), enquanto as propriedades lógicas incluem noções como transitividade, reflexividade e simetria das propriedades, por exemplo.

Como apresentado no terceiro passo desta Etapa de Formalização (Subseção 6.5.3), a Declaração do CIDOC-CRM 7.1.1. se apresenta como um modelo de implementação e não como um guia com restrições; isto advém, principalmente, do tipo de informação com a qual o modelo lida, que inclui informações muitas vezes incertas ou mesmo contraditórias sobre objetos, herança e patrimônio culturais (CIDOC, 2021). Por esse motivo, o modelo trabalha com a noção de mundo-aberto (ou seja, a própria natureza de seu domínio frequentemente possui lacunas que, por um lado, trazem incerteza, mas por outro, permitem maior flexibilidade ao esforço de representação). Dessa forma, a Declaração recomenda<sup>42</sup> que suas propriedades sejam declaradas com a quantificação *many to many* (0,n:0,n), que autoriza que as propriedades sejam tratadas

---

<sup>42</sup> “[...] *Quantifiers for properties are provided for the purpose of semantic clarification only, and should not be treated as implementation recommendations. The CIDOC CRM has been designed to accommodate alternative opinions and incomplete information, and therefore all properties should be implemented as optional and repeatable for their domain and range (“many to many (0,n:0,n)”*). Therefore, the term “cardinality constraints” is avoided here, as it typically pertains to implementations [...]” (CIDOC, 2021, p. 23).

como não-repetíveis tanto para seu domínio quanto para seu *range* (*Ibid*, 2021); além disso, as regras de atributos como reflexividade e simetria, embora constem nas declarações das propriedades, também se tornam adaptáveis conforme a necessidade da proposta de modelagem.

Esta flexibilidade oferecida pela Declaração oferece uma saída para um problema encontrado durante este passo da etapa de formalização da ontologia. Assim que declaramos que as propriedades *P46 is composed of*, *P89 falls within* e *P127 has broader term* eram transitivas, o Protegé detectou inconsistências junto às classes *MP5 Enterramento* e *MP8 Forma do amuleto* sem, no entanto, explicar com clareza qual o problema estava ocasionando esta inconsistência. Após algumas tentativas, identificamos que o problema desaparecia quando deixávamos de marcar as propriedades *P46*, *P89* e *P127* como sendo transitivas. Assim, considerando a recomendação apresentada pela Declaração do CIDOC-CRM, mantivemos essas propriedades sem indicação de transitividade, de forma que a ontologia manteve-se consistente e coerente. À exceção destas três, todas as outras propriedades foram modeladas conforme o que indica a Declaração do CIDOC-CRM, incluindo no que tange atributos como simetria e reflexividade. Quanto ao atributo de transitividade, a Declaração recomenda representar todas as propriedades como intransitivas salvo onde informada a exceção, e seguimos esta recomendação ao longo desta etapa.

## 6.6 Etapa 6: Avaliação da ontologia

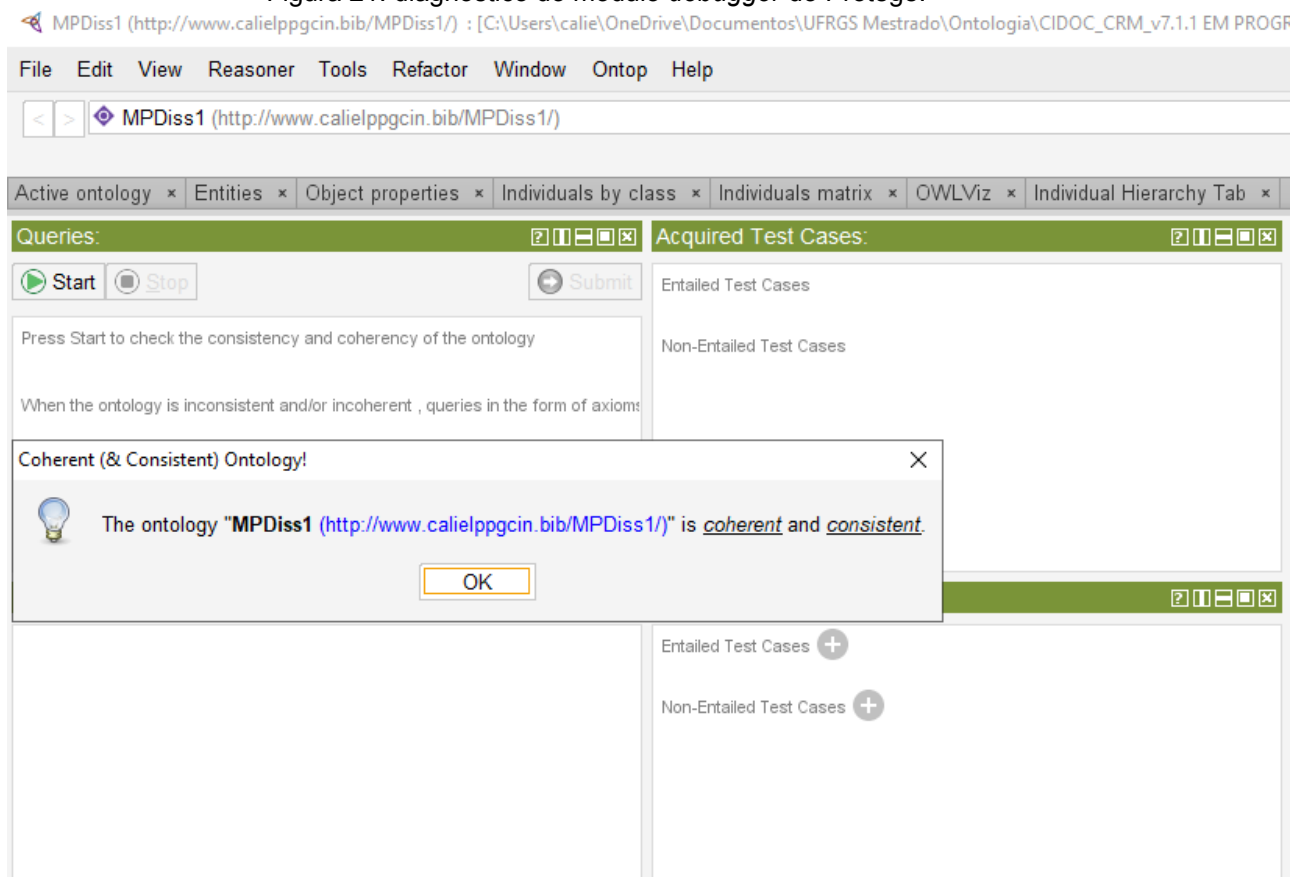
A sexta etapa da metodologia *OntoForInfoScience* consiste na avaliação da ontologia construída. Mendonça (2015) afirma que é nesta etapa que ocorre a checagem das estruturas ontológicas desenvolvidas, bem como a observação de potenciais lacunas onde possam haver melhorias em futuras iterações da ontologia, visando sempre a representação adequada do domínio real que se buscou modelar.

Para a realização do esforço de avaliação, Mendonça (2015) oferece alguns critérios, divididos em duas fases distintas: uma fase de validação (ou seja, de adequação ao domínio) e outra de verificação (ou seja, de corretude ontológica). Além disso, os diferentes conjuntos de critérios dentro de cada fase são pautados por parâmetros que avaliam diferentes aspectos gerais da ontologia.

Antes de apresentar os critérios de avaliação, cabe reforçar o ponto trazido por Mendonça de que diferentes programas de construção de ontologias, como o Protégé,

possuem de dispositivos internos próprios para verificação da consistência e coerência dos modelos neles desenvolvidos. Essas verificações são feitas a partir dos motores de raciocínio (*reasoners*), responsáveis por identificar as inferências que surgem a partir do que é declarado na ontologia. O Protegé 5.5, utilizado nesta pesquisa, faz uso por padrão do *reasoner* Hermit 1.4.3.456, inclusive em seu módulo de *debugging*, que diagnostica erros e inconsistências na construção do artefato ontológico. Nesse sentido, apresentamos na Figura 21 o diagnóstico do módulo *debugger* do Protegé, atestando que a ontologia desenvolvida é coerente e consistente, mesmo se tratando de um modelo híbrido, integrando as classes, indivíduos e relacionamentos desenvolvidos durante esta pesquisa ao modelo já existente do CIDOC-CRM 7.1.1.

Figura 21: diagnóstico do módulo *debugger* do Protegé.



Fonte: elaborado pelo autor (2023)

A seguir, apresentamos a tabela contendo os critérios de avaliação, bem como os resultados da aplicação dos mesmos ao modelo desenvolvido nesta pesquisa. Junto da célula que contém a análise, foi deixada uma pergunta que sumariza o que está sendo avaliado.

**Quadro 2:** critérios para a avaliação do modelo proposto

Critérios de validação: adequação do domínio		
Parâmetro de avaliação	Critério avaliativo	Análise do modelo proposto
Compromisso ontológico	Fundamentação ontológica	<b>As classes utilizadas na ontologia de fundamentação foram importadas adequadamente?</b> No modelo, fizemos uso da estrutura ontológica oferecida pelo CIDOC-CRM 7.1.1; todas as entidades daquele modelo foram importadas e foi feito uso de algumas.
	Natureza ontológica	<b>As entidades representadas existem de fato como entes na realidade?</b> Considerando o escopo de nosso estudo (dados científicos produzidos e ordenados durante a construção de uma dissertação de mestrado na área das Ciências Humanas) e o olhar oferecido pelo modelo ontológico utilizado para fundamentação (perspectiva <i>event-centric</i> do CIDOC-CRM), consideramos que todas as entidades representadas existiram na realidade, ainda que em caráter abstrato/conceitual (na forma de eventos que se desenvolveram ao longo do tempo, por exemplo)
	Não-intuitividade	<b>As classes foram definidas de maneira rigorosa, sem depender da intuitividade humana?</b> Embora as entidades possuam definições concisas, em especial as que se derivam da taxonomia nativa do CIDOC-CRM, alguns relacionamentos entre as classes dependem da intuitividade humana para expressar seu sentido completo (em particular, a relação entre <i>MP22</i> , <i>MP8</i> e <i>MP17</i> ).
	Universalidade de classes	<b>Cada classe engloba todas as instâncias possíveis para si dentro do domínio?</b> Considerando que trabalhamos com um recorte do domínio (por exemplo, não modelamos todas as instâncias possíveis de amuleto ou de suas representações), as classes no modelo são universais pois permitem a representação total de todas as instâncias, havendo tempo-hábil e recursos para tanto.
	Universalidade de relações	<b>As relações que uma classe pode</b>

		<p><b>ter se aplicam a todas as suas instâncias?</b> Após análise, verificamos que todas as relações no modelo proposto possuem caráter de universalidade.</p>
	Formalização das definições das entidades	<p><b>Todas as classes e relações da ontologia possuem definição formal, legível por máquinas?</b> Todas as entidades no modelo possuem definição formal, seja oriunda do CIDOC-CRM ou construída pelo pesquisador durante o decorrer da pesquisa.</p>
	Definição de propriedades	<p><b>Foram definidas propriedades suficientes para o entendimento completo das classes e relações na ontologia?</b> Houve algumas dificuldades para modelar os atributos de algumas propriedades (a saber, <i>P46</i>, <i>P89</i> e <i>P127</i>), mas o uso do CIDOC-CRM prevê inconsistências nesse sentido. Além disso, não foi estritamente necessário declarar equivalências ou disjunções para o bom entendimento do modelo.</p>
Especificação	Resposta às questões de competência	<p><b>Através do que foi representado, a ontologia é capaz de responder às questões de competência?</b><sup>43</sup> Após análise, não identificamos questão de competência que não possa ser respondida a partir do que foi representado no arquivo da ontologia.</p>
	Grau de representatividade do domínio	<p><b>O esforço de representação é significativo considerando o escopo geral do domínio?</b> Considerando que o domínio é especificamente um recorte de uma dissertação de mestrado, focando nos dados produzidos na mesma, observa-se relevância dentro do que foi representado. No entanto, alguns dados potencialmente importantes que não constam em nosso modelo são passíveis de representação.</p>
Validação especializada	Consulta aos especialistas	<p><b>Houve participação de especialistas no processo de validação?</b> Considerando o domínio trabalhado, não houve consulta direta com especialistas. No entanto, tendo em vista a especificidade do domínio e da fonte de conhecimento principal,</p>

<sup>43</sup> Apresentadas na Etapa 1 da seção de Metodologia deste documento.

		acreditamos que o esforço de representação não sofreu maiores prejuízos.
	Grau de participação	<b>Qual o grau de participação dos especialistas na construção do modelo?</b> Similar ao critério anterior, a ausência de especialistas pode ter implicado em alguma falha de representação, mas não identificamos inconsistências nesse sentido.
Expandibilidade	Capacidade de atualização	<b>Considerando o domínio, a documentação da ontologia prevê possíveis atualizações?</b> Se tratando da primeira versão do modelo, criada especificamente para fins de experimentação e desenvolvimento desta pesquisa, a documentação existente permite a atualização do modelo, se assim for necessário.
	Capacidade de integração	<b>Considerando o grau de reutilização envolvido em sua construção, a ontologia pode ser facilmente integrada em outros modelos?</b> Embora faça uso da estrutura pré-existente do CIDOC-CRM, consideramos que o modelo proposto nesta pesquisa ainda não possui grau de integração alto o suficiente para ser utilizado de maneira segura como componente de outras ontologias. Isto se dá devido ao caráter experimental que o modelo possui, bem como devido a se tratar da primeira iteração do modelo em si.
<b>Critérios de verificação: corretude ontológica</b>		
<b>Parâmetro de avaliação</b>	<b>Critério avaliativo</b>	<b>Análise do modelo proposto</b>
Compleitude	Única definição textual	<b>Cada entidade possui uma definição textual única?</b> Após análise, verifica-se que todas as entidades possuem uma definição textual única para si, salvo por aqueles indivíduos cuja classe possui em sua definição informações suficientes para permitir a compreensão dos indivíduos de maneira informal.
	Única definição formal	<b>Cada entidade possui uma definição formal única?</b> Após análise, verifica-se que todas as entidades possuem uma definição formal única para si.

