UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

CENTRO TECNOLÓGICO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Jussara Teixeira

Ontological Representation of Archival Records

Vitória, ES

2025

Jussara Teixeira

Ontological Representation of Archival Records

Master's dissertation presented to the Computer Science Graduate Program of the Federal University of Espírito Santo, as partial requirement to obtain the Computer Science Master's Degree.

Federal University of Espírito Santo – UFES

Technology Faculty

Postgraduate Program in Computer Science

Supervisor: Prof. Dr. Vítor E. Silva Souza

Co-supervisor: Prof. Dr. João Paulo A. Almeida

Vitória, ES

2025

Ficha catalográfica disponibilizada pelo Sistema Integrado de
Bibliotecas - SIBI/UFES e elaborada pelo autor

T266o

Teixeira, Jussara, 1969-
Ontological Representation of Archival Records / Jussara Teixeira. - 2025.
135 f. : il.

Orientador: Vítor Estêvão Silva Souza.
Coorientador: João Paulo Andrade Almeida.
Dissertação (Mestrado em Informática) - Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Documento arquivístico. 2. Arquivologia. 3. Ontologia. I. Souza, Vítor Estêvão Silva. II. Almeida, João Paulo Andrade. III. Universidade Federal do Espírito Santo. Centro Tecnológico. IV. Título.

CDU: 004

Ontological Representation of Archival Records

Jussara Teixeira

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo como requisito parcial para a obtenção do grau de Mestre em Informática.

Aprovada em **28 de fevereiro de 2025**.

Documento assinado digitalmente
gov.br
VITOR ESTEVAO SILVA SOUZA
Data: 06/03/2025 16:21:10-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. VÍTOR E. SILVA SOUZA
Orientador

Documento assinado digitalmente
gov.br
JOAO PAULO ANDRADE ALMEIDA
Data: 11/03/2025 15:57:10-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. JOÃO PAULO A. ALMEIDA
Coorientador

Documento assinado digitalmente
gov.br
ELIAS SILVA DE OLIVEIRA
Data: 12/03/2025 19:52:36-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. ELIAS SILVA DE OLIVEIRA
Membro Interno

Documento assinado digitalmente
gov.br
JEAN REMI BOURGUET
Data: 14/03/2025 18:06:36-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. JEAN-RÉMI BOURGUET
Membro Externo

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Vitória/ES, 28 de fevereiro de 2025

I dedicate this work to those who have profoundly shaped my personal and professional journey, whose legacy continues to inspire me, even though they are no longer in this world. Their memories and teachings live on in every achievement, illuminating my path and strengthening my resolve.

Acknowledgements

First of all, I thank my God for being my strength and guide throughout this journey. His presence has been my foundation in moments of uncertainty and the constant inspiration to overcome challenges and move forward. Without Him, none of this would have been possible.

I would like to thank my supervisors, Prof. Dr. Vítor E. Silva Souza and Prof. Dr. João Paulo A. Almeida, for their patience, wisdom, and dedication throughout this work. Their guidance and knowledge were essential in transforming this project into a solid and meaningful piece of research. Their contribution was invaluable, both academically and personally.

My special thanks go to Prof. Dr. Jean-Rémi Bourguet, who introduced me to the fascinating world of ontologies. His teaching was the starting point for the construction of this work; it awakened in me an interest in and understanding of this complex and essential area.

To colleagues and professionals in the field of archival science who shared their time and knowledge, offering valuable contributions to this work. Their commitment to the preservation and organization of archival records was a constant inspiration and reinforced the importance of every detail in this research.

I am deeply grateful to my research colleagues, particularly Diovani Favoreto and Melissa Zorzanelli, with whom I embarked on this journey, shared ideas, and offered unwavering support. Your company made the journey lighter and more meaningful.

To my husband, for his love, patience, and unconditional support. You have been my anchor throughout this journey, offering me strength at times when I needed it most and reminding me of the value and importance of this achievement.

This project is the result of an individual but collective effort, as each page reflects the affection and dedication of everyone I see shining through in each small achievement and whose light has inspired me to keep going, even in the most challenging moments. To everyone, my deep and eternal gratitude.

In this lifetime, I can't repay all of your kindness.
May in the next life, I'll turn into the soil to grow flowers in spring.
(The Yin-Yang Master: Dream of Eternity)

"Sócrates contou como a divindade que inventou a escrita foi
repreendida por Tamuz, rei do Egito:
Se os homens aprenderem isto, estará implantado o esquecimento em suas almas:
eles deixarão de exercitar a memória porque confiarão no que está escrito e chamarão as
coisas à lembrança não mais de dentro de si, mas por meio de marcas externas;
o que descobriste é um remédio não para a memória, mas para a lembrança."
(Luciana Duranti - Registros documentais contemporâneos como provas de ação.)

Abstract

This master's dissertation presents a conceptual modeling approach for the domain of archival records, motivated by the challenges of interoperability and long-term preservation of archival records in digital environments. The transition from physical to digital or born-digital archival records has significantly increased the complexity of their management, requiring solutions that seek to guarantee the interoperability, accessibility and authenticity of archival records over time. In response to these challenges, this research presents a specific ontology to represent the archival record, with the aim of facilitating semantic interoperability between various public administration information systems. The study begins by establishing a theoretical framework that encompasses the fundamental concepts of archival record characteristics, the role of ontologies in the field of Archival Science and the specific approaches employed in the construction of ontologies. A systematic literature mapping was carried out using the CLeAR approach to identify existing structured resources that could contribute to the proposed ontology. The process of building the Archival Record Ontology (ARO) integrates ontological resources such as OntoUML, SABiO, UFO and gUFO, and the Visual Paradigm tool, which together helped to build the representation of the archival record. The ontology verification process was carried out using Protégé, a tool that not only facilitates verification, but also supports other important tasks in ontology development. SPARQL queries were implemented to verify the ontology's ability to correctly answer the integration questions that were formulated. The results of this research contribute to the advancement of archival science by providing a semantically rich ontology that aids the management, preservation and interoperability of archival records. This work not only enhances theoretical understanding, but also demonstrates the practical use of conceptual modeling, using ontologies to address archival concepts.

Keywords: Archival Record. Archival Science. Ontology. UFO. Conceptual Modeling.

Resumo

Esta dissertação de mestrado apresenta uma abordagem de modelagem conceitual para o domínio dos documentos arquivísticos, motivada pelos desafios da interoperabilidade e preservação a longo prazo de documentos arquivísticos em ambientes digitais. A transição de documentos arquivísticos físicos para digitais ou nato-digitais, aumentou significativamente a complexidade do seu gerenciamento, exigindo soluções que busquem garantir a interoperabilidade, acessibilidade e autenticidade dos documentos arquivísticos ao longo do tempo. Em resposta a esses desafios, esta pesquisa apresenta uma ontologia específica para representar o documento arquivístico, com o objetivo de facilitar a interoperabilidade semântica entre diversos sistemas de informação da administração pública. O estudo inicia-se com o estabelecimento de um quadro teórico que engloba os conceitos fundamentais de características do documento arquivístico, o papel das ontologias no campo da Arquivologia e as abordagens específicas empregadas na construção de ontologias. Um mapeamento sistemático de literatura foi realizado usando a abordagem CLeAR para identificar recursos estruturados existentes que poderiam contribuir para a ontologia proposta. O processo de construção da Ontologia de Registro Arquivístico (ARO), integra recursos ontológicos como OntoUML, SABiO, UFO e gUFO, e a ferramenta Visual Paradigm, que juntos ajudaram a construir a representação do registro arquivístico. O processo de verificação da ontologia foi realizado com o Protégé, uma ferramenta que não apenas facilita a verificação, mas também oferece suporte a outras tarefas importantes no desenvolvimento da ontologia. Para isto consultas SPARQL foram implementadas para verificar a capacidade da ontologia de responder corretamente às questões de integração formuladas. Os resultados desta pesquisa contribuem para o avanço da arquivística, ao fornecer uma ontologia semanticamente rica que auxilia o gerenciamento, preservação e interoperabilidade de registros arquivísticos. Este trabalho não apenas aprimora a compreensão teórica, mas também demonstra o uso prático da modelagem conceitual, utilizando ontologias para abordar conceitos arquivísticos.

Palavras-chaves: Documento Arquivístico. Arquivologia. Ontologia. UFO. Modelagem Conceitual.

List of Figures

List of Tables

List of abbreviations and acronyms

5W1H	Who, What, When, Where, Why, and How
ArCO	Italian Cultural Heritage Knowledge Graph
ArDO	Archive Dynamics Ontology
ARKIVO	ARKIVO ontology
ARO	Archival Record Ontology
BDA	Born-Digital Archives
BFO	Basic Formal Ontology
BibO	Bibliographic Ontology
CAPES	Coordination for the Improvement of Higher Education Personnel
CEDARS	CURL Exemplars in Digital ARchiveS
CIDOC	International Committee for Documentation
CIDOC-CRM	CIDOC Conceptual Reference Model
CLeAR	Conducting Literature Search for Artifact Reuse
CQ	Competency Question
D-Acts	Ontology of Document Acts
DCMI	Dublin Core Metadata Initiative
EDRMS	Electronic Document and Records Management System
EGAD	Expert Group on Archival Description
EU	European Union
FaBiO	FRBR-aligned Bibliographic Ontology
FAIR	Findable, Accessible, Interoperable, and Reusable
FRBR	Functional Requirements for Bibliographic Records
GAO	Chinese Government Archive Ontology

gUFO	A Lightweight Implementation of the Unified Foundational Ontology
HERO	Historical Event Representation Ontology
IAO	Information Artifact Ontology
ICA	International Council on Archives
ICOM	International Council of Museums
IQ	Integration Question
ISO	International Organization for Standardization
KOS	knowledge organization systems
LRM	Linked-Resource-Model
PERICLES	PERICLES Linked Resource Model
OntoUML	Ontologically Well-Founded Profile for UML
OAI	Open Archives Initiative
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
OPMO	Open Provenance Model Ontology
OWL	Web Ontology Language
PAV	Provenance, Authoring and Versioning Ontology
PREMIS-O	PREMIS OWL Ontology
PROV-O	PROV Ontology
RiC-O	Records in Contexts Ontology
RDF	Resource Description Framework
SABiO	Systematic Approach for Building Ontologies
SBA	Software-Based Art
SLR	Systematic Literature Review
SPARQL	SPARQL Protocol and RDF Query Language
TIMBUS	TIMBUS context model
UFO	Unified Foundational Ontology

UML	Unified Modeling Language
UNESCO	United Nations Educational, Scientific and Cultural Organization
URI	Uniform Resource Identifiers
W3C	World Wide Web Consortium
XML	Extensible Markup Language

Contents

1 Introduction

This chapter introduces the context in which the present research is situated, with an emphasis on contemporary challenges related to interoperability and the management of archival records in the digital environment, particularly within public administration systems. The increasing complexity and volume of electronic records demand approaches that ensure not only their preservation but also the long-term integration and efficient access to archival records. The chapter is structured to provide an overview of the current state of the field, establishing the necessary context for understanding the challenges faced. Following this, the motivation for conducting this research is discussed, highlighting the factors that have driven its development. Subsequently, the general objective and specific objectives of the study are outlined, emphasizing the goals that the research aims to achieve. The study's methodology is presented, followed by a discussion of the contribution the study intends to offer, advancing both theoretical knowledge and effective practices in the management of archival record. Finally, this chapter concludes by presenting the structure of this work, offering an outline of how this master's thesis is organized.

1.1 Context

The concept of archive has undergone a long historical evolution, whose roots can be traced back to the ancient Greek term arkheion (Derrida, 1996), referring to the place where public records were kept under the authority of magistrates. Over the centuries, the notion of archive expanded, incorporating different meanings and practices influenced by juridical, administrative, and cultural developments. While an etymological and historical discussion offers valuable insights into the evolution of archival thought, this study adopts a contemporary definition of archive, centered on its institutional role and relevance in the digital age (ISO, 2016; Mbembe, 2002). This focus is aligned with the objective of the research, which seeks to address the current challenges of representing archival records in information systems and promoting semantic interoperability in digital environments.

An archive is the place where records of business activities and other information assets are preserved (ISO, 2016; Mbembe, 2002), but archives, which are essential for preserving institutional and collective memory, face growing challenges in the digital age. The transition from physical to digital records has not only transformed archival practice but has also significantly increased the complexity of the information systems that need to be managed (Mazikana, 1997; Colavizza *et al.*, 2021). For example, in a public administration archive, there are records with information related to public procurement, public health, education, security, agreements between government agencies, strategic planning, etc. An archival record has sufficient constituent elements to serve as evidence of activities performed by a variety of institutional actors and may become part of the archival heritage of a society (SAA, 2005-2023a). It can serve to support the various activities of an organization and can be recognized as a primary source for research work or as a historical record of the institutional memory of an entity.

Recognizing the important role of archives for society, in 1983, the United Nations Educational, Scientific and Cultural Organization (UNESCO) developed a study to provide information to decision makers about the essential character and value of archives and records. That study already identified electronic records as an important element of modern archives (Rhoads, 1983). In commemoration of the International Archives Day 2023, UNESCO reaffirms that archives have a permanent impact and importance for society and reiterates its mission to raise public awareness about archives and the risks and challenges they face (UNESCO, 2023a; UNESCO, 2023b).

The contemporary archival practices have undergone significant transformations in recent years due to advancements in technology, changing social and cultural contexts, and a rethinking of the role of the archivist (Cook, 2001; Cook, 2013; Gilliland; McKemmish; Lau, 2016). After almost 40 years, the electronic management of archival records has become inevitable for organizations of all sorts, and key to areas such as e-Government. Given the critical role of records to the functioning of an organization, there is a need to implement strategies to preserve the myriad of born-digital records created and maintained in various (interconnected) information systems.

1.2 Motivation

The motivation for this study stems from the growing need to develop solutions that enable interoperability and data integration between heterogeneous information systems, particularly in the context of digital archival records and their long-term digital preservation. Preservation of digital records requires special attention to interoperability, mostly because of the diversity of supporting systems. According to Farinelli, Almeida & Melo (2013), "the issue of interoperability among information systems has been on the agenda of several governments around the world for many years, including Brazil". For Bastola & Campus (2020), "the representation of heterogeneous systems with common standards and vocabularies is an effective solution for interoperability". From the same point of view, Tshering & Anutariya (2022) argue that "ontologies give an outstanding way to represent reality".