	Definição de propriedades descritivas	<b>As propriedades descritivas estão presentes junto às entidades?</b> Todas as entidades no modelo possuem pelo menos uma propriedade descritiva ( <i>annotation</i> ).
	Caracterização das propriedades básicas	<b>Os atributos básicos de cada relação foram caracterizados no modelo?</b> A maioria dos atributos foram caracterizados, dentro do recorte de relações que utilizamos em nosso modelo. As exceções são as propriedades <i>P46</i> , <i>P89</i> e <i>P127</i> .
Integridade	Conectividade com o domínio e contexto	<b>Todas as entidades na ontologia possuem conexão com o domínio do conhecimento que se buscou representar?</b> Após análise, verificamos que o modelo proposto nesta pesquisa se caracteriza bem conectado ao domínio do conhecimento que visamos modelar.
	Definição de equivalência	<b>Foram definidas como equivalentes entidades que possuíssem tal característica?</b> O modelo não possui entidades que compartilhem de equivalência entre si.
	Definição de disjunção	<b>Foram definidas como disjuntas entidades que possuíssem tal característica?</b> As relações de disjunção foram feitas apenas nos casos onde pudesse haver equívocos entre as entidades envolvidas no relacionamento.
Integridade e Completude	Definição de domínio e imagem da relação	<b>Para cada relação, foram declaradas quais classes podem lhe servir de domínio e de imagem (<i>Range</i> ou objeto)?</b> Todas as relações no modelo possuem <i>domain</i> e <i>range</i> estabelecidas pela Declaração do CIDOC-CRM 7.1.1.
	Definição de relação inversa	<b>Toda relação tem seu inverso declarado?</b> Todas as relações no modelo possuem seus relacionamentos inversos declarados pela Declaração do CIDOC-CRM 7.1.1.
	Definição de tipos	<b>Foram declarados os tipos de dados aceitos nas propriedades de dados usadas na ontologia?</b> Os tipos de dados aceitos pelas <i>data</i>



		<i>properties</i> utilizadas pelo modelo foram definidos.
	Criação de cardinalidade	<b>Foram definidas as cardinalidades em cada relação ontológica e para cada classe que lide com propriedade de dados?</b> As restrições de cardinalidade foram declaradas junto à declaração formal das classes utilizadas no modelo.
Consistência	Não-polissemia	<b>Existem termos polissêmicos na ontologia construída?</b> Após análise, não localizamos ocorrência de polissemia no modelo proposto.
	Ausência de ciclos na hierarquia	<b>Existe relação na ontologia em que a classe “A” seja tanto superclasse quanto subclasse de “B”?</b> Após análise, não localizamos concorrência de ciclos de hierarquia no modelo proposto.
	Ausência de classes variadas ou genéricas	<b>Existe alguma classe que reúna conceitos díspares sob um mesmo rótulo?</b> Nenhuma classe representa dois ou mais conceitos diferentes sob o mesmo rótulo dentro do modelo proposto.
	Não recursividade na definição	<b>Existe algum termo que faça uso de si mesmo para poder ser definido?</b> Não há casos de recursividade estrita no modelo proposto.
	Consistência na definição da relação inversa	<b>Há certeza de que as relações declaradas inversas estejam corretas?</b> A consistência nas definições das relações e de seus inversos é garantida pela Declaração do CIDOC-CRM 7.1.1, onde se originam, bem como pelos recursos de checagem de coerência embutidos no Protegé 5.5.
Precisão	Definição de sinônimos nas classes	<b>Não há sinônimos dispersos em diferentes classes, que poderiam ser agrupados sob uma única classe?</b> Não foram identificadas ocorrências de sinônimos no modelo proposto.
	Unicidade dos rótulos (identificadores) das classes	<b>Cada classe possui um rótulo único, sem repetição de rótulos?</b> Não há rótulos (representados pela propriedade descritiva <i>label</i> ) em

		repetição no modelo proposto.
	Não criação de instâncias como classes	<p><b>Não existem classes que poderiam ter sido modeladas apenas como instâncias de outra classe já existente, em vez de serem categorias em si?</b></p> <p>Há um único caso que pode se encaixar neste critério: a classe <i>MP10 Grupo responsável pela escavação</i> possui uma única instância (o indivíduo "Gurob Harem Palace Project", de forma que seria possível representar o indivíduo como instância direta da classe <i>E74 GROUP</i>, se preferível.</p>
	Especificação dos diferentes tipos da relação "is_a"	<p><b>Todos os sentidos diferentes para a relação "is_a" estão presentes na ontologia?/Qual o sentido de "is_a" adotado na ontologia?</b></p> <p>O modelo entende a relação "is_a" no sentido de "SubclassOf", conforme seu uso no ambiente do <i>software</i> Protegé. Outros sentidos de "is_a" incluem o de definição de instâncias (igualmente baseado no seu funcionamento no <i>software</i> Protegé) e o de declaração de sinonímia (que não ocorreu em nosso modelo).</p>
	Especificação dos diferentes tipos da relação "part_of"	<p><b>Todos os sentidos diferentes para a relação "part_of" estão presentes na ontologia?/Qual o sentido de "part_of" adotado na ontologia?</b></p> <p>Os diferentes sentidos possíveis para a relação "part_of" (partitude, partitude temporal, etc.) são representados por diferentes relações do modelo CIDOC-CRM, como <i>P5i forms part of</i>, <i>P9i forms part of</i>, etc.</p>
Documentação	Existência de um documento formal da ontologia	<p><b>Há um documento formal que descreva a ontologia?</b></p> <p>O arquivo da ontologia será disponibilizado ao final da pesquisa.</p>
	Único critério de nomeação dos termos da ontologia	<p><b>Foi adotado um critério único para a nomeação dos termos da ontologia?</b></p> <p>Embora não tenhamos adotado um critério específico, a maioria das entidades segue o padrão de uso de letras iniciais maiúsculas, seguido por letras minúsculas. Além disso, expressões compostas foram separadas com <i>underlines</i> quando transplantadas para o ambiente Protegé, como é usual com este</p>

		<i>software.</i>
	Anotação de comentários	<p><b>Foram feitos comentários específicos nos campos de annotations?</b></p> <p>Optamos por não fazer uso dos campos de comentários para outros fins que não o de definição das entidades, seguindo o que vinha sendo feito já no arquivo do CIDOC-CRM. Ainda assim, alguns comentários feitos ao longo deste documento podem ser passíveis de inclusão no arquivo da ontologia em iterações futuras.</p>

Fonte: elaborado pelo autor (2023), a partir do modelo proposto por Mendonça (2015).

### 6.7 Etapa 7: Documentação da ontologia

A sétima etapa da metodologia *OntoForInfoScience* consiste na verificação da documentação da ontologia construída, que deve ser organizada em um documento formalizado, em linguagem natural. Mendonça (2015) afirma que esta etapa é indispensável em um projeto de construção de ontologia, pois registra não apenas as alterações e procedimentos que ocorreram ao longo do projeto, mas também de que maneira e por qual motivo cada passo foi dado.

Para auxiliar nesta etapa, Mendonça (2015) montou um quadro-modelo com o tipo de documentação que se espera, idealmente, ter sido produzida em cada etapa, podendo ser reunida em um único documento formal. A seguir, apresentamos o quadro-modelo e a documentação resultante do desenvolvimento desta pesquisa.

**Quadro 3:** documentação para cada etapa da metodologia

Etapa 1: Documento de especificação	O documento foi produzido?	Observação
Domínio/escopo da ontologia	Sim	Documentos reunidos no <b>Quadro 1:</b> especificações do modelo ontológico proposto, na Seção 6.2.1 deste documento.
Propósito geral	Sim	
Classes de usuários	Sim	
Uso pretendido	Sim	
Tipo da ontologia	Sim	
Grau de formalidade	Sim	

Questões de competência	Sim	
<b><u>Etapa 2: Documentos de referência</u></b>	O documento foi produzido?	<b>Observação</b>
Relação dos documentos específicos do domínio utilizados no trabalho.	Não	Considerando que nosso documento-fonte foi a própria dissertação de mestrado de Adaime (2021), não foi elencada relação de documentos de domínio utilizados.
<b><u>Etapa 3: Modelos conceituais</u></b>	O documento foi produzido?	<b>Observação</b>
Conjunto de modelos conceituais que representam o domínio.	Sim	Corresponde ao mapa conceitual apresentado na Seção 6.4 e disponível no Apêndice A deste documento.
<b><u>Etapa 4: Ontologias reutilizadas</u></b>	O documento foi produzido?	<b>Observação</b>
Apresentação das ontologias que serviram de fundamentação para o desenvolvimento do novo modelo.	Parcialmente.	Ao longo da Seção 6.4, foram apresentados os principais pontos do CIDOC-CRM 7.1.1, modelo ontológico que serviu de fundamentação para esta pesquisa. Não havendo outras ontologias de fundamentação em uso nesta pesquisa, não foram elencados outros modelos.
<b><u>Etapa 5: Conteúdo ontológico</u></b>	O documento foi produzido?	<b>Observação</b>
Taxonomia geral da ontologia	Sim	A taxonomia-geral pode ser observada a partir do arquivo da ontologia, a ser disponibilizado na Seção 6.8.
Dicionário de classes	Sim	Os Apêndices B e C servem como dicionários de classes, trazendo tanto as definições textuais das classes nativas do CIDOC-CRM quanto das classes elaboradas para esta pesquisa.
Dicionário de relações ontológicas	Sim	O apêndice D serve como dicionário e relações; no entanto, cabe ressaltar que nele constam apenas as relações que foram utilizadas em nosso modelo, e não todas as relações presentes no CIDOC-CRM.
Estruturas gráficas de representação	Parcialmente	Embora tenha sido apresentada uma taxonomia parcial da

		ontologia na Seção 6.5, não foi desenvolvida uma representação gráfica geral do modelo, devido ao volume de entidades que nele existem. Tal representação pode ser gerada via módulos específicos do Protegé.
<b>Etapa 6: Métricas de avaliação</b>	<b>O documento foi produzido?</b>	<b>Observação</b>
Conjunto de critérios utilizados para avaliação do conteúdo ontológico	Sim	O quadro com os critérios de avaliação foi apresentado na Seção 6.6.
Conjuntos de respostas para as questões de competência	Não	Não foi elaborado um documento formal contendo as respostas, mas, durante a avaliação feita Seção 6.6, não identificamos inconsistência que impedisse a resolução das mesmas.

Fonte: elaborado pelo autor (2023), a partir do modelo proposto por Mendonça (2015).

### 6.8 Etapa 8: Disponibilização da ontologia

A oitava e última etapa da metodologia *OntoForInfoScience* consiste na disponibilização da ontologia construída em algum formato e canal que facilite seu acesso por usuários interessados. Para tanto, Mendonça (2015) elenca dois passos para facilitar esta disponibilização.

O primeiro passo consiste em gerar a ontologia em uma linguagem lógica, que permita sua fácil interpretação por máquinas. Nesse sentido, Mendonça (2015) afirma que podemos optar por uma linguagem lógica mais descritiva, ideal para o processamento por máquinas (como OWL), ou por uma linguagem lógica mais complexa, como as de Lógica de Primeira Ordem, que permite declarações semanticamente mais ricas. Considerando que o modelo proposto se trata de um experimento realizado durante uma pesquisa de mestrado, optamos por disponibilizar a ontologia em formato de arquivo OWL, facilitando sua edição e verificação pela banca avaliadora.

O segundo passo consiste em apresentar a ontologia em algum formato eletrônico de fácil acesso, preferencialmente já como recurso na *Web*. Mendonça recomenda que isto seja feito via o próprio *software* de construção de ontologias, a fim de se criar uma *homepage* que faça uso dos URIs da ontologia desenvolvida. Em nosso caso, se tratando de um modelo experimental, optamos por não disponibilizar o arquivo amplamente via a construção de uma *homepage*; em vez disso, o arquivo foi depositado junto ao *website*

*Figshare*, que permite a hospedagem de arquivos em diferentes formatos, especificamente voltado a arquivos resultantes do desenvolvimento de trabalhos acadêmicos e de pesquisa. O arquivo pode ser acessado através [deste link](#)<sup>44</sup>.

## 7 CONSIDERAÇÕES FINAIS

Ao término desta pesquisa, observamos tanto resultados importantes quanto lacunas que podem ser preenchidas em novas iterações deste trabalho ou mesmo em pesquisas futuras. A intersecção de diferentes áreas do conhecimento inerente ao desenvolvimento de ontologias, como a Terminologia e a Organização do Conhecimento, somou-se às diferentes bases teóricas que adotamos. O domínio escolhido (dados de pesquisa produzidos pela pesquisadora Victoria Arroyo Adaime, na área da Arqueologia), a fundamentação ontológica utilizada (o modelo CIDOC-CRM), e a metodologia adotada (*OntoForInfoScience*) foram utilizados em consonância a fim de organizar nosso universo de conhecimento e desenvolver nosso modelo, em uma conjunção de fatores que culminou no arquivo da ontologia construída durante a pesquisa.

A partir destes três pilares teóricos e metodológicos, buscamos alcançar nosso objetivo geral, construindo um modelo representacional ontológico que desse conta de, primeiro, organizar o universo de conhecimento e, segundo, representar esse universo organizado em linguagem legível por máquinas. O êxito na busca deste objetivo se deu parcialmente: nossos produtos, o modelo conceitual e a ontologia em si, possuem caráter experimental, e ainda não dispõem de maturidade suficiente para servirem como modelo-padrão, e tampouco oferecem diretrizes formalizadas o bastante para se propor a ser um requisito para trabalhos vindouros.