Ontologies can be used as tools to build machine-interpretable semantic representations of domain knowledge (Liyanage; Krause; Lusignan, 2015), such as in the case of e-Government archival records. Additionally, Barros & Gomes (2018) argue that the relationship between knowledge organization and archival science is fundamental. The authors emphasizes the importance of representing and organizing the information contained in archival records, providing a theoretical foundation essential for understanding the application of ontologies in this context.

Furthermore, domain ontologies provide a framework for representing knowledge about a specific domain, so by defining concepts, relationships and properties within the domain, ontologies help to organize and categorize archival records more effectively (Barros; Sousa, 2019b). In this regard, Barros & Sousa (2019a) discuss the importance of ontologies in the management of archival records, emphasizing that the development of archival classification and description is essential for the construction of ontologies, taxonomies, and controlled vocabularies, among other tools. Hedstrom (1997) argues that effective digital preservation depends on the ability to capture not only the data, but also the context and relationships that give it meaning. In this sense, building a specific ontology for the archival records domain, with a focus on the archival record, can facilitate the interoperability, long-term preservation, and access of these records, ensuring that their integrity and authenticity are maintained.

Ontologies can be used as tools to construct machine-interpretable semantic representations of knowledge in a specific domain (Liyanage; Krause; Lusignan, 2015), such as in the case of archival records produced within e-government initiatives. According to Gruber (1993), an ontology is "an explicit specification of a conceptualization," and its application enables the establishment of a common language across systems, promoting knowledge sharing and reuse. One of the core benefits of using ontologies lies precisely in their ability to align different controlled vocabularies, ensuring that equivalent terms used in distinct contexts are formally reconciled. This aspect is particularly relevant in archival environments, where multiple standards, metadata models, and terminologies coexist and must be integrated to ensure semantic interoperability. Thus, the development of a specific ontology for the domain of archival records provides not only a rigorous conceptual representation but also a mechanism to semantically integrate and align dispersed information.

1.3 Objectives

This work has the main objective *of proposing an ontology to represent the domain of archival records management, with the aim of enabling interoperability between different systems and contributing to long-term digital preservation.* This main objective can be detailed in the following specific objectives:

(i) Identify structured resources: identify existing structured resources, such as ontologies, reference models, and vocabularies, that will serve as references for constructing the ontology, ensuring alignment with practices in the domain of archival records;

(ii) Develop the ontology: construct the ontology, ensuring that it is theoretically grounded and conceptually accurate[1] , to effectively model the domain of archival records management;

(iii) Verify the ontology: carry out tests to ensure that its logical structure is consistent and that it can answer the integration questions formulated, ensuring the coherence of the model within the proposed context.

1.4 Approach

The method used for this work was structured into three stages, each aligned with the specific objectives defined for the research. These stages are: *(i)* a systematic review to identify the structured resources, *(ii)* the application of modeling techniques and principles for constructing the ontology, and *(iii)* conducting evaluations to verify the proposed ontology.

In the first stage of this study, a systematic mapping was be conducted using the Conducting Literature Search for Artifact Reuse (CLeAR) method (Campos *et al.*, 2020; Campos, 2019). Developed to systematically search for reusable structured resources in the literature, this method ensures a reproducible data collection process. The focus of this stage is the identification and analysis of existing structured resources, such as ontologies, reference models, and controlled vocabularies, related to the management, interoperability, and preservation of archival records. Through this mapping, the aim is not only to identify practices and standards utilized in the field but also to extract insights that may assist in the theoretical and practical foundation of the proposed ontology's development.

The second stage involves the construction of the ontology, which was developed following the Systematic Approach for Building Ontologies (SABiO) (Falbo, 2014). This approach provides a systematic process for ontology development, encompassing everything from scope definition to final implementation. In this stage, the Unified Foundational Ontology (UFO) (Guizzardi, 2005; Guizzardi *et al.*, 2015) served as the theoretical foundation, offering the core ontological concepts. UFO has the capability to represent concepts with precision and consistency across different domains. Additionally, the OntoUML (Guizzardi, 2005; Guizzardi *et al.*, 2018) ontologically well-founded profile for UML class diagrams was used as the conceptual modeling language to graphically represent the proposed concepts and relationships for the ontology. The combination of these approaches ensures that the

[1] By *conceptually accurate*, refers to the faithful representation of domain entities, relations, and constraints based on ontological distinctions grounded in a foundational ontology (e.g., UFO). This type of accuracy emphasizes semantic adequacy and theoretical alignment with both the target domain and the modeling methodology adopted.

ontology is well-grounded and appropriately applied to the specific domain intended for modeling. By integrating these approaches, the ontology represents key aspects of the lifecycle of an archival record, such as its creation, signing, sending, and receiving, ensuring a well-founded conceptual structure that aligns with the principles of archival science and supports the long-term preservation and interoperability of electronic archival records. Although the concept of the archival record lifecycle is intrinsic to Archival Science and closely associated with tools such as the Records Retention and Disposition Schedule, this aspect was not explicitly addressed in the scope of the proposed ontology. The modeling focused on representing the key events that characterize the lifecycle trajectory of archival records — including creation, signing, sending, and receiving — in line with the primary objective of ensuring interoperability and long-term digital preservation. The representation of processes related to retention, disposition, and appraisal remains an important direction for future extensions of the ontology.

The third and final stage consists of verifying the constructed ontology. In accordance with the recommendations of the SABiO approach, empirical verification was adopted as the main strategy, through the formulation and execution of SPARQL[2] queries (SPARQL Protocol and RDF Query Language) (Arenas; Pérez, 2011; W3C, 2013). This verification aimed to assess the ontology's ability to adequately represent the domain and to answer the previously defined integration questions. It is important to clarify that, in this work, the term "verification" refers to the empirical validation of the ontology's applicability, and not to formal logical consistency checking, which involves detecting contradictions in the ontological model and is typically performed using automated reasoning mechanisms (reasoners). Therefore, the formulation of SPARQL queries constitutes a reasoning activity aimed at demonstrating the practical utility of the ontology in the archival context, without implying the execution of formal consistency tests in the strict logical sense.

1.5 Structure

The remainder of this dissertation is structured into the following chapters:

- **Chapter 2** presents the theoretical and methodological foundations of this research, exploring the domain of archival records and the use of methods for conceptual modeling using ontologies;
- **Chapter 3** details the application of the CLeAR approach to the archival records domain, describing the execution of the systematic search and an analysis of struc-

[2] SPARQL is the standard language for querying data on the Semantic Web, developed by the W3C (World Wide Web Consortium), and is designed to work on RDF graphs and enable efficient retrieval

tured resources, completing with a comparative analysis between CIDOC Conceptual Reference Model (CIDOC-CRM) and Records in Contexts Ontology (RiC-O). These two models were selected because of their importance in the field of cultural heritage and in the description of archival record resources;

- **Chapter 4** focuses on the ontological representation of an archival record, applying the concepts previously discussed. The chapter details the ontology's conceptual foundation, including its classes, stereotypes, and relationships, ensuring adherence to the characteristics of archival records. Additionally, explores the alignment of Archival Record Ontology (ARO) with structured resources identified through the CLeAR approach and provides a comparative analysis between ARO, RiC-O, and CIDOC-CRM, highlighting their conceptual and structural differences;

- **Chapter 5** details the verification process of the Archival Records Ontology (ARO) conducted using Protégé, where SPARQL queries were developed to evaluate the ontology's ability to answer the integration questions formulated;

- **Chapter 6** presents the final considerations, including a summary of the results, the applicability of the work, the limitations encountered, and suggestions for future research.

of information on the Semantic Web(Arenas; Pérez, 2011).

2 Background

This chapter presents the theoretical and methodological foundations of this research, exploring the domain of archival records and the use of methods for conceptual modeling using ontologies. The chapter begins with a presentation of the main characteristics of an archival record as discussed in the specialized literature. The chapter then introduces the CLeAR approach, applied in this research to systematically map existing literature and identify structured resources that can contribute to the ontology to be developed. In addition, the chapter presents the concept of ontology and its application in the field of Archival Science, followed by the ontological resources used in the development of the ontology, such as OntoUML, SABiO, UFO and gUFO, as well as the Visual Paradigm tool, which together enable the creation of consistent ontological models.

2.1 The Archival Records Domain

The Archival Records Domain, in this research, is a term used to describe the specific focus and expertise involved in managing and preserving archival records. Records management includes activities such as creation or capture, storage, tracking, and audit in order to meet the evidence requirements of an organization's activities. It also encompasses appropriate actions to protect the authenticity, integrity, reliability, usability, and accountability of records as the context and requirements of public administration change over time. Preservation of archival records is the systematic and proactive efforts performed to maintain the long-term survival, accessibility, and usability of significant records retained in archives.

When managing and preserving archival records, archivists and other professionals deal with an extensive list of characteristics present in archival records. These characteristics cover different areas of archival activities and contribute to the features of archival assets. As main characteristics in the domain of archival records, Santos (2012) enumerates *fixity*, *organic nature*, *naturalness*, *uniqueness*, *authenticity*, and *impartiality*.

The *fixity* of an archival record refers to its immutability, which means that it cannot be modified without clear indications of these modifications (Santos, 2012). According to Tikhonov (2019), there is a need to have confidence that the objects we have in storage are the same[1] as they were when we put them in there, which is one of the basic requirements of digital preservation. Closely related to fixity, the term integrity refers to the condition

[1] This concern is directly related to the concept of fixity, one of the essential characteristics of archival records discussed later in this dissertation. The ability to identify whether a digital object remains the same over time is typically addressed through technical mechanisms such as checksums and

of a record being complete, consistent, and reliable throughout its entire life cycle. While fixity addresses the immutability of content, integrity encompasses the preservation of all essential components of the record — including its structure, metadata, and contextual relationships — ensuring that it continues to serve as trustworthy evidence over time (PREMIS Editorial Committee, 2015; Lemieux; Trapnell, 2016). In digital environments, maintaining integrity requires a combination of technical mechanisms (such as checksums) and archival strategies that preserve not only the data itself but also its provenance, structure, and meaning.

Archival records play a crucial role within a comprehensive record keeping system, serving as tangible evidence of the functions, operations, and activities undertaken by an organization over time (Cook, 2013). Cook (2013), emphasizes the *organic* and interconnected *nature* of records, suggesting that each document is not an isolated entity but rather a part of a broader network of information (Cook, 2013). This interconnection implies that understanding a single record requires considering its relationship with other records within the archive. Furthermore, Santos (2012) explains that the *organic nature* describes the characteristic of the archival record of being an integral part of an organic system, reflecting the activities, functions, and structure of the entity or person that produced it and refers to the collection of connections that a record has with other records associated with the same activity (Santos, 2012).

In archival theory, the characteristic of *naturalness* could be interpreted as the arrangement organization, and structure of records as they were originally created, accumulated, or received in the course of personal or organizational activity (Eastwood, 1994; Cline, 2022). This natural order is considered essential for understanding the context and meaning of the records. Santos (2012) explains that *naturalness* is the characteristic of archival records that are not collected artificially but are produced and received in a natural process related to the interests of the institution and the legal conformity of its acts and are thus accumulated (Santos, 2012).

Another foundational principle of Archival Science is “respect des fonds”, which establishes that records must be arranged according to the entity that produced them, thereby preserving their provenance and original order (Schellenberg, 1956; Duranti, 1997). This principle ensures the maintenance of contextual relationships among records and prevents the fragmentation of coherent archival sets. Its adoption is essential not only for the physical and logical organization of archives but also for the conceptual representation of archival records in digital environments.

The digital revolution is reshaping our traditional notions of *uniqueness*. Instead of solely relying on documentary context, distinctions can now be drawn based on the

cryptographic hashes, which allow for fixity verification. These mechanisms will be examined in more detail in the sections on digital preservation.

production, maintenance, and usage contexts of records (Yeo, 2010). Although our understanding of *uniqueness* may require adjustments in the digital age, it is essential not to entirely discard this concept (Yeo, 2010). The concept of *uniqueness* suggests that in some areas, like art, an original work, such as a painting, is distinct and superior to all replicas or copies (Yeo, 2010). This principle has been historically applied by archivists to records, where the original is deemed unique and all other versions are considered copies (Yeo, 2010). Santos (2012) argues[2] that the archival record is *unique* in its function, even in the presence of multiple copies (Santos, 2012). Each record serves a distinct role related to its creation and use within the institution, maintaining its association with other records connected to the same activity. Thus, the characteristic of *uniqueness* is upheld (Santos, 2012).

The characteristic of *uniqueness* may also be related to broader ontological foundations, especially to the principle of identity, formulated in the Greek philosophical tradition. This principle states that "everything is identical to itself" ($A = A$)(McKeon, 1941). In Aristotle's classical logic, this principle serves as a foundation for distinguishing entities. In well-founded ontologies, this idea is operationalized through the notion of *sortal*, which provides ontological criteria for determining identity and distinction between instances (Guizzardi, 2005; Guarino; Welty, 2009).

The concept of *authenticity* of records is fundamental to archival science and has a centuries-old theoretical basis, being a critical concern in the fields of history, jurisprudence, and diplomacy (Rogers, 2016). *Authenticity* is the quality of archival documents to bear reliable testimony to the actions, procedures, and processes that brought them into being (Eastwood, 1992). Another definition explains that *authenticity* refers to the quality of being genuine, not counterfeit, and free of tampering. It is typically determined through internal and external evidence, including the record's physical characteristic, structure, content, and context (Santos, 2012; SAA, 2022), and the authenticity of original data is one of the most important aspects of the preservation strategy (Tikhonov, 2019).

Impartiality[3] merely means that the record accurately reflects what occurred and any tinkering with these documents might in some way interfere with this *impartiality* and create a misrepresentation of the whole (Boles; Greene, 1996). This property is explained by Santos (2012), and refers to its characteristic of being objective and unbiased in recording

[2] Even if digital or physical copies share the same content as the original, they may not fulfill the same archival function. This is because the archival value of a record is tied not only to its content, but also to contextual elements such as provenance, authenticity, and custody. Only the exemplar that retains these attributes can be considered the archival record in the strict sense Santos (2012).