Nosso esforço resultou, no entanto, em uma modelagem de domínio que possui valor como um protótipo, um experimento de construção de ontologias, modelando um domínio cuja representação em ontologias ainda é bastante incipiente. Os resultados desta pesquisa dão margem para futuras investigações quanto à modelagem conceitual dos dados científicos das Humanidades e para a construção de novos artefatos ontológicos que contemplem este domínio. Dados científicos já são substrato de muitos empreendimentos de construção de ontologias, inseridos no contexto maior de valorização dos dados abertos de pesquisas. Este contexto promove a criação de

---

44

[https://figshare.com/articles/dataset/Arquivo\\_da\\_ontologia\\_desenvolvida\\_na\\_disserta\\_o\\_Ontologias\\_como\\_artefatos\\_para\\_a\\_representa\\_o\\_de\\_dados\\_abertos\\_de\\_pesquisa\\_uma\\_explora\\_o\\_acerca\\_da\\_modelagem\\_de\\_dados\\_da\\_Arqueologia\\_atrav\\_s\\_do\\_modelo\\_CIDOC\\_Conceptual\\_Reference\\_Model\\_/22285885](https://figshare.com/articles/dataset/Arquivo_da_ontologia_desenvolvida_na_disserta_o_Ontologias_como_artefatos_para_a_representa_o_de_dados_abertos_de_pesquisa_uma_explora_o_acerca_da_modelagem_de_dados_da_Arqueologia_atrav_s_do_modelo_CIDOC_Conceptual_Reference_Model_/22285885)

protocolos e ambientes próprios para o tratamento dos dados, o que tanto favorece quanto é beneficiado pelos esforços de representação destes dados. Ainda assim, a modelagem de dados oriundos das Humanidades, onde há maior importância na contextualização do dado bruto, ainda é pouco difundida, especialmente no contexto brasileiro. Nesse sentido, nosso trabalho surge como uma tentativa de oferecer alguma contribuição a este segmento de pesquisa. Acreditamos que nosso esforço de modelagem tanto de um recorte dos dados do domínio explorado quanto do contexto em que estes dados estão inseridos se apresenta como uma contribuição importante, ainda que careça de maiores experimentos e de novas iterações para colocar à prova sua eficiência e eficácia naquilo que se propõe.

De fato, o caráter experimental de nosso trabalho é a raiz de muitas das lacunas observadas durante esta pesquisa. Inicialmente, uma maneira mais assegurada de verificar a efetividade de um modelo ontológico como o que foi proposto em nosso trabalho seria a de se construir uma ontologia a partir do zero absoluto. Nossa dependência do modelo CIDOC-CRM e suas particularidades, embora tenha nos servido imensamente para fins de praticidade e economia de tempo, pode ter nos limitado frente a todo potencial de desenvolvimentos que poderiam ocorrer em uma pesquisa do tipo. O procedimento mais tradicional de construção de ontologias, pautado por diferentes ontologias de fundamentação, pode se mostrar mais efetivo na representação da dupla dados/contexto, que buscamos realizar em nossa pesquisa. Igualmente, este tipo de desenvolvimento tradicional permitiria maior aderência às diretrizes propostas pela metodologia *OntoForInfoScience*, muitas das quais tivemos de adaptar, tanto devido ao domínio que escolhemos modelar quanto ao modelo ontológico com o qual nos propusemos a trabalhar. Também o trabalho terminológico poderia ter sido realizado com maior cautela, dada sua imprescindibilidade a qualquer esforço de organização de conhecimento. Apesar de não ter prejudicado nosso resultado final, alguns termos podem produzir ambiguidade na interpretação, caso não houvesse os Apêndices para fins de verificação.

Apesar das lacunas que identificamos ao longo de nosso trabalho, o uso que fizemos do CIDOC-CRM permitiu testar até que ponto sua proposta de modelagem dos domínios de patrimônio e herança cultural podia ser adaptada para a modelagem de dados de pesquisa das Humanidades, bem como o contexto em que estes se encontram inseridos. Dessa forma, a aplicação do CIDOC-CRM que deu origem à nossa modelagem

conceitual oferece algum substrato para pesquisas vindouras, bem como para futuras interfaces entre as Humanidades e a representação dos dados de pesquisa que delas advêm. Também a documentação elaborada durante nossa pesquisa serve como material para análise em pesquisas diversas, ou novas iterações do modelo que desenvolvemos no presente trabalho. O Apêndice A permite visualizar como se deu a modelagem conceitual de nosso domínio, enquanto os Apêndices B, C e D expõem as entidades que figuram em nosso modelo e no resultante arquivo da ontologia, cuja análise e avaliação servirão para aprimorar os procedimentos de representação que nos propusemos a realizar.

Em última instância, este trabalho se apresenta como um experimento em modelagem conceitual e representação formal de um domínio com idiosincrasias próprias, o que fortalece seu caráter de lançar uma perspectiva nova em estudos do tipo. Nossos objetivos específicos foram alcançados, com o desenvolvimento de um artefato ontológico funcional capaz de auxiliar na recuperação de informação especializada que necessite de restrições de domínio e conceitualização dos entes envolvidos. Para além do desenvolvimento metodológico da pesquisa, acreditamos ainda que ela tenha servido como exercício de aproximação teórica. Acreditamos, nesta dissertação, termos sido capazes de promover um diálogo entre a dinâmica de produção de dados nas Humanidades, as particularidades de modelagem do modelo CIDOC-CRM e a metodologia *OntoForInfoScience*, um método para a construção de ontologias que tem a peculiaridade de ser voltada ao cientista da informação interessado neste tipo de atividade. A partir desta interface, pudemos concluir como as ontologias possuem potencial como ferramentas de representação do conhecimento resultante das iniciativas de investigação científica. Em nossa pesquisa, pudemos modelar conceitualmente tanto o conhecimento finalizado contido na pesquisa de Victoria Adaime quanto o conhecimento trazido pelos dados produzidos durante seu trabalho. Dessa forma, esperamos ter conseguido contribuir para os avanços no que se refere à aplicação das ontologias junto à representação de dados de pesquisa enquanto elementos próprios, que carregam por si só conhecimento e potencial científico.

Em resposta ao problema que motivou esta pesquisa, acreditamos haver potencial para que as ontologias sirvam como instrumentos de facilitação da interoperabilidade entre instituições e repositórios que trabalhem com dados oriundos de investigações nas áreas das Humanidades. A partir da padronização do sentido e da facilitação do



intercâmbio de conhecimento promovidas pelo uso das ontologias, a recuperação da informação especializada, quer seja em sua forma finalizada ou ainda no escopo dos dados de pesquisa, é facilitada, aumentando o valor da construção e manutenção de repositórios especializados no tratamento e disponibilização dos dados de pesquisa. E, em áreas tão díspares, mas ainda tão próximas, como são as Humanidades, o consenso no sentido dos saberes especializados é tarefa árdua, envolvendo desde debates intra e interdisciplinares até problemáticas de caráter terminológico. O mesmo pode ser dito quanto aos dados produzidos pelas pesquisas destas áreas, e é nesse panorama que as ontologias podem beneficiar o desenvolvimento de novas investigações nessas disciplinas. A construção de novos padrões de metadados capazes de representar o conhecimento especializado contido nos conjuntos de dados, a representação adequada do sentido dos conceitos e a construção de sistemas capazes de estruturar esses conceitos e seus relacionamentos aumenta o valor dos dados produzidos durante as investigações científicas. Havendo consenso entre o significado dos conceitos e como eles se relacionam, pesquisadores de áreas correlatas podem compartilhar dados entre si, dando origem a novas investigações que façam uso destes conjuntos de dados e justificando aos mantenedores de suas pesquisas a importância da construção de repositórios capazes de abarcar estes dados e oferecê-los de maneira adequada à comunidade acadêmica. E, como esta pesquisa pode atestar, mesmo o esforço de modelagem conceitual dos dados é uma tarefa complexa, perpassando desde procedimentos de mapeamento do domínio que abarca estes até a obtenção dos termos que melhor expressem o sentido que busca-se representar no artefato a ser desenvolvido. Talvez devido justamente a essa complexidade, as conclusões obtidas após esse esforço de representação e construção de ontologia podem trazer *insights* com algum valor para iterações futuras de artefatos de representação e experimentos de modelagem similares. Como nota de encerramento, resta a expectativa que este trabalho, com seus êxitos e suas lacunas, sirva como substrato para pesquisas futuras que promovam o avanço das disciplinas que compõem a Organização e Representação do Conhecimento.

## REFERÊNCIAS

5 STAR OPEN DATA. Página inicial. 2012. Disponível em: <https://5stardata.info/en/>. Acesso em: 10 fev. 2022.

ABADAL, Ernest. Ciencia abierta: un modelo con piezas por encajar. **Arbor**, [s. l.], v.197, n. 799, 2021. Disponível em: <https://arbor.revistas.csic.es/index.php/arbor/article/view/2403>. Acesso em: 04 fev. 2022.

ADAIME, Victoria Arroyo. **Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário**: um estudo sobre o papel dos amuletos. 2021. Dissertação (Mestrado em Arqueologia) - Museu de Arqueologia e Etnologia, Universidade de São Paulo, São Paulo, 2021. Disponível em: <https://doi.org/10.11606/D.71.2021.tde-12072021-151821>. Acesso em: 2023-07-11.

AGANETTE, Elisângela Cristina; ALMEIDA, Maurício Barcellos. Avanços teórico-empíricos na comparação entre Teoria da Ontologia Aplicada e Teoria da Terminologia. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 20, n. 44, p. 3- 24, set./dez., 2015. ISSN 1518-2924. Disponível em: <https://doi.org/10.5007/1518-2924.2015v20n44p3>. Acesso em: 08 jun. 2023.

ALMEIDA, Maurício Barcellos de. **Ontologia em Ciência da Informação**: teoria e método. Coleção Representação do Conhecimento em Ciência da Informação: volume 01. Curitiba: Editora CRV, 2020.

ALMEIDA, Maurício Barcellos. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 242-258, jul./set. 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35762>. Acesso em: 02 fev. 2022.

ALMEIDA, Maurício Barcellos; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ci. Inf.**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003. Disponível em: <https://doi.org/10.1590/S0100-19652003000300002>. Acesso em: 06 fev. 2022.

ALPSP; STM. **Databases, data sets, and data accessibility** – views and practices of scholarly publishers. [s. l.]: ALPSP; STM, 2006. Disponível em:

[https://www.stm-assoc.org/2006\\_06\\_01\\_STM\\_ALPSP\\_Data\\_Statement.pdf](https://www.stm-assoc.org/2006_06_01_STM_ALPSP_Data_Statement.pdf). Acesso em: 10 fev. 2022.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **ANSI/NISO Z39.19-2005 (R2010)**: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. *Online*, 2010. Disponível em: <http://www.niso.org/publications/ansiniso-z3919-2005-r2010>. Acesso em: 08 fev. 2022.

ARAKAKI, Ana Carolina Simionato; ARAKAKI, Felipe Augusto. Dados e metadados: conceitos e relações. **Ciência da Informação**, [S. l.] v. 49, n. 3, 2019. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/163406>. Acesso em: 13 out. 2022.

AUTRAN, Marynice de Medeiros Matos; BORGES, Maria Manuel. Comunicação da ciência: (r)evolução ou crise? **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, Rio de Janeiro, v. 8, n. 2, 2014. Disponível em: <http://dx.doi.org/10.3395/reciis.v8i2.621>. Acesso em: 10 out. 2022.

BARBOSA, Marcia Cristina Bernardes; PAVAN, Cleusa. Financiamento público no Brasil para a publicação de artigos em acesso aberto: alguns apontamentos. **Em Questão**, Porto Alegre, v. 23, n. 2, maio/ago. 2017. Disponível em: <http://hdl.handle.net/10183/159734>. Acesso em 04 fev. 2022.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. **The Semantic Web**. Arquivo PDF *Online*, 2001. Disponível em: [https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American\\_%20Feature%20Article\\_%20The%20Semantic%20Web\\_%20May%202001.pdf](https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf). Acesso em: 06 fev. 2022.

BIAGETTI, Maria Teresa. Ontologies (as knowledge organization systems) *In*: HJØRLAND, Birger; GNOLI, Claudio (ed.). **Encyclopedia of Knowledge Organization**. [S.l.]: International Society for Knowledge Organisation: 2020. Disponível em: <https://www.isko.org/cyclo/ontologies>. Acesso em: 10 out. 2022.

BRANDÃO, Anarosa Alves Franco; LUCENA, Carlos José Pereira de. **Uma introdução à engenharia de ontologias no contexto da web semântica**. Rio de Janeiro. Departamento de Informática, PUC-Rio, 2002. Disponível em: [http://bib-di.inf.puc-rio.br/ftp/pub/docs/techreports/02\\_29\\_brandao.pdf](http://bib-di.inf.puc-rio.br/ftp/pub/docs/techreports/02_29_brandao.pdf). Acesso em: 08 out. 2022.

BRUSEKER, George. **Learning Ontology & CIDOC CRM**. Arquivo PDF *online*, 2019. Disponível em: <https://www.archesproject.org/wp-content/uploads/2020/04/Learning-Ontology-CRM.pdf>. Acesso em: 18 ago. 2022.

BORGES, Maria Manuel; SANZ CASADO, Elias (coord.). **Sob a lente da Ciência Aberta:**

olhares de Portugal, Espanha e Brasil. Coimbra: Imprensa da Universidade de Coimbra; CEIS20, 2021. Disponível em: <http://hdl.handle.net/10316/93276>. Acesso em: 09 fev. 2022.

BUDAPEST OPEN ACCESS INITIATIVE. **Iniciativa de Budapeste pelo Acesso Aberto**. Budapeste, 2002. Página inicial. Disponível em: <https://www.budapestopenaccessinitiative.org/translations/portuguese-translation>. Acesso em: 04 fev. 2022.

BUDIN, Gerhard. A critical evaluation of the state-of-the-art of terminology theory. **Terminology Science and Research**, [s. l.], v. 12, n. 1-2, p. 7-231, 2001. Disponível em: <https://www.researchgate.net/publication/331247775>. Acesso em: 06 jan. 2022.

CABRÉ, Maria Teresa. **La Terminología**: representación y comunicación. Institut Universitari de Lingüística Aplicada Universitat Pompeu Fabra: Barcelona, 2005a.