[3] Some authors argue that the classical notion of impartiality in archival science should be reexamined in light of the challenges posed by digital culture and algorithmic mediation. For instance, Cook (Cook, 2013) and Ketelaar (Ketelaar, 2002) highlight that records are socially constructed and contextually influenced, making it necessary to rethink archival principles such as neutrality and impartiality in contemporary settings.

facts, events, or transactions. It is an attribute that records carry, in contrast to their creators, who inherently have biases in matters related to their personal interests (Santos, 2012). Therefore, this principle is crucial because records often serve as the foundation for historical research and accountability. They can shape our understanding of history, culture, and heritage. Therefore, maintaining *impartiality* helps to protect the integrity of the records and ensures they can be trusted by future generations.

Archival records are primary sources of information that provide evidence of past events, actions, and decisions, and their proper management and preservation is essential for administrative purposes, historical studies, and cultural heritage. However, when an archival record no longer presents some of the above characteristics, it fails to be adequate as a source of reference and evidence.

2.2 The CLeAR Approach

To find candidate structured resources to be reused, we used the CLeAR approach (Campos, 2019; Campos *et al.*, 2020), enabling the identification of structured resources that will serve as references for the domain being modeled, ensuring alignment with its specific requirements and characteristics. CLeAR follows some principles of systematic literature reviews, supporting the search for structured resources which can ultimately be reused (Campos *et al.*, 2020). Figure 1 gives an overview of the CLeAR approach and its activities. The approach is structured in three cycles.

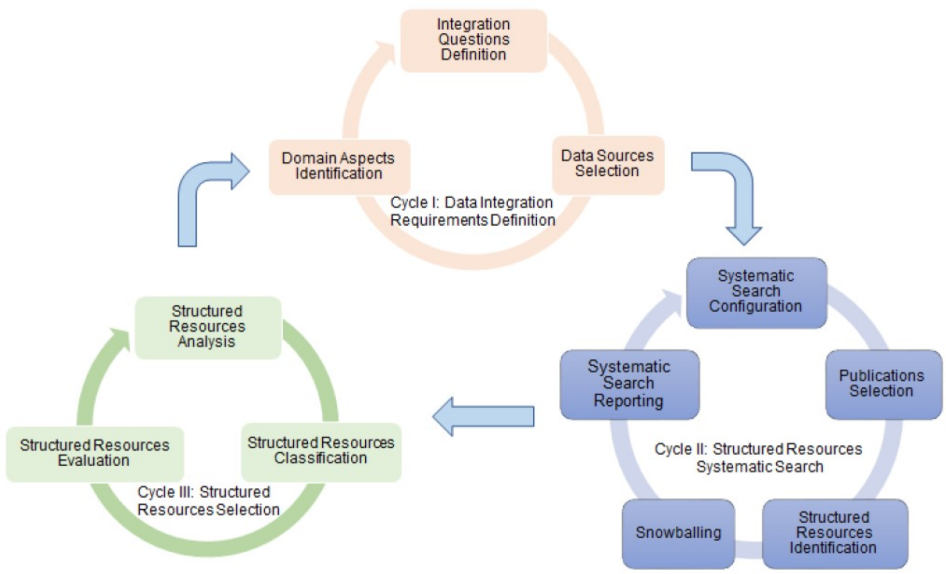

Figure 1 – Overview of CLeAR's cycles and activities (Campos, 2019).

CLeAR was defined with the premise that information needs usually cross-cut several aspects and subdomains of an overall knowledge domain. This materializes itself frequently in the need to integrate data from heterogeneous sources. Because of this, requirements for the literature search are formulated in three cycles that can be visited iteratively.

Cycle I (Data Integration Requirements Definition) focuses on raising integration questions and domain aspects that can be observed in actual data sources. The following activities are performed iteratively in this cycle: (a) Integration Questions Definition, (b) Data Sources Selection, and (c) Domain Aspects Identification.

Cycle II (Structured Resources Systematic Search) is focused on systematically identifying structured resources based on the requirements defined in Cycle I. This cycle has five activities, the first being (a) Systematic Search Configuration, which is focused on planning the search, including defining search objectives, repositories, designing and testing search strings and inclusion/exclusion criteria. This activity is then followed by three other activities: (b) Publications Selection, (c) Structured Resources Identification, and (d) Snowballing. Finally, the fifth activity, (e) Systematic Search Report, presents the results of the systematic search and is evaluated to verify whether the search goals were reached.

Cycle III (Structured Resources Selection) has the purpose of guiding the analysis of the structured resources found in Cycle II and is composed of three activities: (a) Structured Resources Analysis, (b) Structured Resources Classification, and (c) Structured Resources Evaluation.

An example of the use of the CLeAR approach was the identification of structured resources for the development of an ontology for the integration of water quality data (Campos *et al.*, 2020; Campos, 2019). The approach was used because of its unique advantages in ontology engineering such as systematic identification of reusable knowledge resources, objective evaluation based on quality attributes, and alignment with the needs of scientific research data integration (Campos *et al.*, 2020; Campos, 2019). It is important to recognize that the CLeAR approach may have limited applicability in domains where structured resources are scarce or have limited availability. Since it is based on the use of formal specifications and conceptualizations, it may not be suitable for all types of data in integration scenarios, because its effectiveness will depend on the quality and relevance of the literature available in the field of interest (Campos *et al.*, 2020; Campos, 2019). In the case of the archival records domain, there is significant literature on which to base the application of CLeAR.

2.3 Ontology

This subsection discusses the definition of ontology, describing its fundamentals as a formal structure for organizing concepts and their interrelationships in specific domains, as well as presenting its use within the context of Archival Science.

2.3.1 Definition

Ontology, in its philosophical origin, refers to the study of being, that is, the investigation of the entities that make up reality and their interrelations (Smith, 2002; Guizzardi, 2007; Hebeche, 2012). According to Guarino, Oberle & Staab (2009), ontology is defined as the science of "being *qua* being", that is, the study of the attributes that belong to objects and entities because of their inherent nature. Building on this distinction, Guarino, Oberle & Staab (2009) further clarify the difference between "Ontology" with a capital "O", which refers to the philosophical study of the nature of being, and "ontology" with a lowercase "o", which relates to computational artifacts used to formalize conceptual structures in specific domains. Guizzardi (2007) presents a more applied definition of the term, particularly within the fields of Computer Science and Software Engineering where ontologies are employed to formally represent the categories and concepts of a specific domain. As Guizzardi (2007) explains, "an ontology can be seem as the metamodel specification for an ideal language to represent phenomena in a given domain in reality, i.e., a language which only admits specifications representing possible state of affairs in reality".

Gruber (1993) adopts the definition of ontology as "an explicit specification of a conceptualization", playing a central role in establishing a common language to describe the elements of a domain, facilitating the sharing, reuse, and integration of data among heterogeneous systems. In this context, a conceptualization refers to a simplified abstraction of the world that describes the objects, concepts, and relations that exist within a specific domain of discourse. Guarino (1998) clarifies the distinction between the concepts of "ontology" and "conceptualization". While conceptualization represents an abstract and simplified view of the world, ontology is a formal specification of that conceptualization where the terms in the formal vocabulary are rigorously defined and shared in a precise manner. This means that ontology provides a formal foundation for knowledge representation, allowing multiple systems to share and interpret the same concepts consistently.

In addition, Gruber (1993) underscores that, by using a standardized or ontology-aligned formal vocabulary, data produced by one system can be consumed and understood by another without compromising its semantic integrity. In this sense, ontologies become indispensable in the context of interoperable systems, ensuring that information exchange occurs efficiently and without loss of meaning, regardless of differences in the representation mechanisms employed.

Guarino (1994) introduces the idea that, to ensure accuracy and interoperability among knowledge systems, ontologies must formalize these assumptions clearly. Ontological commitment, therefore, involves deciding which entities and relations should be represented and how these entities are organized within an ontology. Moreover, in the same work, Guarino (1994) argues that an ontology must be more than just a taxonomy of concepts; it must specify the fundamental categories that govern how concepts and relationships are structured, in addition to determining the necessary conditions for identity and distinctions between entities. This commitment is essential to avoid ambiguities and to ensure that terms are interpreted uniformly and coherently in different application contexts.

Years later, Guarino, Oberle & Staab (2009) revisit the concept of ontological commitment in the context of 30 years of advancements in knowledge representation, reaffirming the importance of constructing ontologies that incorporate formal ontological distinctions. Despite the progress made, the authors emphasize that many representation systems still fail to explicitly articulate their ontological commitments, leaving assumptions about the meaning of terms implicit. They reinforce the importance of a well-defined ontological commitment for ensuring semantic clarity and interoperability among systems, proposing the formalization of these distinctions through well-founded ontologies.

2.3.2 Ontology in Archival Science

In recent years, Archival Science has increasingly relied on ontologies to enhance the management, organization, and access to administrative and historical records. This has led to more efficient and precise representation, classification, and retrieval of information. This application of ontologies is expressed in various aspects, such as: the improvement of archival management and access, enabling faster and more accurate record retrieval; the classification and organization of knowledge, structuring information in a logical and standardized manner; and finally, the modeling and extraction of knowledge, allowing archival data to be interpreted and reused in different contexts; and ensuring interoperability between systems.

According to Elena *et al.* (2007), Goy *et al.* (2020), and Hawkins (2022), the introduction of modern approaches, such as ontologies, semantic metadata, and Linked Data, enhances both the management of archives and the efficient access and exploration by users. Elena *et al.* (2007) highlight the importance of representing the temporal evolution of classes and instances, which facilitates understanding the history and transformations of entities, thus promoting a more precise contextualization of the data. Goy *et al.* (2020), in turn, demonstrate how the implementation of an ontology-driven user interface, combined with semantic metadata, improves the organization of archives and offers a more intuitive and efficient search experience, allowing for the identification of complex relationships between historical documents. Hawkins (2022) complements this by exploring the use of Linked Data to integrate and interconnect digitized and born-digital archives, adhering to Findable, Accessible, Interoperable, and Reusable (FAIR) principles, which significantly enhances the interoperability and reusability of archival data, enabling more complex queries and the discovery of new connections between documents.

For Medina, Chavez-Aragon & Chavez (2006), Barros & Gomes (2018), and Barros & Sousa (2019b), the use of ontologies and knowledge organization systems (KOS) enhances the ability to access and retrieve information more effectively. They recognize the importance of integrating concepts from fields such as Information Science, Computer Science (in the case of ontologies), and Archival Science, demonstrating that interactions between these disciplines can produce better methodologies for the classification and organization of archives. Medina, Chavez-Aragon & Chavez (2006), in particular, highlight the semi-automatic construction of ontologies of records using data mining techniques and hierarchical clustering algorithms to organize documents and optimize the information retrieval process in digital archives. Barros & Gomes (2018) emphasize that the creation of terminological ontologies prior to the implementation of an archival classification system makes the system more comprehensive, flexible, and user-friendly, allowing archivists to clearly visualize institutional functional structures and dynamically adapt classification systems as needed. Barros & Sousa (2019b), in turn, explore the intersections between archival science and knowledge organization, arguing that the development of archival classification and description systems can benefit from knowledge organization methodologies, such as ontologies and taxonomies, to improve the representation and access to archival documents, especially in the digital context.

Munir & Anjum (2018), Zarubin, Koval & Moshkin (2020), and Eswaraiah & Syed (2023) highlight that ontological modeling is an effective approach to organizing and extracting knowledge from archival documents. This approach helps improve the precision and efficiency of information retrieval by providing a semantic structure that facilitates the extraction of relevant data and the execution of precise queries on large volumes of archival documents. Munir & Anjum (2018), in particular, emphasize the role of ontologies in creating a semantic metadata layer that connects terms, concepts, and relationships within documents, allowing the information retrieval process to go beyond simple keyword matching and employ more sophisticated semantic queries. Zarubin, Koval & Moshkin (2020), in turn, present an ontological model specifically developed to classify large volumes of unstructured information in electronic archives, using algorithms to structure documents and improve data retrieval. The authors demonstrate that this approach enhances the efficiency and quality of archival document classification. Eswaraiah & Syed (2023) complement this view by proposing an optimized ontological model that, together with query execution, improves the accuracy of content extraction from documents. They highlight that ontology is essential for minimizing semantic data loss and improving query performance, achieving high accuracy rates in their experiments.

Another example of research using ontology in the context of semantic representation of archival documentary production is presented by Löw (2021). This research was based on archival concepts and principles, such as provenance and context of document production, analyzing their relationship with Basic Formal Ontology (BFO) entities and derived ontologies, such as the Information Artifact Ontology (IAO) and the Ontology of Document Acts (D-Acts). The experiment was carried out within the context of records production in an oil production sharing contract in the oil and gas sector, illustrating the feasibility of the ontological approach and its capacity to enhance traditional archive description tools. The findings demonstrate that employing ontologies improves the semantic precision of archive context representation, promoting more accurate and interoperable record modeling within information systems. The research carried out by Löw (2021) contributes to the improvement of ontological modeling in the archival domain and to the development of tools for document management, expanding the possibilities for the organization, retrieval and preservation of archival information.

The research of Luz (2016) reinforces the central role of ontologies in Archival Science, particularly in addressing the challenges of interoperability and long-term preservation of archival information. His research demonstrates how integrating descriptive standards, such as General International Standard Archival Description (ISAD(G)) and Brazilian Standard for Archival Description (NOBRADE), into digital ontologies enhances semantic alignment between systems, ensuring that archival data can be reliably exchanged, preserved, and accessed over time. The findings highlight the importance of ontologies not only for structuring archival information but also for maintaining the authenticity and integrity of archival records through the preservation of the chain of custody, ensuring that archival knowledge remains accessible and trustworthy in a rapidly evolving digital landscape (Luz, 2016).

Although this section has presented several innovative directions on the use of ontologies in archival science, existing work tends to focus on limited aspects, such as classification, retrieval or semantic interoperability of archival records, without fully covering the intrinsic characteristics of archival records, such as fixity, organic nature, naturalness, uniqueness, authenticity and impartiality, in a unified and systematic way. In addition, many of these works are specific to certain domains or systems, which further restricts their applicability in the multifaceted contexts that public administration systems present. Thus, this research seeks to fill an existing gap, with the aim of developing an ontology that encompasses these basic archiving principles and, at the same time, guarantees interoperability and adaptability in a framework between public administration systems.

2.4 Ontological Modeling and Development

In the context of ontological modeling and development, the SABiO (Systematic Approach for Building Ontologies) method, UFO (Unified Foundational Ontology), and OntoUML (Ontologically Well-Founded Profile for UML) are applied in conceptual representation projects, promoting semantic precision and consistency in the creation of ontologies. Further, Visual Paradigm plays an important role in this process, as it is used as the tool for diagramming and implementing these models. The tool allows for the creation of OntoUML models with the aid of a specialized plug-in that ensures compliance with ontological principles, thereby enhancing the modeling experience. Additionally, gUFO (Gentle Unified Foundational Ontology) provides a lightweight implementation of UFO, facilitating the practical application of its core concepts while maintaining compatibility with existing Web Ontology Language[4] based tools and environments. The discussion to follow focuses on the fundamental principles of these approaches, their application in conceptual representation projects, and the integration of these methods and tools.

2.4.1 Systematic Approach for Building Ontologies (SABiO)

SABiO (Systematic Approach for Building Ontologies), introduced by Falbo (2014), is a method for ontology construction, comprising five main phases: (i) Purpose Identification and Requirements Elicitation, (ii) Capture and Formalization, (iii) Design, (iv) Implementation, and (v) Testing. The phases mentioned are described in detail below and, in addition, Figure 2 is presented to illustrate the flow and interrelationship between these phases in the methodological process.

(i) **Purpose Identification and Requirements Elicitation**: This phase involves defining the objective of the ontology and outlining its functional and non-functional requirements, in which the requirements are derived from competency questions, which the ontology must be designed to answer (Falbo, 2014). The competency questions guide the development process, ensuring that the ontology meets the specific needs of the domain (Falbo, 2014).

(ii) **Capture and Formalization**: In this phase, the main concepts and relationships within the domain are identified and formalized, usually using a foundational ontology, such as the Unified Foundational Ontology (UFO), to ensure semantic precision (Falbo, 2014). In addition to a foundational ontology, a modeling language such as OntoUML can be used, providing a graphical representation of these con-

[4] Web Ontology Language (OWL) is a language developed by the W3C to represent and share knowledge on the Semantic Web. Designed to describe ontologies, OWL allows classes, properties, relationships and restrictions to be defined, providing a logical formalism for modeling knowledge domains (Horrocks; Patel-Schneider; Harmelen, 2003).

cepts and relationships to ensure that the ontology accurately captures the domain knowledge (Falbo, 2014).

(iii) **Design**: The design phase focuses on organizing the ontology's architecture, in which technological factors and non-functional aspects such as scalability and performance are considered (Falbo, 2014). In this phase, the ontology's modular structure is also reviewed to ensure that it meets implementation standards and supports reuse and integration (Falbo, 2014).

(iv) **Implementation**: This phase consists of translating the designed ontology into an operational language, such as the Web Ontology Language (OWL) (Falbo, 2014). This consists of implementing the ontology in a format that is compatible with computer systems, allowing it to be used effectively in software applications (Falbo, 2014).

(v) **Testing**: In the final phase, testing is conducted to ensure that the ontology is both structurally sound and aligned with the original requirements (Falbo, 2014). This phase includes integration tests, which verify the correct interaction between sub-ontologies, as well as verification and validation tests (Falbo, 2014).

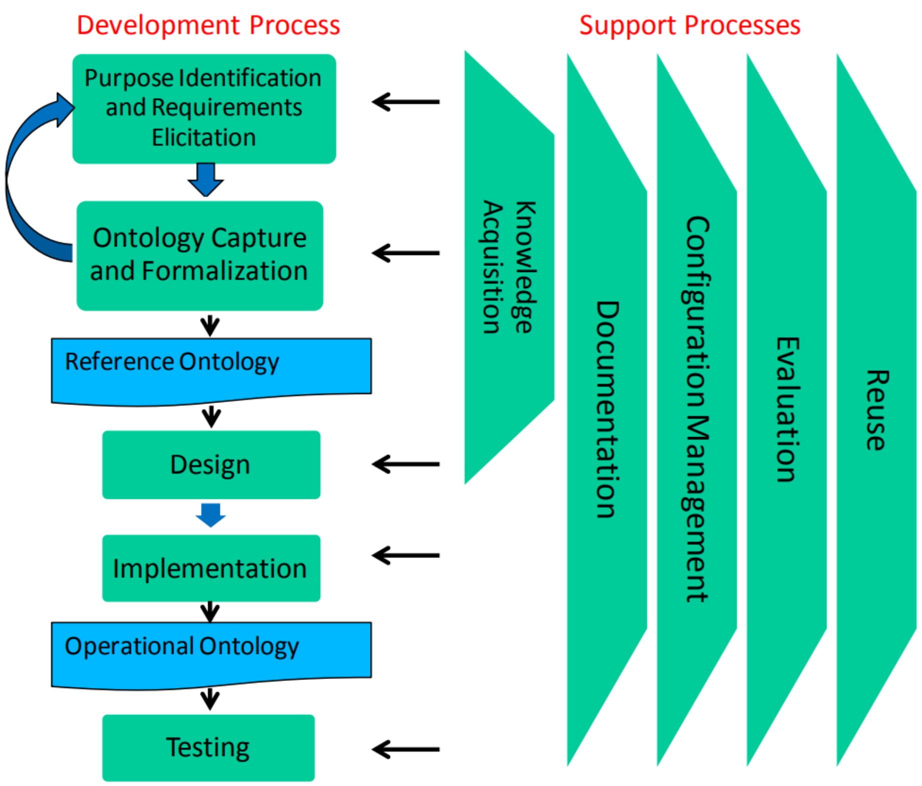

Figure 2 – Overview of the SABiO 2.0 development process, illustrating its five main phases (Falbo, 2014).

The objective of the phases in this approach is to guarantee that the created ontologies are exact, coherent, and faithfully represent the entities and their relationships within the domain. The SABiO methodology distinguishes between reference ontologies and operational ontologies. Reference ontologies offer a solution-independent description, whereas operational ontologies concentrate on computational execution. Moreover, SABiO underscores the utilization of foundational ontologies, such as UFO, in the ontology capture phase, guaranteeing that concepts and relationships are analyzed with ontological precision. This establishes an accurate basis for modeling and mitigates ambiguities that can jeopardize the practical implementation of the ontology (Falbo, 2014).

2.4.2 Unified Foundational Ontology (UFO)

A particular approach to the precise and consistent conceptual representation in information systems is based on the development of foundational ontologies (Guizzardi, 2005). Among those sets of ontologies, the Unified Foundational Ontology (UFO) was defined according to strong theoretical principles rooted in philosophy, formal logic, cognitive psychology, and linguistics (Guizzardi; Falbo; Guizzardi, 2008; Guizzardi *et al.*, 2021). UFO provides with a reference ontology for the analysis and (re)engineering of conceptual models, enforcing semantic interoperability and ontological coherence across domains (Guizzardi, 2005). UFO is divided into three interrelated layers: **UFO-A** (An Ontology of Endurants), **UFO-B** (An Ontology of Perdurants), and **UFO-C** (An Ontology of Social Entities), which provide a comprehensive view of the entities and processes that compose reality, categorizing them according to their fundamental ontological characteristics (Guizzardi, 2005; Guizzardi; Falbo; Guizzardi, 2008; Guizzardi *et al.*, 2021). Figure 3 presents the UFO taxonomy, which systematically organizes and classifies the ontology's concepts into a hierarchical structure, facilitating a clearer understanding of their relationships and interdependencies.

The **UFO-A** layer makes a fundamental distinction between the categories of *Particular (Individual)* and *Universals (Type)*, as can be seen in Figure 4, which shows a fragment of UFO-A. *Particulars* are entities that exist in reality and possess a unique identity, while *Universals* are patterns of properties that can be realized in different *Particulars* (Guizzardi; Falbo; Guizzardi, 2008). *Substances* are examples of existentially independent *Particulars*, such as people or physical objects, whereas *Moments* represent individualized properties that depend on other individuals for their existence, such as color or electrical charge (Guizzardi; Falbo; Guizzardi, 2008). These *Moments* can be classified into *Intrinsic Moments*, which depend on a single individual, or *Relational Moments*, which depend on a plurality of individuals, such as a marriage (Guizzardi *et al.*, 2021). The ontology also explores the relationship between *Intrinsic Moments* and their representation in human cognitive structures, using the theory of "quality spaces", which describes how properties, such as color and height, can be modeled in one-dimensional or multidimensional structures, depending on the type of property represented (Guizzardi *et*

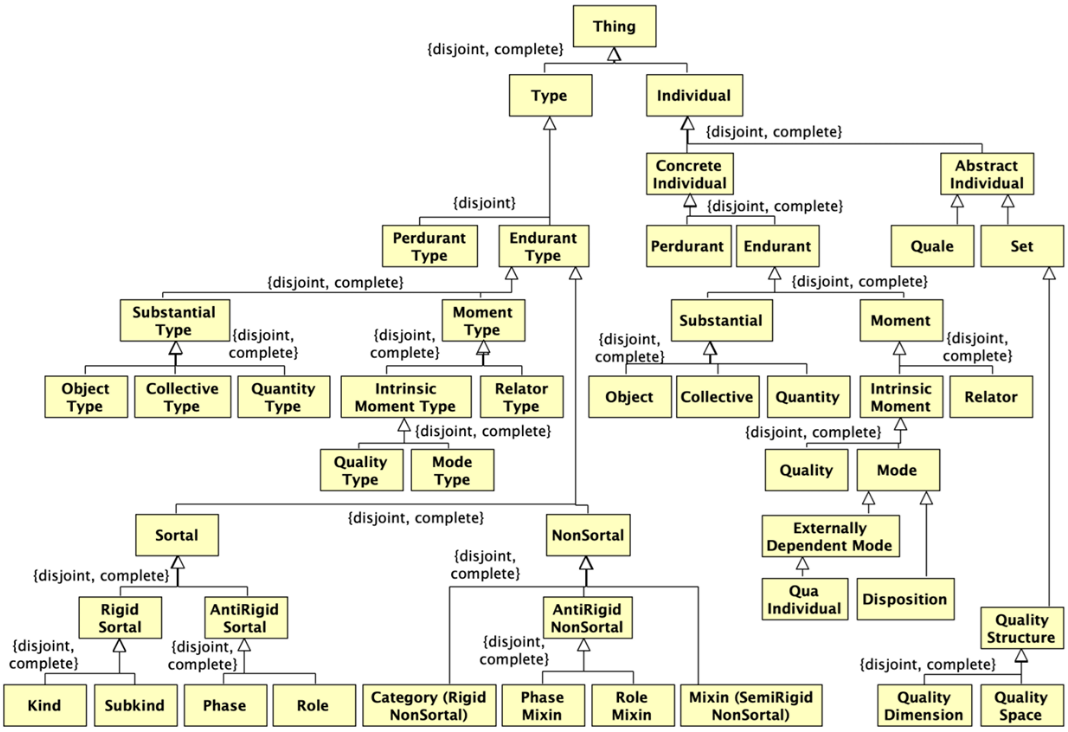

Figure 3 – The Taxonomy of UFO (Guizzardi *et al.*, 2021).

al., 2015; Guizzardi *et al.*, 2021).

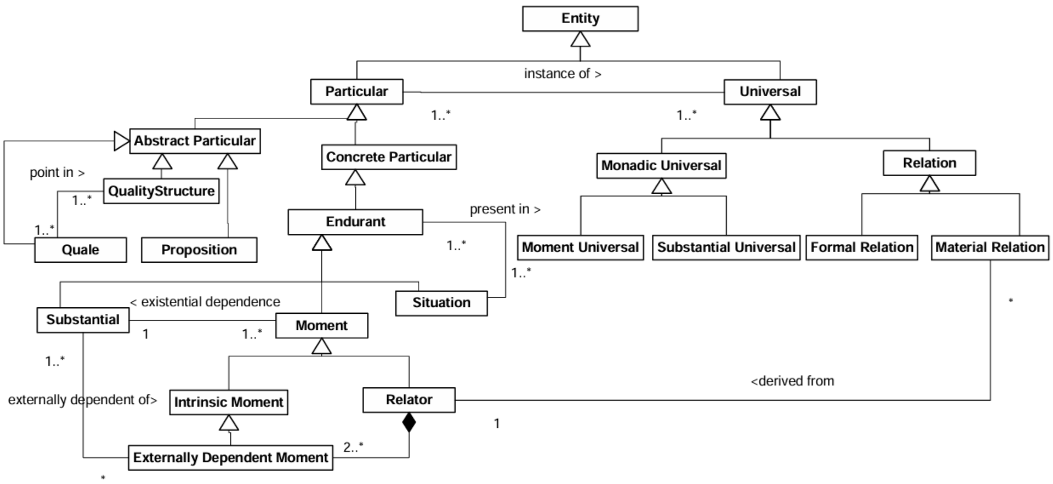

Figure 4 – A Fragment of the Foundational Ontology of Endurants (UFO-A) (Guizzardi; Falbo; Guizzardi, 2008).

Additionally, UFO-A describes two categories of relations: formal relations, which connect entities directly, and material relations, which require a *relator* to mediate the connection between entities (Guizzardi; Falbo; Guizzardi, 2008). For example, a material relation such as medical treatment connects a patient to a medical unit, mediated by a relator (Guizzardi; Falbo; Guizzardi, 2008). These *Relators* play a fundamental role in mediating relations and are composed of *Externally Dependent Moments*, which depend on other individuals for their existence (Guizzardi; Falbo; Guizzardi, 2008). To conclude, UFO-A introduces the concept of *Situations*, which are complex entities formed by various *endurants*, understood as portions of reality that can be analyzed as a whole. *Situations* are viewed as analogous to what is known in the literature as a *state of affairs*, representing combinations of different *endurants* and *moments* (Guizzardi; Falbo; Guizzardi, 2008).

The **UFO-B** layer defines the separations between *endurants* and *perdurants*, in other words, it speaks about the time behaviour (function) of these two. *Endurants* are entities that exist in time and so the whole of the *endurant* will be "present" or existing "all at once", thereby maintaining its identity throughout different situations (Guizzardi; Falbo; Guizzardi, 2008). In contrast, *Perdurants* are said to have the feature that "they persist in time" and they have temporal constituents with non-preserved identity over time (Guizzardi; Falbo; Guizzardi, 2008). Because these individuals exist and persist in time, at any specific moment in time only some of their temporal parts are instantiated, as illustrated by a conversation, a football game or a business process (Guizzardi; Falbo; Guizzardi, 2008).