CABRÉ, Maria Teresa. **La Terminología, una disciplina en evolución**: pasado, presente y algunos elementos de futuro. [S.l.; s.n.]: 2005b. Disponível em: [https://www.researchgate.net/publication/237821583\\_La\\_Terminologia\\_una\\_disciplina\\_en\\_evolucion\\_pasado\\_presente\\_y\\_algunos\\_elementos\\_de\\_futuro](https://www.researchgate.net/publication/237821583_La_Terminologia_una_disciplina_en_evolucion_pasado_presente_y_algunos_elementos_de_futuro). Acesso em: 06 fev. 2022.

CAFÉ, Ligia Maria Arruda.; SANTOS, Raimundo Nonato Macedo dos; BARROS, Camila Monteiro de. Os estudos de Gruber e Guarino sobre ontologias na ciência da informação e nas ciências da computação. **DataGramZero**, Porto Alegre, v. 16, n. 3, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/8169>. Acesso em: 01 mar. 2023.

CARVALHO, Flávia Medeiros de; FERREIRA, Alice Maria Araújo. Da sociolinguística à socioterminologia: definindo conceitos. **Tabuleiro de Letras**, Salvador, n. 05, dez. 2012. Disponível em: <https://doi.org/10.35499/tl.v0i5.176>. Acesso em: 06 fev. 2022.

CINTRA, Anna Maria Marques; TÁLAMO, Maria de Fátima Gonçalves Moreira; LARA, Marilda Lopes Ginez de; KOBASHI, Nair Yumiko. **Para entender as linguagens documentárias**. São Paulo: Polis, 1994.

COMITÉ INTERNACIONAL DE DOCUMENTAÇÃO (CIDOC). **Definition of the CIDOC Conceptual Reference Model**. Versão 7.1.1. [S.l.]: CIDOC-ICOM, 2021. Disponível em: <https://www.cidoc-crm.org/Version/version-7.1.1>. Acesso em: 09 jun. 2023.

CORONEL, Carlos; MORRIS, Steven; ROB, Peter. **Database Systems**: design, implementation and management. Boston, Massachusetts: Cengage Learning, 2011.

CORRÊA, Fabiano Couto. **Gestión de Datos de Investigación**. Barcelona: Editorial UOC, 2016.

COSTA, Sely M.S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ci. Inf.**, Brasília, v. 35, n. 2, p. 39-50, maio/ago. 2006. Disponível em: <https://doi.org/10.1590/S0100-19652006000200005>. Acesso em: 09 fev. 2022.

CURRÁS, Emilia. **Ontologies, Taxonomies and Thesauri in Systems Science and Systematics**. Oxford: Chandos Publishing, 2010.

DAHLBERG, Ingetraut. A referent-oriented, analytical Concept Theory for INTERCONCEPT. *International Classification*, [s.l.], v. 5, n. 3, p. 142-151, 1978b.

DAHLBERG, Ingetraut. Knowledge Organization: A New Science? **Knowledge Organization**, Baden-Baden, Alemanha, v. 33, n.1, 2006. Disponível em: [https://www.ergon-verlag.de/isko\\_ko/downloads/ko3320061c.pdf](https://www.ergon-verlag.de/isko_ko/downloads/ko3320061c.pdf). Acesso em: 10 out. 2022.

DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, 1978a. Disponível em: <https://revista.ibict.br/ciinf/article/view/115>. Acesso em: 10 out. 2022.

DAMASIO, Edilson. Preprints na comunicação científica: uma introdução. **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, Rio Grande, v. 32, n. 2, p. 155-168, 2018. Disponível em: <http://dx.doi.org/10.14295/biblos.v32i2.8635>. Acesso em: 10 out. 2022.

DIAS, Cláudia Augusto. Terminologia: conceitos e aplicações. **Ciência da Informação**, Brasília, v. 29, n. 1, 2000. Disponível em: <http://dx.doi.org/10.18225/ci.inf..v29i1.902>. Acesso em: 06 jan. 2022.

DING, Ying; FOO, Schubert. Ontology research and development. Part 1 - a review of ontology generation. **Journal of Information Science**, [s.l.], v. 28, n. 123, 2002. Disponível em: <https://doi.org/10.1177/016555150202800204>. Acesso em: 06 fev. 2022.

DOERR, Martin. "The new official edition of the CIDOC CRM, version 7.1.1, contains a compact section with title "Reality, Knowledge Bases and CIDOC CRM" [...]. *In: Issue 504*: Formulate the philosophical underpinnings of CRM and its relation to reality and the objectivity of observations. 2021. Fórum de discussão virtual. Disponível em: <https://cidoc-crm.org/Issue/ID-504-formulate-the-philosophical-underpinnings-of-crm-and-its-relation-to-reality-and-the>. Acesso em: 18 ago. 2022.

DOWNS, Robert R. Improving Opportunities for New Value of Open Data: Assessing and Certifying Research Data Repositories. **Data Science Journal**, [s.l.], v. 20, n. 1, 2021, p.1.

Disponível em: <http://doi.org/10.5334/dsj-2021-001>. Acesso em: 10 fev. 2022.

FURNIVAL, Ariadne Chloe; HUBBARD, Bill. Acesso Aberto às Publicações Científicas: vantagens, políticas e advocacy. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 2, n. 2, p. 160-177, 2011. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v2i2p160-177>. Acesso em: 9 fev. 2022.

GIL, Henrique Teixeira. A passagem da Web 1.0 para a Web 2.0 e... Web 3.0: Potenciais consequências para uma humanização» em contexto educativo. **Educativ** : boletim informativo, Castelo Branco, Portugal, n. 5, 2014, p. 1-2. Disponível em: <http://hdl.handle.net/10400.11/2404>. Acesso em: 08 fev. 2022.

GOMES, H. E. Marcos históricos e teóricos da organização do conhecimento. **Informação & Informação**, Londrina, v. 22, n. 2, p. 33-66, 2017. Disponível em: <https://doi.org/10.5433/1981-8920.2017v22n2p33>. Acesso em: 08 fev. 2022.

HEATH; Tom; BIZER, Christian. **Linked Data**: Evolving the Web into a Global Data Space. San Rafael, EUA: Morgan, 2011. Página inicial. Disponível em: <https://linkeddatabook.com/editions/1.0/>. Acesso em: 9 fev. 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 1087:2019**: Terminology work and terminology science — Vocabulary. [S.l.]: ISO, 2019. Disponível em: <https://www.iso.org/standard/62330.html>. Acesso em: 06 fev. 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO). **ISO 25964-2:2013**: Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies [S.l.]: ISO, 2013. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:25964:-2:ed-1:v1:en>. Acesso em: 06 out. 2022.

ISOTANI; Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados**: em Busca da Web do Conhecimento. São Paulo: Novatec Editora, 2015. Disponível em: <http://dx.doi.org/10.13140/RG.2.1.4355.6329>. Acesso em: 09 fev. 2022.

KREBS, L. M.; CAREGNATO, S. E. Implicações do acesso aberto aos dados de pesquisa: questões em debate. *In*: Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, n. XVIII. **Anais** [...]. Marília: ENANCIB, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/104452>. Acesso em: 10 fev. 2022.

KRIEGER, Maria da Graça. Terminologia e seus objetos de investigação. *In*: X Simposio Iberoamericano de Terminología, 2006, Montevideo. **Actas** [...]. Montevideo, 2008. p. 1-8.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à Terminologia**:

teoria e prática. São Paulo: Contexto, 2004. p. 24 a 65.

KURAMOTO, Hélio. Acesso Livre: uma solução adotada em todo o globo; porém, no Brasil parece existir uma indefinição. **RECIIS**, Rio de Janeiro, v. 8, n. 2, 2014. Disponível em: <http://eprints.rclis.org/23478/1/935-6274-1-PB.pdf>. Acesso em 04 fev. 2022.

LA REFERENCIA. **Políticas para la ciencia abierta y los datos científicos em américa latina**. [s.l.]: La referencia, 2018. 11 p. Disponível em: <http://www.lareferencia.info/pt/recursos/ciencia-abierta-documentos-externos/77-la-referencia-politicas-para-la-ciencia-abierta-y-los-datos-cientificos-en-america-latina-es>. Acesso em: 09 fev. 2022.

LAUFER, Carlos. **Guia de Web Semântica**. São Paulo: Centro de Estudos sobre Tecnologia Web, 2015. *E-book*. Disponível em: <http://ceweb.br/guias/web-semantica/>. Acesso em: 08 fev. 2022.

MACULAN, Benildes Coura Moreira dos Santos; AGANETTE, Elisângela Cristina. Desambiguação de relações em tesouros e o seu reúso em ontologias. **Ci.Inf.**, Brasília, v.46 n.1, p.102-119, jan./abr. 2017. Disponível em: <http://revista.ibict.br/ciinf/article/view/4017>. Acesso em: 06 fev. 2022.

MAYERNIK, Matthew. Metadata *In*: HJØRLAND, Birger; GNOLI, Claudio (ed.). **Encyclopedia of Knowledge Organization**. [S.l.]: International Society for Knowledge Organisation: 2020. Disponível em: <https://www.isko.org/cyclo/metadata>. Acesso em: 06 out. 2022.

MENDONÇA, Fabrício Martins. **Ontoforinfoscience**: metodologia para construção de ontologias pelos cientistas da informação. 2015. Tese (doutorado em Ciência da Informação) - Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUBD-A35H3K>. Acesso em: 10 fev. 2022.

MENDONÇA, Fabrício Martins; SOARES, Antônio Lucas. Construindo ontologias com a metodologia ontoforinfoscience: uma abordagem detalhada das atividades do desenvolvimento ontológico. **Ci.Inf.** Brasília, v. 46 n.1, p.43-59, jan./abr. 2017. Disponível em: <http://dx.doi.org/10.18225/ci.inf..v46i1.4013>. Acesso em: 06 fev. 2022.

MONTEIRO, Elizabete Cristina de Souza de Aguiar; SENA, Priscila Machado Borges; SANT'ANA, Ricardo César Gonçalves; BLATTMANN, Ursula. Repositório de dados como forma de organização do conhecimento e desenvolvimento científico. 14º Congresso ISKO España (4º ISKO España-Portugal). **Anais [...]**. Barcelona: ISKO, 2010. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=7393114>. Acesso em: 10 fev. 2022.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. **Ontologias**: conceitos,

usos, tipos, metodologias, ferramentas e linguagens. Relatório técnico. Goiânia: UFG (Instituto de Informática), 2007. Disponível em: [https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-07.pdf](https://ww2.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-07.pdf). Acesso em: 06 fev. 2022.

MOREIRA, Walter. **Sistemas de organização do conhecimento: aspectos teóricos, conceituais e metodológicos**. 2018. Tese (Livre-Docência em Sistemas de Organização do Conhecimento) — Departamento de Ciência da Informação – Faculdade de Filosofia e Ciências — Universidade Estadual Paulista, Marília, 2018. Disponível em: <https://repositorio.unesp.br/handle/11449/190878>. Acesso em: 08 fev. 2022.

MORESI, Eduardo (org). **Metodologia da pesquisa**. Brasília, DF: Universidade Católica de Brasília, 2003. Disponível em: <http://www.inf.ufes.br/~pdcosta/ensino/2010-2-metodologia-de-pesquisa/MetodologiaPesquisa-Moresi2003.pdf>. Acesso em: 08 fev. 2022.

MURRAY-RUST, Peter. Open Data in Science. **Nature Precedings**, [s.l.], 2008. Disponível em: <https://doi.org/10.1038/npre.2008.1526.1>. Acesso em: 04 fev. 2022.

NATIONAL INFORMATION STANDARDS SOCIETY (NISO). **Understanding Metadata**. Tucson, Arizona: NISO; University of Arizona, 2004. Disponível em: <https://repository.arizona.edu/handle/10150/105486>. Acesso em: 06 fev. 2022.

NATIONAL INFORMATION STANDARDS SOCIETY (NISO). **Understanding Metadata: What is Metadata, and What is it For?: A Primer**. Baltimore, Maryland: NISO, 2017. Disponível em: <https://www.niso.org/publications/understanding-metadata-2017>. Acesso em: 06 fev. 2022.

PAVÃO, Caterina Marta Groposo; PASSOS, Paula Caroline Schifino Jardim; GABRIEL JUNIOR, Rene Faustino; VANZ, Samile Andrea de Souza; BORGES, Eduardo Nunes; AZAMBUJA, Luís Alberto Barbosa. **Acesso aberto a dados de pesquisa no Brasil: repositórios brasileiros de dados de pesquisa: relatório 2018**. Porto Alegre: UFRGS, 2018. Disponível em: <https://lume.ufrgs.br/handle/10183/185138>. Acesso em: 10 fev. 2022.

PEIS, E; HERRERA-VIDEIRA; Enrique; HASSAN; Yang Rafidah; HERRERA; J.C. Análisis de la web semántica: estado actual y requisitos futuros. **El Profesional de la Información**, [s.l.], v.12, n.5, Set. 2003. Disponível em: <http://dx.doi.org/10.1076/epri.12.5.368.17567>. Acesso em: 09 fev. 2022.

RAFFERTY, Pauline. Tagging *In*: HJØRLAND, Birger; GNOLI, Claudio (ed.). **Encyclopedia of Knowledge Organization**. [S.l.]: International Society for Knowledge Organisation: 2020. Disponível em: <https://www.isko.org/cyclo/tagging>. Acesso em: 06 out. 2022.

ROCHE, Christopher. Ontological definition. *In*: KOCKAERT, H. J.; STEURS, F. (ed.).



**Handbook of Terminology.** Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015. p. 128-152.

RODRIGUES, Eloy; SARAIVA, Ricardo; RIBEIRO, Cristina; FERNANDES, Eugenia Matos. **Os repositórios de dados científicos: estado da arte.** Portugal: Universidade do Porto, 2010. Disponível em: <https://hdl.handle.net/10216/23806>. Acesso em: 09 fev. 2022.