The main ontological category in UFO-B is the *Event* (*Perdurant, Occurrent*), which can be classified as *Atomic* or *Complex*, depending on its mereological structure (Guizzardi; Falbo; Guizzardi, 2008). While *Atomic* events have no proper parts, *Complex* events are formed by the aggregation of two or more events, which can themselves be *Atomic* or *Complex* (Guizzardi; Falbo; Guizzardi, 2008). These events represent transformations in reality, changing "states of affairs" from a previous situation to a subsequent one. Moreover, events are ontologically dependent on their participants to exist (Guizzardi; Falbo; Guizzardi, 2008). A classic example is the assassination of Caesar by Brutus, where the event includes the participation of Caesar, Brutus, and the knife used in the act (Guizzardi; Falbo; Guizzardi, 2008; Almeida; Falbo; Guizzardi, 2019). Each of these participations constitutes events that, in turn, can be simple or complex but are dependent on the existence of their participating entities (Guizzardi; Falbo; Guizzardi, 2008; Almeida; Falbo; Guizzardi, 2019).

According to Almeida, Falbo & Guizzardi (2019), UFO-B is composed of five sub-theories:

(i) *mereology* – events can form partonomies. In this sub-theory, the relations between events and their parts are characterized by the axioms of the so-called *extensional mereology* (Almeida; Falbo; Guizzardi, 2019);

(ii) *participation* – events are existentially dependent on endurants. The maximal part of an event that is exclusively dependent on a particular endurant is called a participation (of that endurant in that event) (Almeida; Falbo; Guizzardi, 2019);

(iii) *temporal relations* – events occur in time, accumulating temporal parts (Almeida; Falbo; Guizzardi, 2019);

(iv) *events as manifestations of dispositions* – this sub-theory connects endurants and events by characterizing how events are *manifestations* of particular endurants called *dispositions*, which can themselves inhere in other *endurants* (Almeida; Falbo; Guizzardi, 2019);

(v) *change* – events map the world from situations (that *activate* the dispositions of which they are manifestations) to situations (which are *brought about* by the occurrence of that event) (Almeida; Falbo; Guizzardi, 2019).

This ontology also explores the nature of events as entities ontologically dependent on *endurants* and as manifestations of the dispositions of these *endurants* (Almeida; Falbo; Guizzardi, 2019). Furthermore, events can be analyzed as responsible for the transformation of situations, where an event leads to the activation of dispositions that manifest in other events, characterizing a causal relationship between them (Almeida; Falbo; Guizzardi, 2019). Additionally, UFO-B explores the temporal characteristics of events, employing a quality structure grounded in *Time Intervals* and *Time Points* (Guizzardi; Falbo; Guizzardi, 2008). This approach may accommodate diverse temporal structures, including linear, branching, parallel, or circular time, without committing ontologically to any particular representation (Guizzardi; Falbo; Guizzardi, 2008). Consequently, UFO-B offers considerable adaptability in representing events and their temporal relationships, allowing various kinds of temporal ordering to be evaluated based on the applicable context (Guizzardi; Falbo; Guizzardi, 2008). Figure 5 provides a summary of UFO-B, illustrating its key concepts providing a comprehensive view of how the ontology models temporal entities.

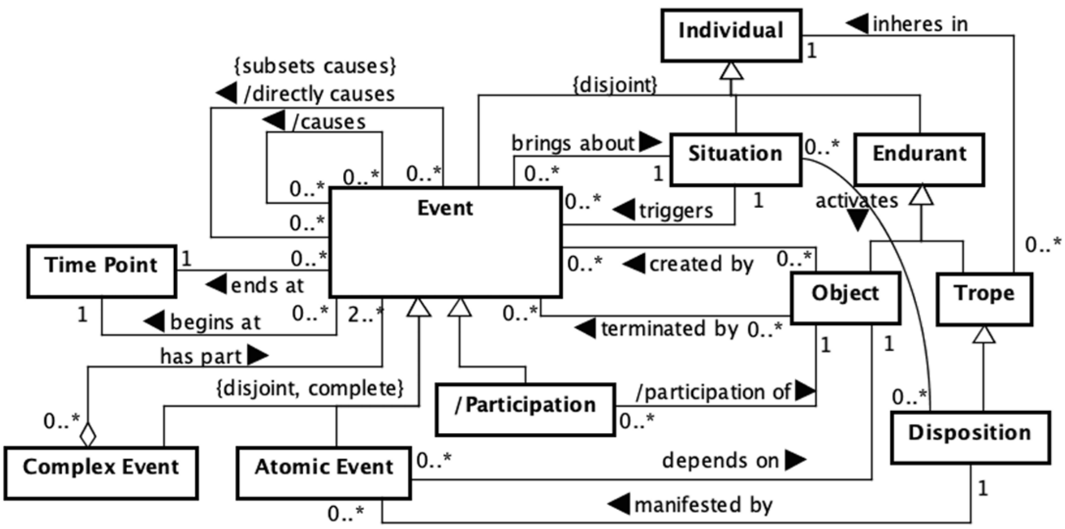

Figure 5 – A summary of UFO-B and its relations (Almeida; Falbo; Guizzardi, 2019).

The ontological representation of the archival record, proposed here, explores the resources offered by UFO-B, especially concerning the modeling of events as *Perdurants* and its ability to represent temporal changes and state transformations. Since UFO-B distinguishes between complex and simple events, which are ontologically dependent on *Endurants* such as agents or objects involved, it provides a solid foundation for capturing the interactions and processes involved in the creation, use, and archiving of records. Moreover, UFO-B provides elements of flexible temporal structures that are *Time Intervals* and *Time Points*, which make it possible to represent phases and states of a record along the way, describing not only its existence through time but also its evolution. Therefore, the usage of UFO-B to represent archival records results in a detailed and comprehensive ontological modeling that ensures temporal coverage of events as well as their causal dependencies over the life cycle of the record.

The **UFO-C** layer is an ontology dedicated to modeling social entities and introduces the distinction between Agentive and Non-agentive substantial particulars, termed *Agents* and *Objects* (Guizzardi; Falbo; Guizzardi, 2008). Agents, such as individuals, can be physical or social, such as organizations or societies (Guizzardi; Falbo; Guizzardi, 2008). Objects, however, can be physical, like books or trees, or social, like money and *Normative Descriptions*. A normative description defines one or more rules recognized by at least one social agent. It can define nominal universals, such as social commitments or social roles, such as president or PhD candidate (Guizzardi; Falbo; Guizzardi, 2008). A fragment of this ontology is shown in Figure 6.

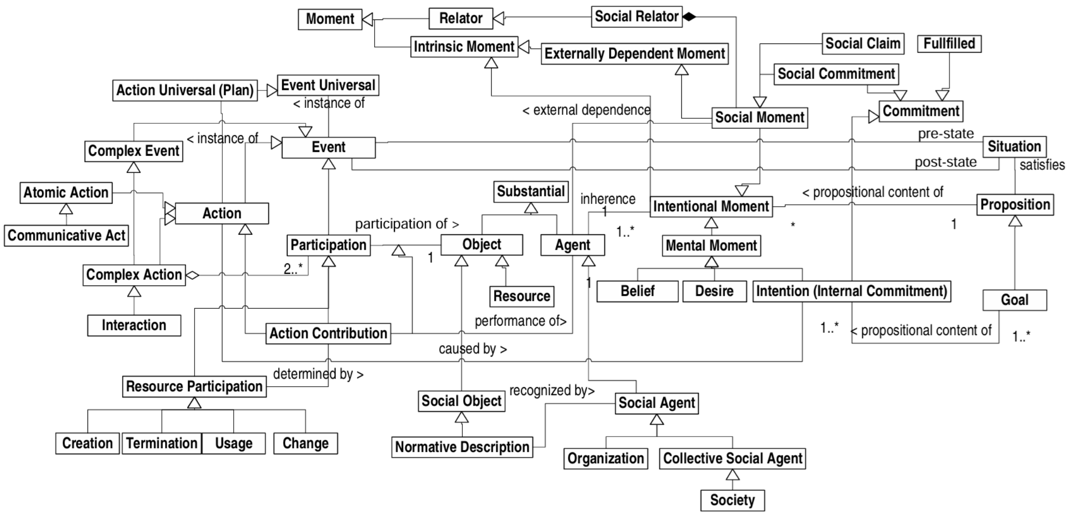

Figure 6 – A Fragment of the Foundational Ontology of Social Entities (UFO-C) (Guizzardi; Falbo; Guizzardi, 2008).

UFO-C also presents the concept of *Intentional Moments*, which are properties carried by agents and have the capacity to refer to possible situations in reality (Guizzardi; Falbo; Guizzardi, 2008). A type, such as *Belief*, *Desire*, or *Intention*, characterizes each intentional moment, while the propositional content serves as an abstract representation of the situation to which it refers (Guizzardi; Falbo; Guizzardi, 2008). Real situations can justify beliefs, and they can fulfill or frustrate desires and intentions. An intention, in turn, is the desire of the agent to commit to performing an *Action* (Guizzardi; Falbo; Guizzardi, 2008).

Actions are intentional events that instantiate a *Plan* (*Action Universal*) to satisfy the propositional content of an intention (Guizzardi; Falbo; Guizzardi, 2008). These actions can be atomic or complex, with complex actions involving multiple intentional or non-intentional participation of other agents or objects (Guizzardi; Falbo; Guizzardi, 2008). The objects participating in actions play the role of resources, which can be used in different ways, such as *Creation*, *Termination*, *Modification*, or *Usage*, depending on their participation in the action (Guizzardi; Falbo; Guizzardi, 2008).

Additionally, UFO-C addresses *Social Commitments* created by communicative acts, such as promises (Guizzardi; Falbo; Guizzardi, 2008). These acts generate *Social Moments* that establish obligations and claims between the agents involved, as exemplified by a rental contract where the renter acquires the commitment to return the car in a specified condition (Guizzardi; Falbo; Guizzardi, 2008). In contrast, the rental company has a claim regarding that commitment, forming pairs with the renter's obligations that refer to specific propositional content (Guizzardi; Falbo; Guizzardi, 2008).

Following the explanation of UFO-A, UFO-B and UFO-C, it can be noted that UFO plays an important role in conceptual modeling, providing a solid theoretical foundation for representing entities, events, and social interactions. Its modular structure allows for accurate modeling of different aspects of reality, and its integration with methods such as SABiO and languages such as OntoUML promotes the creation of semantically rich and consistent ontologies for a wide range of domains (Guizzardi *et al.*, 2021).

Considering the wide range of stereotypes proposed by UFO's micro-theories and their ongoing evolution over time, this work adopts the strategy of introducing each stereotype's concept only at the point where it is actually applied within the ontology. This approach seeks to maintain the coherence and readability of the exposition while avoiding overly anticipatory theoretical discussions. It is important to emphasize that the adopted stereotypes are grounded in authoritative sources according to their original formulations, as many of them are not consolidated in a single reference, but rather disseminated across various specialized academic publications.

2.4.3 OntoUML

The development of conceptual ontologies plays an important role in information systems modeling by enabling the creation of abstract but accurate representations of complex realities. The semantics of concepts are well understood (Guizzardi, 2005), and a representation language that captures these concepts precisely is OntoUML — an ontologically grounded extension of UML.[5] Guizzardi (2005) developed OntoUML to offer instruments for modeling ontology with conceptual precision. Unlike other modeling languages, OntoUML proposes a distinction among ontological categories such as *Kind*, *Role*, and *Phase* for this modeling goal, enabling them to express entities that can undergo state changes over time while retaining their identities (Guizzardi, 2005). This is especially significant in contexts such as archival records management, where records and their associated entities undergo cycles of change over time, but still face challenges and must preserve their integrity and identity throughout these transformations.

In addition, OntoUML introduces **relators** and **mediations** in order to model complex interdependent interaction between several entity types (Falbo, 2017). A relator is used to mediate two different things (a pair of participant relations) so that the relationship based upon them is not simply accidental but ontologically grounded. For example, in events of contracting, the contract is a **relator** that connects parties and defines some of their obligations and responsibilities (Falbo, 2017). Mediations, in contrast, represent how entities are involved in the relationships with the **relator**. This is particularly useful in information systems dealing with contractual or business processes, allowing for the detailed modeling of institutional relations and commitments (Falbo, 2017). Therefore, OntoUML is not only a modeling notation, but also a formalization of UFO's ontological distinctions.

2.4.4 Visual Paradigm

Visual Paradigm serves as a tool for creating a variety of software diagrams, such as class diagrams that adhere to the Unified Modeling Language (UML), an important component of object-oriented modeling (Visual Paradigm, 2024). The academic community employs Visual Paradigm for teaching and research, as the tool is offered via academic licenses that allow students and researchers to utilize the platform for educational initiatives (Visual Paradigm, 2024).

Over the years, OntoUML users recognized the necessity for tools that facilitate the verification of model consistency and ensure accurate representation of intended conceptualizations. Guizzardi *et al.* (2018) developed the OntoUML plug-in for Visual Paradigm in response to this demand, assisting modelers in diagramming models using Unified Foundational Ontology (UFO) stereotypes and enabling automatic detection of inconsistencies. There are numerous benefits associated with using this plug-in, including the following:

- **Support for Ontological Stereotypes**: The plug-in enhances Visual Paradigm by adding ontological stereotypes specific to UFO, aiding in the accurate conceptual

[5] The OntoUML vocabulary is available at: <https://dev.ontouml.org/ontouml-vocabulary/>.

representation of the domain being modeled (Guizzardi *et al.*, 2018);

- **Automatic Consistency Verification**: The plug-in allows modelers to automatically verify whether the model complies with the semantic and syntactic rules defined by UFO, ensuring that the modeling aligns with ontological principles (Guizzardi *et al.*, 2018);
- **Detailed Logs**: In the event of a rule or restriction violation, the plug-in generates detailed logs that highlight the entities involved, facilitating the analysis and correction of models (Guizzardi *et al.*, 2018).

The necessity to provide computational assistance for the development of OntoUML models, ensuring adherence to the ontological distinctions and the semantic and syntactic constraints of UFO, drove the design of the plug-in (Guizzardi *et al.*, 2018). This enhances conceptual precision and allows sharing and application of solid ontological approaches in ontology-driven conceptual modeling (Guizzardi *et al.*, 2018), including the application of OntoUML stereotypes, support for intelligent diagram painting, automatic model consistency checking, and Model Transformation to OWL with gUFO. To work with models using OntoUML features in Visual Paradigm, it is necessary to install the *OntoUML Plugin*.[6]

By using an OntoUML-compliant tool, such as Visual Paradigm with the OntoUML plugin, modelers benefit from built-in validation mechanisms that ensure adherence to the ontological rules defined in UFO's micro-theories. These include constraints on identity, parthood, dependency, and the nature of events and agents. As a result, the diagrams produced maintain ontological integrity, and inconsistencies with UFO's foundational commitments can be detected and corrected during the modeling process.