SACHS, Eliza. **Noções básicas para uso do Protégé.** Tradução: Maurício Barcellos Almeida. Manual Eletrônico. 2006. Disponível em: [https://mba.eci.ufmg.br/onto\\_frames/](https://mba.eci.ufmg.br/onto_frames/). Acesso: 19 fev, 2023.

SALES, Rodrigo de; CAFÉ, Lígia. Tesouros e ontologias sob o olhar da Teoria Comunicativa da Terminologia. Congresso ISKO-Spain, IX, 2009, Valência, Espanha. **Anais [...].** Valência: ISKO, 2009. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/176537>. Acesso em: 06 fev. 2022.

SANTARÉM SEGUNDO, José Eduardo; CONEGLIAN, Caio Saraiva. Web Semântica e ontologias: um estudo sobre construção de axiomas e uso de inferências. **Informação & Informação**, Londrina, v. 21, n. 2, p. 217 – 244, maio/ago., 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p217>. Acesso em: 08 fev. 2022.

SANTOS, Monick Trajano dos. **Estudo do processo de apropriação da ontologia pela Ciência da Informação no Brasil.** 2014. Dissertação (mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Recife, 2014. Disponível em: <https://repositorio.ufpe.br/handle/123456789/12945>. Acesso em: 06 fev. 2022.

SANTOS, Monick Trajano dos; CORRÊA, Renato Fernandes; LAPA, Remi Correia. Estudos sobre a apropriação da Ontologia pela Ciência da Informação. Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), XIV, 2013, Florianópolis. **Anais [...].** Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/2333>. Acesso em: 06 jan. 2022.

SANTOS, Claudia; COSTA, Rute. Domain specificity: Semasiological and onomasiological knowledge representation. *In*: KOCKAERT, H. J.; STEURS, F. (ed.). **Handbook of Terminology.** Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015. p. 153-179.

SAYÃO, Luis Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli:** Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 15, n. 30, p. 1-31, 2010. Disponível em: <https://doi.org/10.5007/1518-2924.2010v15n30p1>. Acesso em: 12 out. 2022.

SAYÃO, Luis Fernando; SALES, Luana Farias. Algumas considerações sobre os

repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90 – 115, maio/ago., 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p90>. Acesso em: 08 fev. 2022.

SAYÃO, Luis Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS**, Rio de Janeiro, v. 8, n. 2, jun. 2014. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/611>. Acesso em: 02 fev. 2022.

SCHIESSL, Marcelo; BRÄSCHER, Marisa. Ontologia: ambiguidade e precisão. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 17, n. esp.1, p. 125-141, 2012. Disponível em: <https://doi.org/10.5007/1518-2924.2012v17nesp1p125>. Acesso em: 06 fev. 2022.

SCHIESSL, Marcelo; SHINTAKU, Milton. Sistemas de organização do conhecimento. *In*: ALVARES, Lillian (Org.). **Organização da informação e do conhecimento**: conceitos, subsídios interdisciplinares e aplicações. São Paulo: B4 Editores, 2012. p. 49-118.

SHAW, Ian; NICHOLSON, Paul. **Dictionary of Ancient Egypt**. Londres, Reino Unido: The British Museum Press, 1993.

SILVA, Marcel Ferrante; MARTINS, Dalton Lopes; SIQUEIRA, Joyce. Web semântica em repositórios: ontologia para representação de bibliotecas digitais. **Ciência da Informação em Revista**, Maceió, v. 6, n. 1, p. 99-113, 2019. Disponível em: <https://brapci.inf.br/index.php/res/v/113732>. Acesso em: 08 fev. 2022.

SILVA, Fabiano Couto Corrêa da; SILVEIRA, Lúcia da. O ecossistema da Ciência Aberta. **Transinformação**, São Paulo, v. 31, 2019. Disponível em: <http://dx.doi.org/10.1590/2318-0889201931e190001>. Acesso em: 04 fev. 2022.

SMITH, Barry; WELTY, Christopher. **Ontology**: Towards a New Synthesis. [S.l.: s.n.]. 2001. Disponível em: <https://philpapers.org/archive/SMIOTA-9.pdf>. Acesso em: 02 out. 2022.

SOUZA JÚNIOR, Mario Bastos de. **Análise de tipos de ontologias nas áreas de Ciência da Informação e Ciência da Computação**. 2014. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Santa Catarina, Florianópolis, 2014. Disponível em: <https://repositorio.ufsc.br/handle/123456789/123418>. Acesso em: 06 fev. 2022.

SOWA, John F. **Building, Sharing, and Merging Ontologies**. *Online*, 2001. Disponível em: <http://users.bestweb.net/~sowa/ontology/ontoshar.htm>. Acesso em: 06 fev. 2021.

USP (Universidade do Estado de São Paulo). **Resolução Nº 7900, de 11 de dezembro**

**de 2019.** Estabelece normas para a Gestão de Dados Científicos na Universidade de São Paulo. Normativa. Disponível em: <http://www.leginf.usp.br/?resolucao=resolucao-no-7900-de-11-de-dezembro-de-2019>. Acesso em: 10 fev. 2022.

VICKERY, Brian C. Ontologies. **Journal of Information Science**, [s.l.], v. 23, n. 277, 1997. Disponível em: <https://doi.org/10.1177/016555159702300402>. Acesso em: 06 fev. 2022.

WEISS, Leila Cristina. Interoperabilidade semântica: uma análise das perspectivas teóricas dos estudos desenvolvidos na área de ciência da informação. **Em Questão**, Porto Alegre, v. 27, n. 3, p. 431-457, 2021. Disponível em: <http://dx.doi.org/10.19132/1808-5245273.431-457>. Acesso em: 13 out. 2022.

ZENG, Marcia Lei. Introperability *In*: HJØRLAND, Birger; GNOLI, Claudio (ed.). **Encyclopedia of Knowledge Organization**. [S.l.]: International Society for Knowledge Organisation: 2020. Disponível em: <https://www.isko.org/cyclo/interoperability>. Acesso em: 06 out. 2022.

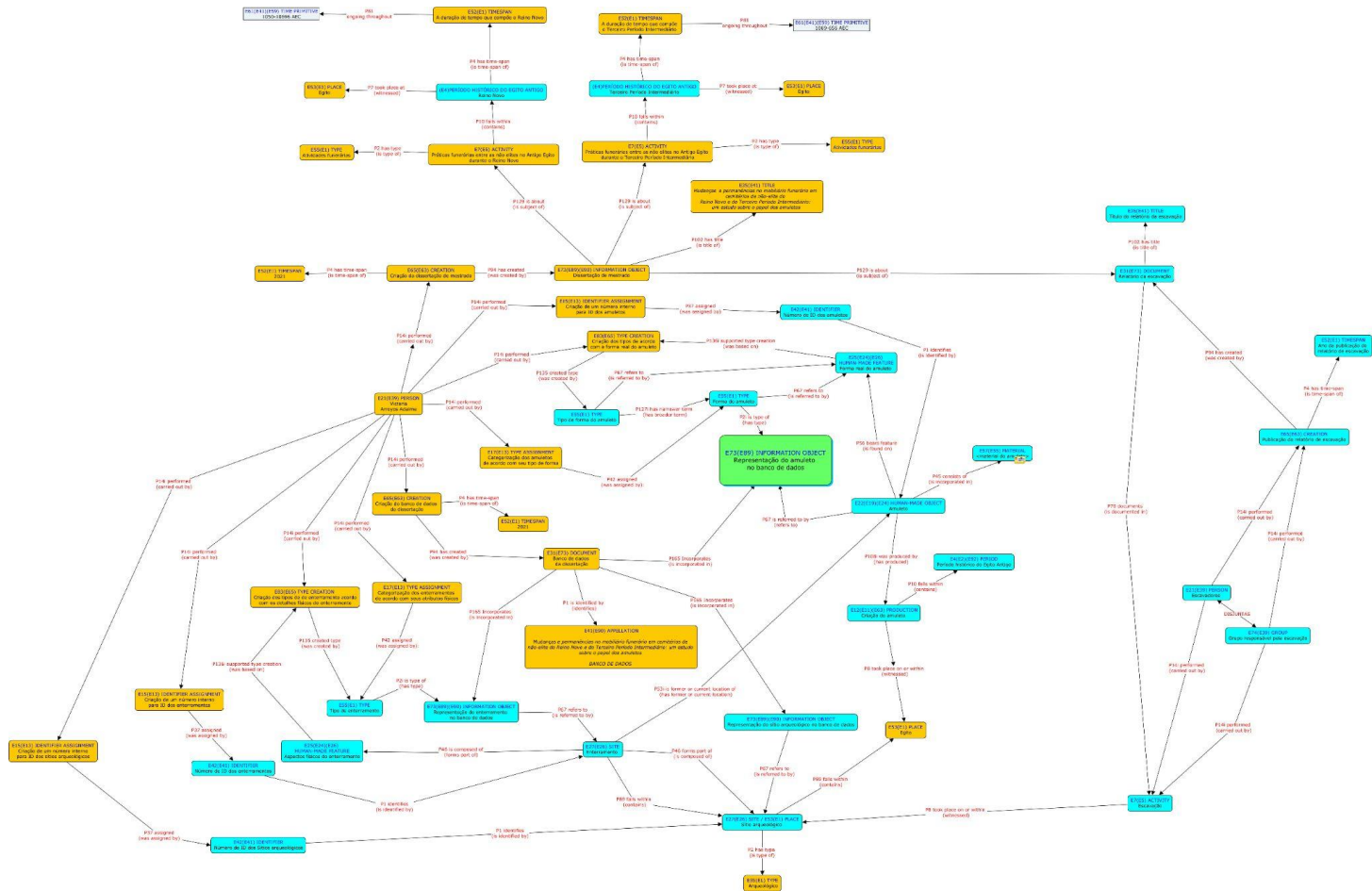
ZENG, Marcia Lei. Knowledge Organization Systems (KOS). **Knowledge Organization**, Baden-Baden, Alemanha, v. 35, n. 3, 2008. Disponível em: <https://www.doi.org/2010.5771/0943-7444-2008-2-3-160>. Acesso em: 06 out. 2022.

ZHAO, Bo. Web Scraping *In*: Schintler, L., McNeely, C. (eds) **Encyclopedia of Big Data**. [s.l.]: Springer, Cham. Disponível em: [https://doi.org/10.1007/978-3-319-32001-4\\_483-1](https://doi.org/10.1007/978-3-319-32001-4_483-1). Acesso em: 13 out. 2022.

ZIEGLER, Maria Fernanda. **FAPESP lança Rede de Repositórios de Dados Científicos do Estado de São Paulo**. *Online*, 2019. Disponível em: <https://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251/>. Acesso em: 10 fev. 2022.

## APÊNDICE A - MAPA CONCEITUAL DESENVOLVIDO DURANTE A PESQUISA

Abaixo observa-se uma versão menor do mapa conceitual desenvolvido para esta pesquisa. Essa versão apresenta-se em baixa qualidade devido às limitações do documento, de forma que é possível fazer seu *download* através deste link, no portal FigShare.



### APÊNDICE B - QUADRO DE CLASSES CIDOC-CRM

O quadro a seguir traz as classes nativas do CIDOC-CRM utilizadas em nosso modelo. Além de apresentar sua superclasse imediata, também são apresentadas definições para cada classe. Finalmente, apresenta-se as instâncias de cada classe, bem como uma definição para as instâncias que sejam indivíduos. As instâncias que são novas subclasses são apresentadas no Apêndice C.

CLASSE	SUPERCLASSE	DEFINIÇÃO RESUMIDA	INSTÂNCIAS	DEFINIÇÃO DAS INSTÂNCIAS
<b>E4(E2)(E92) PERIOD</b>	<b>E2 Temporal Entity, E92 Spacetime Volume</b>	This class comprises sets of coherent phenomena or cultural manifestations occurring in time and space.	<b><u>MP14</u></b> <b><u>Período</u></b> <b><u>Histórico do</u></b> <b><u>Egito Antigo</u></b>	Ver Apêndice C
<b>E7(E5) ACTIVITY</b>	<b>E5 Event</b>	This class comprises actions intentionally carried out by instances of E39 Actor that result in changes of state in the cultural, social, or physical systems documented	<u>Práticas</u> <u>funerárias</u> <u>entre as não</u> <u>elites no</u> <u>Antigo Egito</u> <u>durante o</u> <u>Reino Novo</u>	Conjunto de atividades relativas ao tratamento dos mortos no <u>Egito Antigo</u> , com base nas tradições religiosas locais, durante o Período do Reino Novo.

			<u>Práticas funerárias entre as não elites no Antigo Egito durante o Terceiro Período Intermediário</u>	Conjunto de atividades relativas ao tratamento dos mortos no <u>Egito</u> Antigo, com base nas tradições religiosas locais, durante o Terceiro Período Intermediário.
			<b>MP6 Escavação</b>	Ver Apêndice C
<b>E12(E11)(E63) PRODUCTION</b>	<b>E11 Modification, E63 Beginning of Existence</b>	This class comprises activities that are designed to, and succeed in, creating one or more new items.	<b>MP4 Criação do amuleto</b>	Ver Apêndice C

<b>E15(E13) IDENTIFIER ASSIGNMENT</b>	<b>E13 Attribute Assignment</b>	This class comprises activities that result in the allocation of an identifier to an instance of E1 CRM Entity. Instances of E15 Identifier Assignment may include the creation of the identifier from multiple constituents, which themselves may be instances of E41 Appellation.	<u>Criação de um número interno para ID dos enterramentos</u>	Atividade de criação de um <b>MP12 Número de ID de enterramento</b> , realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> .
			<u>Criação de um número interno para ID dos amuletos</u>	Atividade de criação de um <b>MP11 Número de ID de amuleto</b> , realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> .
			<u>Criação de um número interno para ID dos sítios arqueológicos</u>	Atividade de criação de um <b>MP13 Número de ID de sítio arqueológico</b> , realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> .
<b>E17(E13) TYPE ASSIGNMENT</b>	<b>E13 Attribute Assignment</b>	This class comprises the actions of classifying items of whatever kind. Such items include objects, specimens, people, actions and concepts.	<u>Categorização dos amuletos de acordo com seu tipo de forma</u>	Atividade de categorização das instâncias de <b>MP17 Representação do amuleto no banco de dados</b> . Realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> , com base na <b>MP9 Forma real do amuleto</b> encontrada nas instâncias de <b>MP1 Amuleto</b> , a partir do conteúdo de uma instância de <b>MP16 Relatório de escavação</b> .
			<u>Categorização dos enterramentos de acordo com seus atributos físicos</u>	Atividade de categorização das instâncias de " <b>MP18 Representação do enterramento no banco de dados</b> ". Realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> , com base nos <b>MP3 Aspectos físicos do enterramento</b> encontrada nas instâncias de <b>MP5 Enterramento</b> , a partir do conteúdo de uma instância de <b>MP16 Relatório de escavação</b> .