2.4.5 A Lightweight Implementation of the Unified Foundational Ontology (gUFO)

Developed for Semantic Web[7] applications, particularly in OWL 2 DL applications, *A Lightweight Implementation of the Unified Foundational Ontology* (gUFO) is a simplified version of the Unified Foundational Ontology (UFO) (Almeida *et al.*, 2020). This resource was created with the aim of offering a reusable and lightweight ontology base, allowing ontology developers to specialize and instantiate essential elements of UFO in a more

[6] The *OntoUML Plugin for Visual Paradigm* is available at: <https://github.com/OntoUML/ontouml-vp-plugin>.

[7] This concept came from the need to transform the Web into a space where information is not only accessible to humans, but also intelligible to computer agents, allowing tasks to be automated, data to be integrated and information to be retrieved more accurately and efficiently (Lassila; Hendler; Berners-Lee, 2001).

practical and efficient way (Almeida *et al.*, 2020). As a "lightweight ontology," gUFO retains important computational properties and uses limited expressivity to maintain compatibility with applications that require high performance (Almeida *et al.*, 2020). The "g" in gUFO stands for "gentle", at the same time, "gufo" is the Italian word for "owl" (Almeida *et al.*, 2020).

The structure of gUFO is composed of two main taxonomies:

1. **Taxonomy of Individuals**: This category includes classes whose instances are concrete individuals, such as `gufo:Object` and `gufo:Event`. According to UFO, individuals are classified as either endurants or perdurants (Guizzardi, 2005).

   `gufo:Endurant`, known as *endurants*, are entities that persist in time and can undergo qualitative changes without losing their essential identity (Guizzardi, 2005). Examples of endurants include everyday objects such as a person, a car or a contract. It also includes events and situations that can involve multiple individuals, such as a soccer match or a marriage contract (Guizzardi, 2005; Almeida *et al.*, 2020).

   `gufo:Event`, known as *perdurants*, are entities that unfold in time, meaning that their identity is inherently linked to their temporal extension. Examples include a soccer match, a marriage ceremony, or a business transaction (Guizzardi, 2005; Almeida *et al.*, 2020);

2. **Taxonomy of Types**: This covers classes whose instances are types, such as `gufo:Kind`, `gufo:Phase`, and `gufo:Category`. These types make it possible to classify entities based on well-defined ontological criteria, such as *sortality* and *rigidity*, which allows an ontology to specify how types are applied to their instances (Guizzardi, 2005). This is particularly useful for ensuring consistency and avoiding representation errors, such as assigning multiple rigid categories to a single entity (Guizzardi, 2005; Guizzardi *et al.*, 2018).

In addition, gUFO makes it easier to model the temporal aspects of entities, making it possible to track changes and establish historical dependencies between events (Almeida *et al.*, 2020). For example, it is possible to define that a temporary situation, such as an employment status, has a specific beginning and end, which enriches the temporal representation and helps to capture the evolution of an entity over time (Almeida *et al.*, 2020). This feature is of particular importance when modeling systems where continuity and temporality play central roles, such as in record management and digital preservation systems.

With the OntoUML plugin installed in Visual Paradigm, models created in OntoUML can be transformed into gUFO-based ontologies. This feature enables users to export their conceptual models as OWL ontologies, following the lightweight implementation of the Unified Foundational Ontology. By converting OntoUML models to gUFO, the plugin facilitates semantic interoperability and ensures that the model's ontological distinctions are preserved in a format easily processed by computational systems.

2.5 Summary

Chapter 2 provided a theoretical and methodological basis for this research, addressing the domain of archival records and the application of relevant methods. This chapter explored the fundamental concepts and defining characteristics of archival records, which serve as guidelines for constructing the ontology. It also introduced the CLeAR approach, emphasizing its role in systematically reviewing literature to identify and reuse structured resources relevant to the archival domain. Additionally, the chapter discussed the use of ontologies within archival science, highlighting their importance in enhancing the management, organization, and access to archival records, as well as ensuring interoperability across systems. Finally, the chapter outlined a number of elements utilized in the remainder of this work, including OntoUML, SABiO, UFO, gUFO and the Visual Paradigm tool with its OntoUML plugin.

3 Applying CLeAR to the archival records domain

In this chapter, we report on the application of CLeAR to the domain of archival records, covering Integration Questions Definition, Systematic Search Configuration, Publications Selection, Structured Resources Identification and Snowballing. CLeAR was proposed in the scope of data integration projects, and hence there are also activities concerning Data Sources Selection and corresponding Domain Aspects Identification in Cycle I as discussed in the previous chapter. These are not addressed explicitly here, as our aims are more general; however, the integration questions identified are based on the practices of archival records management observed across various public administration systems. Within this context, systems for managing purchases, ombudsman services, finances, human resources, and others coexist. Hence, there are integration issues that are largely independent of specific organizational or application domains, and the management and preservation of archival records cuts across many systems utilized by the public administration.

3.1 Integration Questions Definition (Cycle I)

The starting point is to define Integration Questions (IQs), which are questions about the domain to be explored and that may cut across its various aspects. This activity is the basis for carrying out the subsequent activities, and requires the participation of domain experts, such as archivists, managers, end users, and product owners. Likewise, it requires firm grounding in the relevant literature. Here, we seek an intersection between the areas of Computer Science and Archival Science.

The characteristics of fixity, organic nature, naturalness, uniqueness, authenticity, and impartiality in the Archival Science's literature are reflected into concrete questions. When we ask "Who is the producer of the record?", "Which activity gave origin to the record?", or "What is the classification scheme assigned to the record?", we refer to the characteristic of organic nature, which relates to the record's connection with the context in which it was created. This broad aspect of "organic nature" can then be translated into a more concrete aspect: "classification scheme". A classification scheme serves as a logical representation of the structure and operations of a public or private organization and should be associated with the producer or creator of the record. These questions can also be associated with aspects of naturalness and authenticity.

An electronic archival record must have a clearly defined storage location to ensure its accessibility. However, identifying where the digital object that materializes it is stored is essential to preserving its chain of custody, a determining factor in verifying its authenticity. The question "Where is the digital object located?" refers to the need to identify the repository, server or system in which it is stored, ensuring not only its retrieval, but also the possibility of verifying its authenticity.

When discussing the characteristic of authenticity, the focus is on elements that can ensure the presumption of authenticity of an archival record. These elements are characterized by the presence of answers to questions such as: "Who is the producer of the record?", "What is the date and place of production of the record?", "Is the record subject to signature? What type of signature? What is the method of signature? Who signs it? When was it signed?", "Which administrative body or authority is the record addressed to?", "What is the classification scheme assigned to the record?", "Who are the parties concerned?". Such responses aim to enhance the reliability of the record and, consequently, increase the trustworthiness. However, the answers to these questions are not only limited to the feature of authenticity, they can also contribute to characterizing the organic nature and naturalness. When these questions are answered, we can obtain important information that helps to characterize the naturalness of the archival record, revealing its connection to the events, processes, and people involved in its creation and use.

Fixity, understood as the guarantee that the record remains unaltered since its creation, is a fundamental characteristic of the archival record, regardless of the medium in which it is held, whether physical or electronic. However, in the context of electronic records, fixity has taken on new challenges. To ensure not only fixity but also aspects such as the authenticity of the archival record, it is essential that certain questions are addressed during the management and preservation of the archival records process. Questions such as "What is the title of the record?", "What is the file format of the record?", "Who are the parties concerned?", "What is the administrative motivation for producing the record?" and "What is the method of signature?" have a role in leaving traces that allow identifying whether the information recorded in the archival record has been manipulated or edited.

The implementation of technological resources can have an important role in ensuring the impartiality of archival records in electronic format. Answers to questions such as: "What is the administrative motivation for producing the record?" and "Who are the parties concerned?", can help to understand the purpose or motive behind the creation of the record and help to assess its objectivity, as well as reveal possible affiliations or interests.

An archival record requires a unique identifier to facilitate its identification and retrieval, however, it does not mean that there will be no other physical or digital copies. When asking "What is the unique identifier of the record?", reference is made to uniqueness, in which the assignment of a code, number, or exclusive identifier to that specific archival record allows it to be linked to other information systems, facilitating its reference in different contexts.

Table 1 presents a list of integration questions (IQs). As can be seen, these questions are related to the context of the production of the archival record, access restrictions and conditions, concerned parties, and others. The questions listed in Table 1 are not intended to cover all possible ways of supporting the aspects of fixity, organicity, naturalness, uniqueness, authenticity, and impartiality of archival records. However, when using the CLeAR approach, there is the possibility for continuous refinement of the integration questions throughout its application. This can be accomplished through the ongoing process of adding, grouping, and updating the integrations questions, allowing for a dynamic and adaptive approach.

Table 1 – Integration Questions Definition

Identifier	**Integration Question**
IQ01	Who is the producer of the record?
IQ02	What is the date and place of production of the record?
IQ03	What is the administrative motivation for producing the record?
IQ04	What is the unique identifier of the record?
IQ05	What is the title of the record?
IQ06	Who are the parties concerned?
IQ07	Is the record subject to signature? What type of signature? What is the method of signature? Who signs it? When was it signed?
IQ08	Does the content of the record have access restrictions?
IQ09	What is the file format of the record?
IQ10	Does the record have any special access conditions?
IQ11	What is the classification scheme assigned to the record?
IQ12	Which administrative body or authority is the record addressed to?
IQ13	Where is the digital object located?
IQ14	What activity gave origin to the record?

3.2 Systematic Search Configuration (Cycle II)

Once the activities of Cycle I have been concluded, we move on to the activities of Cycle II, inspired by Systematic Literature Review (SLR) practices. However, while SLR focuses on systematically reviewing existing studies, CLeAR investigates the scientific literature and technical papers to find structured resources available in the domain of interest. In the Systematic Search Configuration activity, the following steps were performed: (i) describe the search objectives, (ii) select the words that will compose the search string, (iii) build the search string, (iv) select the repositories to be searched, (v) define selection criteria, (vi) list the selected publications, and (vii) identify the structured resources. A detailed description of the results of these activities is available as supplementary material.[1]

The first activity of Cycle II is to define the goal of the research, which is to find candidate structured resources to be reused in the development of ontologies with the purpose of semantic interoperability. Moving on, the selection of keywords reflects the dual nature of the search goals: there are keywords related to structured resources and keywords related to the research domain. To compose the keywords related to structured resources, the selected terms were: (i) *ontology* and correlated terms *reference model*, *knowledge base*, *schema* and (ii) *vocabulary* and correlated terms *taxonomy*, *thesaurus*. To complete the second part, the keywords related to the research domain, the following terms were selected: (iii) *archival studies* and correlated term *Archival Science*, (iv) *record* and correlated terms *Archival Document*, *Archival Record*, *Digital preservation*, *Digital Record*, *Document Management*, *Electronic Record*, *Record Keeping*, *Record Management* and (v) *Archive* and correlated terms *Archival*, *Archives*.

Based on the IQs, search strings were built to reflect the domain to be searched and also the types of structured resources to be found, targeting search engines Scopus Elsevier and Web of Science. To construct the search strings, the fields title, abstract, and author keywords were considered. From the keywords, a first version of the search string was built, which was refined after selecting the electronic repositories Scopus and Web of Science. These repositories provide access to a large volume of scientific publications, abstracts, citations, and also provide resources for bibliometric indicators.

Initially, two search strings were built, one for Scopus and another for Web of Science. However, during the process of testing and refining the search string, it was identified that the TITLE-ABS-KEY-AUTH label as a search field only exists in the Scopus repository, while for the Web of Science repository it was necessary to use three different search fields: Title (TI), Author's keywords (AK), and Abstract (AB). Another difference found was in the construction of the negation structure between the repositories. The Scopus repository allows the use of negation in the TITLE-ABS-KEY-AUTH search field, while the Web of Science search repository only allows negation in the title search field. Due to this limitation and to ensure greater clarity in our search strategy, it was decided to create three separate queries tailored to the Web of Science repository.

To help in the process of filtering the results, it was decided to insert a negation feature in the search string to operationalize exclusion criteria and thus reduce upfront the number of publications with terms related to health, decision support, learning, etc. To simplify extraction from the Web of Science repository, no language restriction was inserted into the search string, but it was necessary to create an inclusion criteria to filter the publications extracted from this repository. For the specific search string in the Scopus

[1] Supplementary material available at: <https://zenodo.org/records/8335807>

Table 2 – Strings used in search engines

Identifier	Search string	Repository	Number of publications returned
S01	TITLE-ABS-KEY-AUTH ((ontolog* OR vocabular* OR schema? OR taxonom* OR thesaur* OR "reference model" OR "knowledge base") AND ("Archival Stud*" OR "Archival Science" OR "Electronic Record?" OR "Record Management" OR "Document Management" OR "Archival document" OR "Digital preservation" OR "record keeping" OR "Digital record" OR "Record? and archive?" OR "Archival record?") AND NOT (diabetes OR *health* OR ehr OR *medical* OR "Sentiment analysis" OR *learning)) AND LANGUAGE ("English") AND SUBJAREA (soci OR comp)	Scopus Elsevier	364
S02	TI=((ontolog* OR vocabular* OR schema? OR taxonom* OR thesaur* OR "reference model" OR "knowledge base") AND ("Archival Stud*" OR "Archival Science" OR "Electronic Record?" OR "Record Management" OR "Document Management" OR "Archival document" OR "Digital preservation" OR "record keeping" OR "Digital record" OR "Record? and archive?" OR "Archival record?")) NOT TI=((diabetes OR *health* OR ehr OR *medical* OR "Sentiment analysis" OR *learning))	Web of Science	10
S03	AB=((ontolog* OR vocabular* OR schema? OR taxonom* OR thesaur* OR "reference model" OR "knowledge base") AND ("Archival Stud*" OR "Archival Science" OR "Electronic Record?" OR "Record Management" OR "Document Management" OR "Archival document" OR "Digital preservation" OR "record keeping" OR "Digital record" OR "Record? and archive?" OR "Archival record?"))	Web of Science	136
S04	AK=((ontolog* OR vocabular* OR schema? OR taxonom* OR thesaur* OR "reference model" OR "knowledge base") AND ("Archival Stud*" OR "Archival Science" OR "Electronic Record?" OR "Record Management" OR "Document Management" OR "Archival document" OR "Digital preservation" OR "record keeping" OR "Digital record" OR "Record? and archive?" OR "Archival record?"))	Web of Science	23

repository, a further restriction by language and by subject area was included, restricting the search for articles in Information Science and Computer Science. This was motivated by the broad scope and interdisciplinary nature of the publications found, thus making necessary to initiate the exclusion process during the search. To achieve the final version of the search string, wildcard and lemmatization features were used in addition to the search fields available in the electronic repositories. In the end, we obtained four search strings that can be seen in Table 2.