<b>E21(E39) PERSON</b>	<b>E39 Actor</b>	This class comprises real persons who live or are assumed to have lived.	<u>Victoria Arroyos Adaime</u>	sua <u>dissertação de mestrado</u> , intitulada " <i>Mudanças e permanências no mobiliário funerário</i> " e pela maior parte dos eventos representados no <u>banco de dados da dissertação</u>
			<b>MP7 Escavador</b>	Ver Apêndice C
<b>E22(E19)(E24) HUMAN-MADE OBJECT</b>	<b>E19 Physical Object, E24 Physical Human-Made Thing</b>	This class comprises all persistent physical objects of any size that are purposely created by human activity and have physical boundaries that separate them completely in an objective way from other objects.	<b>MP1 Amuleto</b>	Ver Apêndice C
<b>E25(E24)(E26) HUMAN-MADE FEATURE</b>	<b>E24 Physical Human-Made Thing, E26 Physical Feature</b>	This class comprises physical features that are purposely created by human activity, such as scratches, artificial caves, artificial water channels, etc	<b>MP9 Forma real do amuleto</b>	Ver Apêndice C
			<b>MP3 Aspectos físicos do enterrament</b>	Ver Apêndice C
<b>E27(E26) SITE</b>	<b>E26 Physical Feature</b>	This class comprises pieces of land or sea floor.	<b>MP20 Sítio Arqueológico</b>	Ver Apêndice C
			<b>MP5 Enterrament</b>	Ver Apêndice C



E31(E73) DOCUMENT	E73 Information Object	This class comprises identifiable immaterial items that make propositions about reality. These propositions may be expressed in text, graphics, images, audiograms, videograms or by other similar means. Documentation databases are regarded as instances of E31 Document.	<p><b>MP16</b> <b>Relatório de escavação</b></p> <p><u>Banco de dados da dissertação</u></p>	<p>Ver Apêndice C</p> <p>Banco de dados contendo as instâncias de <b>MP17 Representação do amuleto no banco de dados, MP18 Representação do enterramento no banco de dados e MP19 Representação do sítio arqueológico no banco de dados.</b> Elaborado pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u>, possui como título genérico <u>"Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos BANCO DE DADOS"</u></p>
E35(E41) TITLE	E41 Appellation	This class comprises textual strings that within a cultural context can be clearly identified as titles due to their form	<p><u>Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos</u></p>	<p>Título da <u>dissertação de mestrado</u> da pesquisadora brasileira <u>Victoria Arroyo Adaime</u>.</p>

			<b><u>MP23 Título do relatório da escavação</u></b>	Ver Apêndice C
<b>E41(E90) APPELLATION</b>	<b>E90 Symbolic Object</b>	This class comprises signs, either meaningful or not, or arrangements of signs following a specific syntax, that are used or can be used to refer to and identify a specific instance of some class or category within a certain context.	<i><u>Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos</u></i>  <b><u>BANCO DE DADOS</u></b>	Título genérico atribuído ao banco de dados da <u>dissertação de mestrado</u> da pesquisadora brasileira <u>Victoria Arroyo Adaime</u> .
<b>E42(E41) IDENTIFIER</b>	<b>E41 Appellation</b>	This class comprises strings or codes assigned to instances of E1 CRM Entity in order to identify them uniquely and permanently within the context of one or more organisations	<b><u>MP11 Número de ID de amuletos</u></b>	Ver Apêndice C
			<b><u>MP12 Número de ID de enterramentos</u></b>	Ver Apêndice C

			<b><u>MP13</u></b> <b><u>Número de</u></b> <b><u>ID de sítio</u></b> <b><u>arqueológico</u></b> <b><u>o</u></b>	Ver Apêndice C
<b>E52(E1)</b> <b>TIMESPAN</b>	<b>E1 CRM Entity</b>	This class comprises abstract temporal extents, in the sense of Galilean physics, having a beginning, an end and a duration. Instances of E52 Time-Span have no semantic connotations about phenomena happening within the temporal extent they represent. They do not convey any meaning other than a positioning on the “time-line” of chronology	<b><u>A duração de</u></b> <b><u>tempo que</u></b> <b><u>compõe o</u></b> <b><u>Reino Novo</u></b>	Extensão de tempo convencionalizada pela historiografia do <u>Egito</u> , tratando do <b>MP14</b> <b><u>Período Histórico do Egito Antigo</u></b> conhecido como Reino Novo
			<b><u>A duração de</u></b> <b><u>tempo que</u></b> <b><u>compõe</u></b> <b><u>o Terceiro</u></b> <b><u>Período</u></b> <b><u>Intermediário</u></b>	Extensão de tempo convencionalizada pela historiografia do <u>Egito</u> , tratando do <b>MP14</b> <b><u>Período Histórico do Egito Antigo</u></b> conhecido como <u>Terceiro Período Intermediário</u> .
			<b><u>2021</u></b>	Representação do ano 2021 da Era Comum.
			<b><u>MP2 Ano da</u></b> <b><u>criação do</u></b> <b><u>relatório de</u></b> <b><u>escavação</u></b>	Ver Apêndice C

<b>E53(E1) PLACE</b>	<b>E1 CRM Entity</b>	This class comprises extents in space, in particular on the surface of the earth, in the pure sense of physics: independent from temporal phenomena and matter.	<b><u>MP20 Sítio Arqueológico</u></b>	
			<b><u>Egito</u></b>	País do Norte da África conhecido por sua história antiga. É lar dos <b><u>MP20 Sítio Arqueológico</u></b> e de diversos entes relacionados a eles apresentados neste modelo. Também é o local onde ocorrem diversos eventos relacionados a esses entes, como os eventos de <b><u>MP4 Criação do amuleto</u></b> e <b><u>MP7 Escavação</u></b> .
<b>E55(E28) TYPE</b>	<b>E28 Conceptual Object</b>	This class comprises concepts denoted by terms from thesauri and controlled vocabularies used to characterize and classify instances of CIDOC CRM classes. Instances of E55 Type represent concepts in contrast to instances of E41 Appellation which are used to name instances of CIDOC CRM classes. independent from temporal phenomena and matter.	<b><u>Atividades funerárias</u></b>	Designação geral para as práticas religiosas e culturais envolvendo o tratamento dos mortos no Antigo <b><u>Egito</u></b> . É o tema de pesquisa da pesquisadora <b><u>Victoria Arroyo Adaime</u></b> em sua <b><u>dissertação de mestrado</u></b> .
			<b><u>Arqueológico</u></b>	Relativo à disciplina da Arqueologia; nesse contexto, serve como descritor à classe <b><u>MP20 Sítio Arqueológico</u></b> e suas instâncias.
			<b><u>MP22 Tipo de forma do amuleto</u></b>	Ver Apêndice C
			<b><u>MP8 Forma do amuleto</u></b>	Ver Apêndice C
			<b><u>MP21 Tipo de enterramento</u></b>	Ver Apêndice C

<b>E57(E55) MATERIAL</b>	<b>E55 Type</b>	<p>This class is a specialization of E55 Type and comprises the concepts of materials.</p> <p>Instances of E57 Material may denote properties of matter before its use, during its use, and as incorporated in an object [...]</p> <p>This type is used categorically in the model without reference to instances of it, i.e., the Model does not foresee the description of instances of instances of E57 Material [...]</p>	<u>Vidro</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas. Feito de areia cristalizada, era visto como uma pedra preciosa artificial no Antigo <u>Egito</u> .
			<u>Cerâmica:</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas. É elemento particularmente comum na <u>Arqueologia do Antigo Egito</u> .
			<u>Faiança egípcia</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas. É uma espécie de cerâmica, feita de fragmentos de quartos misturados a cinzas e cal. comum na <u>arqueologia do Antigo Egito</u> .
			<u>Jaspe</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
			<u>Cornalina</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
			<u>Esteatita</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
			<u>Ouro</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas. Metal raro, ainda assim surge como material de algumas instâncias de <b>MP1 Amuleto</b> que datem do período do <u>Novo Reino</u> .
			<u>Lápis-lázuli (imitação)</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas. Surge como imitação mais acessível às não-elites do lápis-lazúli verdadeiro, pedra considerada elegante no Antigo <u>Egito</u> .
			<u>Concha</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
			<u>Osso</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
			<u>Calcário</u>	Material do qual algumas instâncias de <b>MP1 Amuleto</b> são feitas.
<u>Sem informação de material</u>	Ausência de informações acerca do tipo de material que compunha uma instância de <b>MP1 Amuleto</b> .			

<b>E61(E41)(E59)</b> <b>TIME PRIMITIVE</b>	<b>E41 Appellation;</b> <b>E59 Primitive Value</b>	This class comprises instances of E59 Primitive Value for time that should be implemented with appropriate validation, precision and references to temporal coordinate systems to express time in some context relevant to cultural and scientific documentation	<u>1550-1069</u> <u>AEC</u>	Representação dos período de tempo no qual se desenrolaram as <u>Práticas funerárias entre as não elites no Antigo Egito durante o Reino Novo.</u>
			<u>1069-656</u> <u>AEC</u>	Representação dos período de tempo no qual se desenrolaram as <u>Práticas funerárias entre as não elites no Antigo Egito durante o Terceiro Período Intermediário" durante o Reino Novo</u>
<b>E65(E63)</b> <b>CREATION</b>	<b>E63 Beginning of Existence</b>	This class comprises events that result in the creation of conceptual items or immaterial products.	<u>Criação da dissertação de mestrado</u>	Representação concisa do conjunto de atividades no qual a pesquisadora brasileira <u>Victoria Arroyo Adaime</u> elaborou sua <u>dissertação de mestrado</u> , no ano de <u>2021</u> .
			<b>MP15</b> <b>Publicação do relatório de escavação</b>	Ver Apêndice C
			<u>Criação do banco de dados da dissertação</u>	Representação concisa do conjunto de atividades no qual a pesquisadora brasileira <u>Victoria Arroyo Adaime</u> criou o banco de dados de sua <u>dissertação de mestrado</u> , no ano de <u>2021</u> .

<b>E73(E89)(E90) INFORMATION OBJECT</b>	<b>E89 Propositional Object; E90 Symbolic Object</b>	This class comprises identifiable immaterial items, such as poems, jokes, data sets, images, texts, multimedia objects, procedural prescriptions, computer program code, algorithm or mathematical formulae, that have an objectively recognizable structure and are documented as single units.	<u>Dissertação de mestrado</u>	Esforço acadêmico pelo qual a pesquisadora brasileira <u>Victoria Arroyo Adaime</u> alcançou seu título de Mestra. Tem por título " <u>Mudanças e permanências no mobiliário funerário em cemitérios de não-elite do Reino Novo e do Terceiro Período Intermediário: um estudo sobre o papel dos amuletos</u> ". Tem por assunto as <u>Práticas funerárias entre as não elites no Antigo Egito durante o Reino Novo e as Práticas funerárias entre as não elites no Antigo Egito durante o Terceiro Período Intermediário</u> . Tem por apêndice o banco de dados da dissertação.
			<b>MP17</b> <b>Representação do amuleto no banco de dados</b>	Ver Apêndice C
			<b>MP18</b> <b>Representação do enterramento no banco de dados</b>	Ver Apêndice C
			<b>MP19</b> <b>Representação do sítio arqueológico no banco de dados</b>	Ver Apêndice C

<b>E74(E39) GROUP</b>	<b>E39 Actor</b>	This class comprises any gatherings or organizations of human individuals or groups that act collectively or in a similar way due to any form of unifying relationship. In the wider sense this class also comprises official positions which used to be regarded in certain contexts as one actor, independent of the current holder of the office, such as the president of a country.	<b><u>MP10 Grupo responsável pela escavação</u></b>	Ver Apêndice C
<b>E83(E65) TYPE CREATION</b>	<b>E65 CREATION</b>	This class comprises activities formally defining new types of items.	<u>Criação dos tipos de amuleto com base na forma real do amuleto</u>	Atividade de criação de categorias para <b><u>MP17 Representação do amuleto no banco de dados</u></b> realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> . É baseado nas instâncias de <b><u>MP9 Forma real do amuleto</u></b> , e teve como resultado as instâncias de <b><u>MP22 Tipo de forma do amuleto</u></b> e suas especificações em <b><u>MP8 Forma do amuleto</u></b> .



			<u>Criação dos tipos de enteramento com base nos detalhes físicos do enteramento</u>	Atividade de criação de categorias para <b>MP18 Representação do enteramento no banco de dados</b> realizada pela pesquisadora <u>Victoria Arroyo Adaime</u> durante a elaboração de sua <u>dissertação de mestrado</u> . É baseado nas instâncias de <b>MP3 Aspectos físicos do enteramento</b> , e teve como resultado as instâncias de <b>MP21 Tipo de enteramento</b> .
--	--	--	--	---

### APÊNDICE C - QUADRO DE CLASSES NATIVAS DO MODELO PROPOSTO

O quadro a seguir traz as classes desenvolvidas nesta pesquisa, que são instâncias de classes nativas do CIDOC-CRM. Além de apresentar sua superclasse imediata, também são apresentadas definições para cada classe. Finalmente, apresenta-se as instâncias de cada classe. Consideramos que as definições das classes contêm informação suficiente para que seja possível compreender do que se trata cada indivíduo pertencente àquela classe, razão pela qual não apresentamos definição para cada instância individualmente.