3.3 Publications Selection (Cycle II)

After conducting the systematic search, we initially obtained 533 publications. We first applied the following *exclusion criteria*:

- PEC01: The publication is not available; which in this case means that we were unable to retrieve the full text using the Brazilian Portal of Journals by the Coordination for the Improvement of Higher Education Personnel (CAPES);
- PEC02: The terms health, medicine, decision support, and natural language were added to the exclusion criteria to limit the scope of the mapping. We exclude articles dealing specifically with record management in health care (electronic health record or EHR, medical, diabetes), record use for decision support in business processes, and the use of natural language for record analysis.

After that, the following *inclusion criteria* were applied:

- PIC01: The publication is written in English;
- PIC02: The publication presents or mentions structured resources about the semantic interoperability using archival records domain or their aspects.

In the first stage, after joining the results from Scopus Elsevier and Web of Science, it was possible to identify 132 (24.8%) duplicated publications, since many studies are available from more than one source. In the second stage, the remaining 401 (75.2%) publications, were filtered using the PEC01 exclusion criteria, resulting in 253 publications. In a third stage, the PEC02 exclusion criterion was applied, which reduced the set of publications to 216. Although the PEC02 criterion is part of the S01 and S02 strings, it could not be used in strings S03 and S04 in the Web of Science electronic database, because this search engine does not allow negation structure for advanced search field labels AK and AB. In the fourth and last stages, the publications were refined by applying the PIC01 and PIC02 criteria, where each publication was analyzed to identify whether it presented or mentioned structured resources on semantic interoperability using the archival record domain or its aspects. Of these, only 39 presented the structured resources used. Table 3 summarizes this process.

Table 3 – Results of the publications selection process stages

Stage	**Criteria**	**Initial number of publications**	**Number of publications after stage**	**Reduction per stage (%)**
1st Stage	Eliminating duplications	533	401	24.8%
2nd Stage	PEC01	401	253	36.9%
3rd Stage	PEC02	253	216	14.6%
4th Stage	PIC01 and PIC02	216	39	81.9%

3.4 Snowballing (Cycle II)

Snowballing consists of exploring references (backward snowballing) and subsequent citations (forward snowballing) from an initial set of relevant publications, with the aim of identifying additional works that may have been missed by automated search strategies (Wohlin, 2014). As applied in the CLeAR approach, this technique complements the systematic search and strengthens the coverage of the review, even when its quantitative results are limited.

In this study, the technique was applied to the lists of references and citations of the 39 publications selected in the Selected Publications activity (Cycle II), resulting in the retrieval of 1,417 publications. These publications were filtered to eliminate publications whose records were incomplete (e.g., no title, no author, no abstract, no keywords), and also the duplicates, resulting in 503 publications. After applying the criteria PEC01, PEC02, PIC01 and PIC02, six new publications were obtained, however 5 publications had already been identified in the publication selection activity (Cycle II). Thus, only one new publication remained to be added to the final set of the publications selected, resulting in an increase in the total number of selected publications from 39 to 40. For improved organization and understanding, the new publication from snowballing were listed in the new tabs “Reference List Selection” and “Citations Selection” and they can be obtained by accessing the supplementary material.

3.5 Structured Resources Identification (Cycle II)

After performing the publication selection and Snowballing activities (Cycle II), the remaining 40 publications were analyzed to identify the structured resources used by the authors. Initially, 42 structured resources were identified in the publications. These 42 structured resources were analyzed and selected by applying the exclusion criterion **SREC01**: The structured resource is not available; and the inclusion criterion **SRIC01**: The structured resource addresses semantic interoperability using the archival records domain or its aspects. After applying these criteria, 19 structured resources were obtained. The complete list of structured resources identified can be viewed under the “Structured Resource Identification” tab in the supplementary material. A list of the selected structured resources can be found under the “Structured Resources Selection” tab in the supplementary material. In Cycle III, the remaining structured resources are analyzed, classified, and evaluated, resulting in a set of structured resources relevant to the research domain.

3.6 Data Synthesis and Results

This section presents the synthesis and findings obtained after applying the CLeAR approach to the archival records domain. The results obtained provide a comprehensive overview of the current landscape of structured resources in the domain of archival records, allowing the identification of important structured resources for improvement and reuse.

3.6.1 Systematic Search Reporting (Cycle II)

To give an overview of the results, Figure 7 shows the distribution of the 40 selected publications over the years. We have found publications covering the period from 2002 to 2021. One factor that could be considered as a driver for the growth of research in archival records over this period was the emergence of new legislation and compliance regulations introduced by Sarbanes-Oxley, Basel II, and the International Organization for Standardization (ISO) 15489 Records Management published in 2001. ISO 15489 is the first global standard for records management.

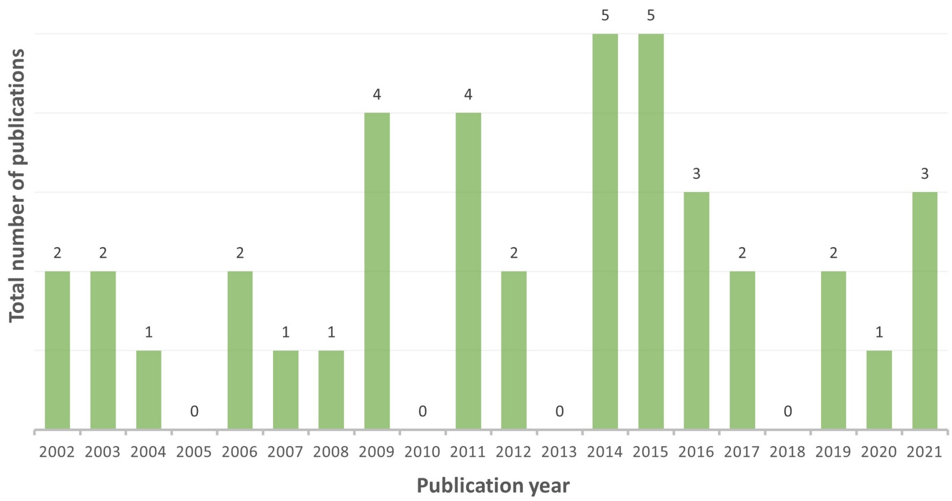

Figure 7 – Final distribution of publications per year.

Table 4 shows the complete list of the 19 structured resources obtained after Cycle II. The vast majority of these resources (85%) are framed as *ontologies*, with only a few framed as *conceptual reference model* (notably CIDOC-CRM), *vocabularies* and/or *data model*.

The identified structured resources use a variety of formalisms and techniques for their representation or even a combination of them. As we can see in Figure 8, the Web Ontology Language (OWL), developed by the World Wide Web Consortium (W3C), is used by 68% of the selected structured resources. If we add up the other arrangements that use OWL, we can summarize that this language has become the “lingua franca” (Lohmann, Steffen and Haag, Florian and Negru, Stefan, 2016) for ontologies. Besides OWL, there are

Table 4 – List of selected structured resources

Identifier	Structured Resource Name	Structured Resource Type	Language	Grouped by usage type
SR01	CIDOC-CRM	Conceptual Reference Model	RDF, OWL	Cultural heritage
SR02	Archive Dynamics Ontology (ArDO)	Ontology	RDF	Cultural heritage
SR03	Bibliographic Ontology (BibO)	Ontology	RDF, OWL	Bibliographic reference
SR04	FRBR-aligned Bibliographic Ontology (FaBiO)	Ontology	OWL	Bibliographic reference
SR05	Records in Contexts-Ontology (RiC-O)	Ontology	OWL	Archival description
SR06	Cultural Event Ontology (ArCo network)	Vocabulary	RDF, OWL	Cultural heritage
SR07	ARKIVO ontology	Ontology	OWL	Archival description
SR08	Chinese Government Archive Ontology (GAO)	Ontology	OWL	Semantic interoperability
SR15	PERICLES Linked Resource Model (LRM)	Ontology	OWL	Digital preservation
SR20	Born-Digital Archives (BDA)	Ontology	OWL	Digital preservation
SR21	Software-Based Art (SBA)	Ontology	OWL	Digital preservation
SR22	TIMBUS context model	Ontology	OWL	Digital preservation
SR26	PREMIS OWL Ontology	Ontology	OWL	Digital preservation
SR28	PROV Ontology (PROV-O)	Ontology	OWL	Provenance
SR29	OPM OWL Ontology (OPMO)	Ontology	OWL	Provenance
SR30	Provenance, Authoring and Versioning Ontology (PAV)	Ontology	OWL	Provenance
SR31	BBC Provenance Ontology (bbcprov)	Ontology	RDF	Provenance
SR34	Open Archives Initiative Object Reuse and Exchange (OAI-ORE)	Data model and vocabulary	RDF/XML, RDFa, JSON-LD	Description of Web resources
SR41	Historical Event Representation Ontology (HERO)	Ontology	OWL	Cultural heritage

structured resources expressed directly in the underlying Resource Description Framework (RDF), or in other RDF serialization variants.

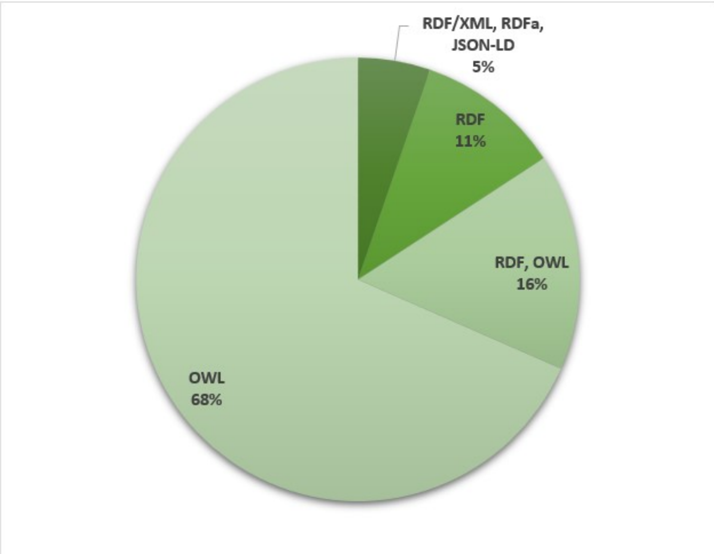

Figure 8 – Distribution of structured resources per language used for development or serialization.

The identified structured resources can also be classified according to the application domain as shown in Figure 9. The first most common application domain is Digital Preservation (26%). In this domain, authors emphasize the need for preservation to ensure the presumption of authenticity of archival records by proposing a model for authenticity in digital archives (Tennis; Rogers, 2012). Another emphasized element is the requirement for provenance registration as an important issue in digital preservation (Li; Sugimoto, 2014).

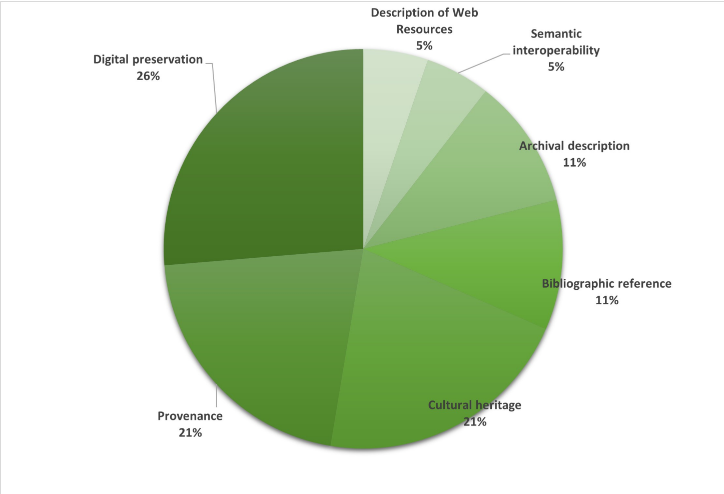

Figure 9 – Distribution of selected structured resources per application domain.

The second most common application domains in the identified structured resources are Provenance (21%) and Cultural Heritage (21%), occurring with the same frequency. Provenance is considered one of the most relevant concepts in archives (Lagos; Vion-Dury, 2016). Li & Sugimoto (2014) and Lagos, Waddington & Vion-Dury (2015) argue that without proper provenance registration, digital preservation may be impacted. Cultural heritage aims to preventing the extinction of the way of making artifacts that are part of a society (Isa *et al.*, 2019), protecting historical archives (Goy *et al.*, 2020), preserving collections that concern society and economy nowadays (Marketakis; Tzitzikas, 2009), recording, organizing and preserving aspects of a specific cultural heritage (Pramartha; Davis; Kuan, 2017), preserving a specific area of science (Cortese; Mantegari, 2011), record keeping of historical artifacts (Matousek; Kouba, 2002), etc. In this domain, the authors demonstrate the need for efforts to be made to protect cultural heritage to support historical memory, cultural identity, understanding of social change, and the ability to manage it (Goy *et al.*, 2020).

Archival description and Bibliographic reference have appeared as application domains of 11% of the resources. Archival description is considered as one of the seven functions of Archival Science (Rousseau; Couture, 2004). The use of structured resources in the domain of bibliographic references by the authors (Vsesviatska *et al.*, 2021) indicates the use of the BibO ontology for assisting in the construction of digital object identifiers, uniform resource identifier, and uniform resource location (Mandal, 2021).

Finally, Semantic Interoperability and the Description of Web Resources both amount to 5% of the remaining resources. Despite the small percentage of these application domains, it is important to understand that they address issues that have recently become part of the electronic record management and web archiving agenda.