CLASSE	SUPERCLASSE	DEFINIÇÃO RESUMIDA	<u>Instâncias</u>
<b><u>MP1 Amuleto</u></b>	<b>E22(E19)(E24) HUMAN-MADE OBJECT</b>	Amuleto funerário egípcio, presente em dos <b>MP5 Enterramento</b> localizados em um dos <b>MP20 Sítio Arqueológico</b> estudados pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u> . É composto por algum <b>E57(E55) MATERIAL</b> e possui uma <b>MP9 Forma real do amuleto</b> .  Possui uma <b>MP17 Representação do amuleto</b> no <u>banco de dados da dissertação</u> . e um <b>MP11 Número de ID de amuleto</b> .	Amuleto 115 Amuleto 155 Amuleto 187 Amuleto 205 Amuleto 232 Amuleto 318 Amuleto 348 Amuleto 364 Amuleto 738 Amuleto 694
<b><u>MP2 Ano da publicação do relatório de escavação</u></b>	<b>E52(E1) TIMESPAN</b>	Ano em que foi publicado qualquer um dos <b>MP16 Relatório de escavação</b> que serviram como fonte bibliográfica para a pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u> .	1890 1891 1892 1893

			1905
			1923
			1927
			1948
<b><u>MP3 Aspectos físicos do enterramento</u></b>	<b>E25(E24)(E26) HUMAN-MADE FEATURE</b>	Aspectos físicos presentes em um <b><u>MP5 Enterramento</u></b> estudado pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u> . Sua existência serviu de base para a <u>Categorização dos enterramentos de acordo com seus atributos físicos</u> .	Presença de Fosso
			Presença de Nicho lateral duplo
			Presença de Túmulo de poço
			Presença de Mastaba
			Presença de Fosso com cavidade (reutilizada)
			Aspectos físicos desconhecidos
			Presença de Pirâmide (reutilizada)
			Presença de Fosso profundo com câmaras
			Presença de Nicho
			Presença de Câmara com acessos
			Presença de Túmulo de poço (reutilizado)

<p><b><u>MP4 Criação do amuleto</u></b></p>	<p><b>E12(E11)(E63) PRODUCTION</b></p>	<p>Evento de criação de um <b><u>MP1 Amuleto</u></b> presente em um <b><u>MP5 Enterramento</u></b> localizado em qualquer um dos <b><u>MP20 Sítio arqueológico</u></b> estudados pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u>.</p>	<p>Criação do Amuleto 115, Criação do Amuleto 155, Criação do Amuleto 187, Criação do Amuleto 205, Criação do Amuleto 232, Criação do Amuleto 318, Criação do Amuleto 348, Criação do Amuleto 364, Criação do Amuleto 738, Criação do Amuleto 694</p>
<p><b><u>MP5 Enterramento</u></b></p>	<p><b>E27(E26) SITE</b></p>	<p>Local contido em um <b><u>MP20 Sítio Arqueológico</u></b> onde se encontram contidos os restos mortais e a mobiliária funerária de indivíduos de não-elite que morreram em dado <b><u>MP13 Período Histórico do Egito Antigo</u></b>. Dentre essa mobiliária, encontram-se exemplos de <b><u>MP1 Amuleto</u></b>. Estudado pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u>. Possui uma <b><u>MP17 Representação do enterramento</u></b> no <u>banco de dados da dissertação</u> e um <b><u>MP12 Número de ID de enterramento</u></b>.</p>	<p>Enterramento 38, Enterramento 48, Enterramento 61, Enterramento 69, Enterramento 79, Enterramento 115, Enterramento 125, Enterramento 129, Enterramento 200, Enterramento 205.</p>

<b><u>MP6 Escavação</u></b>	<b>E7(E5) ACTIVITY</b>	<p>Atividade de demarcação de um <b><u>MP20 Sítio Arqueológico</u></b>, seguida pelo desenterramento, remoção e catalogação de artefatos arqueológicos, como os amuletos funerários em <b><u>MP1 Amuleto</u></b>.</p> <p>É conduzida por um <b><u>MP7 Escavador</u></b> ou um <b><u>MP10 Grupo responsável pela escavação</u></b>, e seus resultados são apresentados em um <b><u>MP16 Relatório de escavação</u></b>.</p>	Escavação Petrie (Gurob)
			Escavação Loat
			Escavação Brunton-Engelbach;
			Escavação Gurob Harem Palace Project
			Escavação Brunton
			Escavação Petrie (Matmar)
			Escavação Petrie/Brunton/Murray
<b><u>MP7 Escavador</u></b>	<b>E21(E39) PERSON</b>	<p>Indivíduo responsável por realizar ou coordenar uma <b><u>MP6 Escavação</u></b> e executar a <u>Publicação do relatório de escavação</u>, dando origem a um <b><u>MP16 Relatório de escavação</u></b>.</p>	Flinders Petrie
			Leonard Loat
			Guy Brunton
			Margaret Alice Murray
			Richard Engelbach;
<b><u>MP8 Forma do amuleto</u></b>	<b>E55(E1) TYPE</b>	<p>Representação mais precisa da <b><u>MP9 Forma real</u></b> do amuleto de um <b><u>MP1 Amuleto</u></b> no <u>banco de dados da dissertação</u> elaborado por Victoria Arroyo Adaime.</p> <p>Possui uma relação de especificação em relação a <b><u>MP22 Tipo de forma do amuleto</u></b>.</p>	Hathor, Thoth, Ísis e Hórus, Nefertum, Sekhmet, Taweret, Bes, Hórus, Ísis, Nehebkau, Hórus criança, Pataikos, Néftis, Mut
			Nefer, udjat, pilar djed, lua crescente, menat, égide, tyt, uraeus, pássaro-ba, cetro-was;

			<p>Sapo, crocodilo, tartaruga, lebre, porca, falcão, pato, mosca, babuíno, peixe, gato, chacal, íbis, serpente, carneiro, leão, golfinho;</p> <p>Coluna de papiro, flor de lótus;</p> <p>Perna, cabeça, rosto, coração, mão</p> <p>Mesa de oferenda, sítula, lira, Forma não-identificada</p>
<b><u>MP9 Forma real do amuleto</u></b>	<b>E25(E24)(E26) HUMAN-MADE FEATURE</b>	<p>Forma real de um <b><u>MP1 Amuleto</u></b>, contendo inscrições, imagens e detalhes físicos particulares.</p> <p>Serviu como base para a <u>Categorização dos amuletos de acordo com seu tipo de forma.</u></p>	<p>Forma da deusa Hathor, forma do deus Thoth, etc.</p> <p>Forma de Nefer, forma de udjat, etc.</p> <p>Forma de Sapo, forma de crocodilo, etc</p> <p>Forma de Coluna de papiro, forma de flor de lótus</p> <p>Forma de Perna, forma de cabeça</p> <p>Forma de Mesa de oferenda, Forma não-identificada, etc.</p>
<b><u>MP10 Grupo responsável pela escavação</u></b>	<b>E74(E39) GROUP</b>	<p>Grupo de indivíduos responsável por realizar ou coordenar uma <b><u>MP6 Escavação</u></b> e realizar a <u>Publicação do relatório de escavação</u>, dando origem a um <b><u>MP16 Relatório de escavação.</u></b></p>	<p>Gurob Harem Palace Project</p>

<p><b><u>MP11 Número de ID de amuletos</u></b></p>	<p><b>E42(E41) IDENTIFIER</b></p>	<p>Número de identificação de um <b><u>MP1 Amuleto.</u></b> Atribuído por <u>Victoria Arroyo Adaime</u> num esforço de <u>Criação de um número interno</u> <u>para ID dos amuletos</u> com base em um <b><u>MP16 Relatório de escavação.</u></b></p>	<p>115 155 187 205 232 318 348 364 738 694</p>
<p><b><u>MP12 Número de ID de enterramentos</u></b></p>	<p><b>E42(E41) IDENTIFIER</b></p>	<p>Número de identificação de um <b><u>MP5 Enterramento.</u></b> Atribuído por <u>Victoria Arroyo Adaime</u> num esforço de <u>Criação de um número interno</u> <u>para ID dos enterramentos</u> com base em um <b><u>MP16 Relatório de escavação.</u></b></p>	<p>38 48 61 69 79 115e 125 129 200 205e</p>

<b><u>MP13 Número de ID de sítio arqueológico</u></b>	<b>E42(E41) IDENTIFIER</b>	Número de identificação de um <b><u>MP20 Sítio Arqueológico</u></b> . Atribuído por <u>Victoria Arroyo Adaime</u> num esforço de <u>Criação de um número interno para ID dos sítios arqueológicos com base em um MP16 Relatório de escavação.</u>	1 2 3
<b><u>MP14 Período Histórico do Egito Antigo</u></b>	<b>E4(E2)(E92) PERIOD</b>	Período histórico utilizado na Historiografia do <u>Egito</u> , durante o qual ocorreram muitas das atividades descritas em um <b><u>MP16 Relatório de escavação.</u></b>	Reino Novo Terceiro Período Intermediário
<b><u>MP15 Publicação do relatório de escavação</u></b>	<b>E65(E63) CREATION</b>	Evento de publicação de um <b><u>MP16 Relatório de escavação</u></b> pelo <b><u>MP7 Escavador</u></b> ou <b><u>MP10 Grupo de escavadores</u></b> responsável. Ocorreu em um <b><u>MP2 Ano da publicação do relatório de escavação.</u></b>	Publicação do relatório <i>Kahun, Gurob, and Hawar</i> Publicação do relatório <i>Illahun, Kahun and Gurob</i> Publicação do relatório <i>Gurob, BSAE/ERA X</i> Publicação do relatório <i>Gurob.</i> Publicação do relatório <i>The Gurob Harem Palace Project</i> Publicação do relatório <i>Matmar</i> Publicação do relatório <i>Lahun II, BSAE/ERA XXXII</i>



			Publicação do relatório <i>Ten years' digging in Egypt, 1881-1891</i>
<b><u>MP16 Relatório de escavação</u></b>	<b>E31(E73) DOCUMENT</b>	Documento de carácter científico que regista uma <b><u>MP6 Escavação</u></b> e suas descobertas, publicado num <b><u>MP2 Ano da publicação do relatório de escavação</u></b> por um <b><u>MP7 Escavador</u></b> ou um <b><u>MP10 Grupo de Escavadores.</u></b>	Relatório da escavação Petrie em Kahun, Gurob, e Hawar
			Relatório da escavação Petrie em Lahun e Gurob
			Relatório da escavação Loat em Gurob
			Relatório da escavação Brunton/Engelbach em Gurob.
			Relatório da escavação realizada The Gurob Harem Palace Project em Gurob
			Relatório da escavação Brunton em Matmar
			Relatório da escavação Petrie/Brunton/Murray em Lahun

<p><b><u>MP17</u></b>  <b><u>Representação</u></b>  <b><u>do amuleto</u></b>  <b><u>no banco de</u></b>  <b><u>dados</u></b></p>	<p><b>E73(E89)(E90)</b>  <b>INFORMATION</b>  <b>OBJECT</b></p>	<p>Representação de um <b><u>MP1 Amuleto</u></b> no <u>banco de dados da dissertação</u>, construída por <u>Victoria Arroyo Adaime</u>. Possui um <b><u>MP11 Número de ID de amuleto</u></b> e possui como tipo alguma categoria de <b><u>MP8 Forma do amuleto</u></b>.</p>	<p>Representação do Amuleto 115  Representação do Amuleto 155  Representação do Amuleto 187  Representação do Amuleto 205  Representação do Amuleto 232  Representação do Amuleto 318  Representação do Amuleto 348  Representação do Amuleto 364  Representação do Amuleto 738  Representação do Amuleto 694</p>
<p><b><u>MP18</u></b>  <b><u>Representação</u></b>  <b><u>do enterramento</u></b>  <b><u>no banco de</u></b>  <b><u>dados</u></b></p>	<p><b>E73(E89)(E90)</b>  <b>INFORMATION</b>  <b>OBJECT</b></p>	<p>Representação de um <b><u>MP5 Enterramento</u></b> no <u>banco de dados da dissertação</u>, construída por <u>Victoria Arroyo Adaime</u>. Possui um <b><u>MP12 Número de ID de enterramento</u></b> e possui como tipo alguma categoria de <b><u>MP21 Tipo de enterramento</u></b>.</p>	<p>Representação do Enterramento 38  Representação do Enterramento 48  Representação do Enterramento 61  Representação do Enterramento 69  Representação do Enterramento 79  Representação do Enterramento 115  Representação do Enterramento 125  Representação do Enterramento 129  Representação do Enterramento 200  Representação do Enterramento 205</p>

<p><b><u>MP19</u></b> <b><u>Representação</u></b> <b><u>do sítio</u></b> <b><u>arqueológico no</u></b> <b><u>banco de dados</u></b></p>	<p><b>E73(E89)(E90)</b> <b>INFORMATION</b> <b>OBJECT</b></p>	<p>Representação de um <b><u>MP20 Sítio Arqueológico</u></b> no <u>banco de dados da dissertação</u>, construída por <u>Victoria Arroyo Adaime</u>. Possui um <b><u>MP13 Número de ID de sítio arqueológico</u></b>.</p>	<p>Representação do sítio arqueológico 1, Representação do sítio arqueológico 2, Representação do sítio arqueológico 3</p>
<p><b><u>MP20 Sítio</u></b> <b><u>Arqueológico</u></b></p>	<p><b>E27(E26) SITE</b></p>	<p>Local de importância arqueológica, demarcado deliberadamente, onde ocorreu uma <b><u>MP6 Escavação</u></b>, contendo diversos <b><u>MP5 Enterramento</u></b>. Localizado no <u>Egito</u>. Estudado pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u>. Possui uma <b><u>MP18 Representação do sítio arqueológico</u></b> no <u>banco de dados da dissertação</u> e um <b><u>MP13 Número de ID de sítio arqueológico</u></b>.</p>	<p><u>Gurob</u></p> <p><u>Lahun</u></p> <p><u>Matmar</u></p>
<p><b><u>MP21 Tipo de</u></b> <b><u>enterramento</u></b></p>	<p><b>E55(E1) TYPE</b></p>	<p>Categorias utilizadas para organizar as instâncias de <b><u>MP18 Representação do enterramento no banco de dados</u></b> dentro do <u>banco de dados da dissertação</u> em um evento de <u>Categorização dos enterramentos de acordo com seus atributos físicos</u>. São fruto de um evento de <u>Criação dos tipos de enterramento com base nos detalhes físicos do enterramento</u> executado pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u>, a partir dos <b><u>MP3 Aspectos físicos do enterramento</u></b>.</p>	<p><u>Nicho lateral duplo</u></p> <p><u>Túmulo de poço</u></p> <p><u>Fosso com cavidade (reutilizada)</u></p> <p><u>Desconhecido</u></p> <p><u>Pirâmide (reutilizada)</u></p> <p><u>Fosso profundo com câmaras</u></p> <p><u>Nicho</u></p> <p><u>Câmara com acessos</u></p>

			<u>Túmulo de poço (reutilizado)</u>
			<u>Mastaba</u>
<b>MP22 Tipo de forma do amuleto</b>	<b>E55(E1) TYPE</b>	Categorias utilizadas para organizar as instâncias de <b>MP17 Representação do amuleto no banco de dados</b> dentro do <u>banco de dados da dissertação</u> em um evento de <u>Categorização dos amuletos de acordo com seu tipo de forma</u> . São fruto de um evento de <u>Criação dos tipos de acordo com a forma real do amuleto</u> executado pela pesquisadora <u>Victoria Arroyo Adaime</u> em sua <u>dissertação de mestrado</u> , a partir da <b>MP9 Forma real do amuleto</b> .	<u>Divindade</u>
			<u>Símbolos divinos</u>
			<u>Animais</u>
			<u>Formas naturais</u>
			<u>Partes do corpo</u>
			<u>Outras formas</u>
<b>MP23 Título do relatório da escavação</b>	<b>E35(E41) TITLE</b>	Título atribuído a uma instância de <b>MP16 Relatório de escavação</b> .	<u><i>Kahun, Gurob, and Hawar</i></u>
			<u><i>Illahun, Kahun and Gurob</i></u>
			<u><i>Gurob, BSAE/ERA X</i></u>
			<u><i>Gurob.</i></u>
			<u><i>The Gurob Harem Palace Project</i></u>
			<u><i>Matmar</i></u>
			<u><i>Lahun II, BSAE/ERA XXXII</i></u>
			<u><i>Ten years' digging in Egypt. 1881-1891</i></u>

### APÊNDICE D - QUADRO DE RELACIONAMENTOS

O quadro a seguir apresenta as propriedades do CIDOC-CRM que permitiram o estabelecimento das relações entre as classes e instâncias em nosso modelo. O quadro apresenta, além das propriedades, o *domain* e o *range* possíveis para cada uma, além de sua definição, sua quantificação, seus atributos e se seu uso foi na configuração direta ou inversa.

Propriedade	Domain	Range	Definição	Quantificação direta	Atributos	Direta ou inversa
P1 is identified by (identifies)	E1 CRM Entity	E41 Appellation	<i>This property describes the naming or identification of any real-world item by a name or any other identifier [...]</i>	many to many (0,n:0,n)		
P2 has type (is type of)	E1 CRM Entity	E55 Type	<i>This property allows sub typing of CIDOC CRM entities – a form of specialisation – through the use of a terminological hierarchy, or thesaurus. [...]</i>	many to many (0,n:0,n)		Direta e Inversa

P4 has time-span (is time-span of)	E2 Temporal Entity	E52 Time-Span	<i>This property associates an instance of E2 Temporal Entity with the instance of E52 Time-Span during which it was on-going. The associated instance of E52 Time-Span is understood as the real time-span during which the phenomena making up the temporal entity instance were active. [...]</i>	many to one, necessary (1,1:0,n) (funcional)		Direta
P7 took place at (witnessed)	E4 Period	E53 Place	<i>This property describes the spatial location of an instance of E4 Period. The related instance of E53 Place should be seen as a wider approximation of the geometric area within which the phenomena that characterise the period in question occurred, see below. P7 took place at (witnessed) does not convey any meaning other than spatial positioning (frequently on the surface of the earth). [...]</i>	many to many, necessary (1,n:0,n)		Direta

P8 took place on or within (witnessed)	E4 Period	E18 Physical Thing	<p><i>This property describes the location of an instance of E4 Period with respect to an instance of E19 Physical Object.</i></p> <p><i>[...]</i></p> <p><i>It describes a period that can be located with respect to the space defined by an E19 Physical Object such as a ship or a building. The precise geographical location of the object during the period in question may be unknown or unimportant.</i></p>	many to many (0,n:0,n)		Direta
P10 falls within (contains)	E92 Spacetime Volume	E92 Spacetime Volume	<p><i>This property associates an instance of E92 Spacetime Volume with another instance of E92 Spacetime Volume that falls within the latter. In other words, all points in the former are also points in the latter.</i></p> <p><i>This property is transitive and reflexive.</i></p>	many to many, necessary, dependent (1,n:0,n)	Esta propriedade é transitiva e reflexiva.	Direta

P14 carried out by (performed)	E7 Activity	E39 Actor	<i>This property describes the active participation of an instance of E39 Actor in an instance of E7 Activity. [...]</i>	many to many, necessary (1,n:0,n)		Inversa
P37 assigned (was assigned by)	E15 Identifier Assignment	E42 Identifier	<i>This property records the identifier that was assigned to an item in an instance of P37 Identifier Assignment. The same identifier may be assigned on more than one occasion. An Identifier might be created prior to an assignment.</i>	many to many (0,n:0,n)		Direta
P42 assigned (was assigned by)	E17 Type Assignment	E55 Type	<i>This property records the type that was assigned to an entity by an E17 Type Assignment activity. [...] A Type may be intellectually constructed independent from assigning an instance of it.</i>	many to many, necessary (1,n:0,n)		Direta



P45 consists of (is incorporated in)	E18 Physical Thing	E57 Material	<p><i>This property identifies the instances of E57 Materials of which an instance of E18 Physical Thing is composed. 140. All physical things consist of physical materials.</i></p> <p><i>[...]</i></p>	many to many, necessary (1,n:0,n)		Direta
P46 is composed of (forms part of)	E18 Physical Thing	E18 Physical Thing	<p><i>This property associates an instance of E18 Physical Thing with another instance of Physical Thing that forms part of it. The spatial extent of the composing part is included in the spatial extent of the whole.</i></p> <p><i>[...]</i></p>	many to many (0,n:0,n)	Essa propriedade é transitiva e não-reflexiva	Direta e Inversa

P53 has former or current location of (is former or current location)	E18 Physical Thing	E53 Place	<p><i>This property identifies an instance of E53 Place as the former or current location of an instance of E18 Physical Thing.</i></p> <p><i>In the case of instances of E19 Physical Object, the property does not allow any indication of the Time-Span during which the instance of E19 Physical Object was located at this instance of E53 Place, nor if this is the current location.</i></p> <p><i>In the case of immobile objects, the Place would normally correspond to the Place of creation.</i></p> <p><i>[...]</i></p>	many to many, necessary (1,n:0,n)		Inversa
---	--------------------	-----------	--	-----------------------------------	--	---------

P56 bears feature (is found on)	E19 Physical Object	E26 Physical Feature	<p><i>This property links an instance of E19 Physical Object to an instance of E26 Physical Feature that it bears.</i></p> <p><i>An instance of E26 Physical Feature can only exist on one object. One object may bear more than one E26 Physical Feature. An instance of E27 Site should be considered as an instance of E26 Physical Feature on the surface of the Earth.</i></p> <p><i>[...]</i></p>	one to many, dependent (0,n:1,1)		Direta
------------------------------------	---------------------------	----------------------	---	-------------------------------------	--	--------

P67 refers to (is referred to by)	E89 Propositional Object	E1 CRM Entity	<p><i>This property documents that an instance of E89 Propositional Object makes a statement about an instance of E1 CRM Entity. P67 refers to (is referred to by) has the P67.1 has type link to an instance of E55 Type. This is intended to allow a more detailed description of the type of reference. This differs from P129 is about (is subject of), which describes the primary subject or subjects of the instance of E89 Propositional Object.</i></p>	many to many (0,n:0,n)		Direta e Inversa
-----------------------------------	--------------------------	---------------	--	------------------------	--	------------------

P70 documents (is documented in)	E31 Document	E1 CRM Entity	<p><i>This property describes the CRM Entities documented as instances of E31 Document.</i></p> <p><i>Examples:</i></p> <p><i>Documents may describe any conceivable entity, hence the link to the highest-level entity in the CIDOC CRM class hierarchy. This property is intended for cases where a reference is regarded as making a proposition about reality. This may be of a documentary character, in the scholarly or scientific sense, or a more general statement.</i></p>	many to many, necessary (1,n:0,n)		Direta
----------------------------------	--------------	---------------	---	-----------------------------------	--	--------

P81 ongoing throughout	E52 Time-Span	E61 Time Primitive	<p><i>This property associates an instance of E52 Time-Span with an instance of E61 Time Primitive specifying a minimum period of time covered by it. Since Time-Spans may not have precisely known temporal extents, the CIDOC CRM supports statements about the minimum and maximum temporal extents of Time-Spans. This property allows a Time-Span's minimum temporal extent (i.e., its inner boundary) to be assigned an E61 Time Primitive value. Time Primitives are treated by the CIDOC CRM as application or system specific date intervals, and are not further analysed. If different sources of evidence justify different minimum extents without contradicting each other, the smallest interval including all these extents</i></p>	many to one, necessary (1,n:0,n)		Direta
------------------------	---------------	--------------------	---	----------------------------------	--	--------

			<i>will be the best estimate. This should be taken into account for information integration.</i>			
P89 falls within (contains)	E53 Place	E53 Place	<i>This property identifies an instance of E53 Place that falls wholly within the extent of another instance of E53 Place. It addresses spatial containment only and does not imply any relationship between things or phenomena occupying these places.</i>	many to many, necessary, dependent (1,n:0,n)	Essa propriedade é transitiva e reflexiva	Direta

P94 has created (was created by)	E65 Creation	E28 Conceptual Object	<p><i>This property links an instance of E65 Creation to the instance of E28 Conceptual Object created by it.</i></p> <p><i>It represents the act of conceiving the intellectual content of the instance of E28 Conceptual Object. It does not represent the act of creating the first physical carrier of the instance of E28 Conceptual Object. As an example, this is the composition of a poem, not its commitment to paper.</i></p>	one to many, necessary, dependent (1,n:1,1)		Direta
-------------------------------------	-----------------	--------------------------	--	---	--	--------



P102 has title (is title of)	E71 Human-Made Thing	E35 Title	<p><i>This property associates an instance of E35 Title has been applied to an instance of E71 Human-Made Thing.</i></p> <p><i>The P102.1 has type property of the P102 has title (is title of) property enables the relationship between the title and the thing to be further clarified, for example, if the title was a given title, a supplied title etc.</i></p> <p><i>It allows any human-made material or immaterial thing to be given a title. It is possible to imagine a title being created without a specific object in mind.</i></p>	many to many (0,n:0,n)		Direta
------------------------------	-------------------------	-----------	---	---------------------------	--	--------

<p>P108 has produced (was produced by)</p>	<p>E12 Production</p>	<p>E24 Physical Human-Made Thing</p>	<p><i>This property identifies the instance of E24 Physical Human-Made Thing that came into existence as a result of the instance of E12 Production.</i></p> <p><i>The identity of an instance of E24 Physical Human-Made Thing is not defined by its matter, but by its existence as a subject of documentation.</i></p> <p><i>An E12 Production can result in the creation of multiple instances of E24 Physical Human-Made Thing.</i></p>	<p>one to many, necessary, dependent (1,n:1,1)</p>		<p>Inversa</p>
--	-----------------------	--------------------------------------	--	--	--	----------------

P127 has broader term (has narrower term)	E55 Type	E55 Type	<p><i>This property associates an instance of E55 Type with another instance of E55 Type that has a broader meaning.</i></p> <p><i>It allows instances of E55 Types to be organised into hierarchies. This is the sense of "broader term generic (BTG)" as defined in ISO 25964-2:2013 (International Organization for Standardization 2013).</i></p> <p><i>This property is transitive</i></p>	many to many (0,n:0,n)	Essa propriedade é transitiva.	Inversa
P129 is about (is subject of)	E89 Propositional Object	E1 CRM Entity	<p><i>This property documents that an instance of E89 Propositional Object has as subject an instance of E1 CRM Entity.</i></p> <p><i>This differs from P67 refers to (is referred to by), which refers to an instance of E1 CRM Entity, in that it describes the primary subject or subjects of an instance of E89 Propositional Object.</i></p>	many to many (0,n:0,n)		Direta

P135 created type (was created by)	E83 Type Creation	E55 Type	<i>This property identifies the instance of E55 Type, which is created in an instance of E83 Type Creation activity</i>	one to many, necessary (1,n:0,1)		Direta
P136 was based on (supported type creation)	E83 Type Creation	E1 CRM Entity	<i>This property identifies one or more instances of E1 CRM Entity that were used as evidence to declare a new instance of E55 Type. The examination of these items is often the only objective way to understand the precise characteristics of a new type. Such items should be deposited in a museum or similar institution for that reason. The taxonomic role renders the specific relationship of each item to the type, such as "holotype" or "original element".</i>	many to many (0,n:0,n)		Inversa

P165 incorporates (is incorporated in)	E73 Information Object	E90 Symbolic Object	<i>This property associates an instance of E73 Information Object with an instance of E90 Symbolic Object (or any of its subclasses) that was included in it. [...]</i>	many to many (0,n:0,n)	Essa propriedade é transitiva quando restrita a instâncias de E73 Information Object	Direta
---	------------------------------	---------------------	---	---------------------------	---	--------