3.6.2 Structured Resources Analysis (Cycle III)

In CLeAR, domain coverage analysis is performed by constructing a matrix, in which each row refers to a structured resource and each column refers to an aspect of the domain under analysis. To this matrix are added the total aspects of the domain covered, then the coverage percentage for each structured resource is calculated. Table 5 presents the result of the domain coverage analysis. The symbols (+) and (±), used in Table 5, indicate coverage of a domain aspect by a structured resource. The symbol (+) denotes full coverage, and the symbol (±) denotes some level of coverage. In any case, a value of 1 is assigned for any level of coverage in order to compute a total coverage score for each resource.

The domain aspects used Table 5 were directly derived from the integration questions defined in Cycle I (Section 3.1). Each question expresses a specific conceptual representation need related to the production and management of archival records — such as authorship, administrative motivation, file format, among others. From the analysis of these questions, the main domain aspects were abstracted and then used as criteria to organize the evaluation of the structured resources identified during the systematic search. This categorization was essential for the construction of Table 5, which presents a domain coverage matrix, allowing for a more structured comparison of the potential contributions of each structured resource. This strategy is aligned with the methodological proposal of the CLeAR approach, as described by Campos (2019), Campos *et al.* (2020), and aims to ensure clarity and coherence in the analysis of resources available within the studied domain.

As part of the application of the CLeAR approach (Campos, 2019; Campos *et al.*, 2020), this study adopts the 5W1H framework — an acronym for Who, What, When, Where, Why, and How — to support the conceptual organization of the domain aspects derived from the integration questions. This classical technique for structuring information has its origins in rhetoric and investigative journalism and was later systematized and widely applied in management, engineering, and quality control contexts, particularly in industrial problem-solving (Ohno, 1988). Although the 5W1H structure is incorporated into the CLeAR methodology, it is important to clarify that this framework was not originally developed by it. In this work, the 5W1H serves as the basis for organizing the domain aspects in Table 5, grouping them into six guiding dimensions to support the comparative analysis of the structured resources.

The structured resource CIDOC-CRM (SR01) has 100% coverage in relation to aspects of the domain and is also one of the resources that has been in continuous development for the longest time. The first version of the CIDOC-CRM (SR01) was released in 2004; in 2006 it became an ISO 21127:2006 standard. The current version is the 2014 version, which is under review and will be replaced by ISO/PRF 21127.

Table 5 – Structured resources domain coverage matrix.

		Aspects Group 1 - What							Aspects Group 2 - Who					Aspects Group 3 - Where			Aspects Group 4 - When			Aspects Group 5 - Why		Aspects Group 6 - How		Number of Covered Aspects	
Identifier	**Structured Resource**	**Record Identifier**	**Record title**	**File format**	**Administrative context**	**Authenticity Note**	**Restriction access**	**Access condition**	**Producer**	**Concerned parties**	**Receiving authority**	**Restriction access classifier**	**Signatory**	**Location of the physical record**	**Receiving place**	**Producer place**	**Receiving date**	**Producer date**	**Subscription date**	**Access restriction legal basis**	**Administrative motivation**	**Classification scheme**	**Subscription method**	**Total**	**Total %**
SR01	CIDOC-CRM	+	+	+	+	±	±	+	±	+	+	±	±	+	±	±	±	±	±	±	±	±	±	22	100%
SR05	RiC-O	+	+	+	+	+	±	±	+	+	+	±	+	+		+		+	+	±	+	+	+	20	91%
SR02	ArDO	+	+	+	+				+	+		+		+		+		±		+	+	±		13	59%
SR04	FaBiO	+	+	+			+	+	+	±		+				+	+	+		+		+		13	59%
SR26	PREMIS	+	+	+		+	+		+				+	+				+	+	+			+	12	55%
SR03	BibO	+	+				+		+	+	+			±		±		+		+				10	45%
SR08	GAO	+	+	+			+		+					+		+		+		+		+		10	45%
SR31	BBCprov	+	+	+			+		+			+				+		+		+		±		10	45%
SR34	OAI-ORE	+	+	+			+		+			+				+		+		+				9	41%
SR07	ARKIVO	+	+		±				+	±						±		+				+		8	36%
SR29	OPMO	+	+		±				±	±			±					±			±			8	36%
SR06	ArCo	+	±					±	±	+				+								+		7	32%
SR28	PROV-O				+				+	+				+		+		+			+			7	32%
SR21	SBA	+	+	+					+					+				+						6	27%
SR30	PAV	+	+		±				+							+		+						6	27%
SR41	HERO								±	±				±	±	±								5	23%
SR22	TIMBUS				±				±	±														3	14%
SR15	PERICLES	±																						1	5%
SR20	BDA								+															1	5%

In addition to its coverage percentage, the recognition of this structured resource as an ISO standard increases its adoption, as it adds characteristics such as consistency, standardization, continuous improvement, and global recognition to the resource. However, due to its comprehensiveness, this resource can be complex to understand and implement for institutions with a lack experience in adopting standards aimed at interoperability. This research includes a comparative analysis of CIDOC-CRM (SR01) and RiC-O (SR05), given the relevance of these two structured resources to the study, which can be viewed in Section 3.7.

Structured resource Archive Dynamics Ontology (ArDO) (SR02), like CIDOC-CRM, is also applied in the field of cultural heritage but aims specifically at organizing multimedia resources present in historical records that are being digitized. It is a recent resource, introduced in 2021. Its narrower focus justifies a lower coverage ranking (59%). This ontology can be enhanced and developed in a number of ways, including: linking ArDO to a classification framework to automatically generate annotations for archive records to support the manual annotation process, evolving the design of the ontology to represent more refined data from archive records, and in the long term being a substantial first step towards providing FAIR archive documents that support all four FAIR principles (Vsesviatska *et al.*, 2021).

Another resource that operates in the field of cultural heritage is Italian Cultural Heritage Knowledge Graph (ArCo network) (SR06), which is structured in seven vocabularies for the description of cultural heritage and a large dataset. It includes modules for cultural events, denotative descriptions, location, and context-description. Given that many of our aspects refer to administrative dimensions of archives that are not particularly relevant to cultural artifacts, ArCo has lower coverage in our analysis (32%). It should be noted that ArCo focuses mainly on Italian Cultural Heritage which means that its coverage of cultural entities from other countries could be limited. However, there are some areas for improvement that should be considered, for example, to facilitate the reuse of ArCo ontologies, the project team plans to develop additional support for the tool for cultural heritage data owners and to meet the requirements coming from the library and archive domains (Carriero *et al.*, 2019).

Structured resources FRBR-aligned Bibliographic Ontology (FaBiO) (SR04) and Bibliographic Ontology (BibO) (SR03) are focused on the semantic description of bibliographic references. FaBiO covers 59% of the aspects, given its (relatively broad) aim of "describing entities that are published or potentially publishable". They cover slightly different ground (Shotton, 2011; Biagetti, 2018). FaBiO is explicitly designed to conform to the Functional Requirements for Bibliographic Records (FRBR), an internationally recognized standard for describing bibliographic resources. In contrast, BibO offers a broader ontology for modeling bibliographic information, covering diverse aspects of bibliographic resources and metadata without being confined to the constraints of the FRBR model. FaBiO closely aligns with the concepts and relationships defined by FRBR, whereas BibO offers a more adaptable framework for the description of bibliographic resources. Bibo and FaBiO offer only partial insights into the complexities of archival records. This is because archival records differ significantly from bibliographic records in their intended use and scope of information. Furthermore, these ontologies often fail to adequately address the diverse range of archival record types and their intricate relationships (Carriero *et al.*, 2019).

RiC-O (SR05) is an OWL ontology to describe archival record resources and their contextual entities. It is a recently developed standard; the first beta version of RiC-O was developed in 2017, provided by International Council on Archives. It expands the archival description by allowing the representation of not only individual records, but also the relationships between records, their creators, and the activities they document. Because of this, it also ranks high in the coverage of the various aspects of the archival domain (with a coverage of 91%). One of the most relevant potentials of the structured feature of RiC-O is its capacity to provide a semantic framework that enables the representation of complex relationships between records and their contexts. This directly assists in enhancing the retrieval and access features of records, as well as aiding users in better understanding the lifecycle of records. However, RiC-O, being a resource in development, has its challenges, with the most relevant being the introduction of an ontological approach to represent archival records. Until now, the archival community has relied on representing archival records according to ISAD(G). This shift will require archival professionals to undergo a new learning process in order to effectively utilize this resource. See Section 3.7 for a comparative analysis of CIDOC-CRM (SR01) and RiC-O (SR05).

ARKIVO (SR07), like RiC-O, is intended for archival description, but its scope is limited to collections of historical documents, which makes it difficult to fully capture the unique requirements of different archival resources and their relationships (Carriero *et al.*, 2019). ARKIVO was developed in 2018 and provides classes to model the structure of archives as well as historical events, however, its class hierarchy for describing archive records may only partially fit in the scenario where an archive record may belong to several other archive records (Carriero *et al.*, 2019).

The Government Archival Ontology (GAO) (SR08) was developed based on the theories and methods of the semantic web with the purpose of implementing semantic interoperability between government records and performing the representation of knowledge and reasoning about its content, contributing to more efficient and effective management of government records, but its scope is limited to government archives of China. However, it introduces an innovative approach by linking archival metadata with the 5W1H framework, resulting in a logical event description of the archive content, facilitating organization and management of archive resources based on content (Wang *et al.*, 2021).

Another structured resource in this domain is the Historical Event Representation Ontology (HERO) (SR41), which presents a process for developing a semantic model through interaction between computer scientists, historians, and archivists. The top-level semantic model defined in HERO has guided the construction of a system that must "know" the documents and understand their content. This ontology serves as a valuable semantic knowledge system, aiming to offer a comprehensive framework for organizing and connecting historical events, participants, places, and time intervals (Baldo; Goy; Magro, 2018), which provides the basis for representing things that happen in the physical or social world at the scale of human affairs (Goy; Magro; Baldo, 2019). HERO is a central ontology with a modular design, mainly aimed at describing historical events, thus allowing for a better representation of the records in historical archives, facilitating their access, analysis, and interpretation. However, representing current events can be challenging for HERO, as historical events are static, while current events are subject to continuous change.

PERICLES (SR15) was developed to advocate a new concept to address digital preservation, in which preservation is part of the life of the digital object. Born-Digital Archives (BDA) (SR20) and Software-Based Art (SBA) (SR21) are also related to the PERICLES project. In PERICLES, the digital media objects area covers three different subdomains; among them, BDA (SR20) may be applied to a digital item, and SBA (SR21) can be applied to acquired artworks. PERICLES was a four-year initiative (2013-2017) funded by the European Union (EU) as part of its Seventh Framework Programme (ICT Call 9). The project aimed to address the challenges associated with ensuring the ongoing accessibility of digital content within a dynamic and evolving landscape. A fundamental component of PERICLES' methodology involved the development of a model-driven approach, centered around the Linked-Resource-Model (LRM), which delineates dependencies within digital ecosystems (Kontopoulos *et al.*, 2016). Within the PERICLES framework, three domain-specific ontologies have been developed to address the complexities of digital preservation risks. This framework highlights two of these ontologies: SBA and BDA. The SBA ontology focuses on assessing the risks associated with newly acquired works of art, considering not only their technical dependencies, but also their functional, conceptual, and aesthetic intentions regarding significant properties (Daranyi *et al.*, 2015). On the other hand, in the context of BDA, the emphasis is on ensuring the original accessibility and maintenance of digital documents, covering various technical, aesthetic, and permission-related characteristics (Daranyi *et al.*, 2015). The PERICLES project encountered limitations, including the necessity to manage technological changes and the obsolescence of formats and software, as well as semantic and social changes that impact the interpretation and utility of digital content (Kontopoulos *et al.*, 2016; Daranyi *et al.*, 2015). It is important to note that the PERICLES project has been concluded and, thus, limiting the opportunities for further improvements.

Similar to the PERICLES project, TIMBUS (SR22) introduces the need to analyze and sustain accessibility in business processes, and aims to introduce a new approach to digital preservation. In the context of TIMBUS, seven ontologies built during the period 2011 to 2014 are available. This EU co-funded project focuses on capturing the context of a business process, to enable the documentation, preservation and later analysis and redeployment of the process in a new environment (Mayer *et al.*, 2015). TIMBUS takes advantage of an ontology-based meta-model to structure the relevant information needed to preserve processes in various domains, and this meta-model consists in modeling the process context and dependencies, and their semantics (Mayer *et al.*, 2015). Integrating multiple preservation metadata models with different frameworks and standards can be challenging. Moreover, maintaining cohesion and consistency among these models requires significant coordination and continuous updates (TECO, 2024). Consequently, this complexity highlights the need for robust integration strategies to ensure that contextual information is accurately preserved and accessible across various systems (TECO, 2024).

PREMIS or PREservation Metadata: Implementation Strategies is a widely used international metadata standard for the preservation of digital objects and the PREMIS OWL Ontology (SR26) is an ontology developed to represent the PREMIS metadata standard for the preservation of digital objects (PREMIS Editorial Committee, 2022). This metadata, which includes information about the object's format, provenance, rights, and technical characteristics, is captured by the PREMIS OWL Ontology to ensure that essential preservation information is documented and accessible (PREMIS Editorial Committee, 2022). Furthermore, this ontology aims to introduce theories and methods of the Semantic

Web to formalize knowledge about digital preservation; by incorporating this feature it is possible to achieve interoperability and query capabilities (PREMIS Editorial Committee, 2022). The goal is to simplify the ontology by using elements from established ontologies when their meanings are aligned. It forms connections between PREMIS entities and elements from other ontologies, uses Uniform Resource Identifiers (URIs) for identification, follows RDF structures instead of XML (Extensible Markup Language), and categorizes objects as classes, among other principles (PREMIS Editorial Committee, 2022). The PREMIS 3.0 ontology faces limitations due to the community's lack of necessary knowledge in semantic web and digital preservation, highlighting the need to encourage widespread use within the community (Iorio; Caron, 2016). To address these challenges, developing simpler versions of the ontology for basic use cases and providing comprehensive documentation and implementation guides can facilitate the understanding and adoption of PREMIS. Additionally, creating mappings and integrations with other metadata standards, such as Dublin Core Metadata Initiative (DCMI)[2] and PROV Ontology (PROV-O), can improve interoperability and encourage broader adoption (Iorio; Caron, 2016).

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness (W3C, 2011-2013). Structured resources as PROV-O (SR28), OPMO (SR29), PAV (SR30), and BBCPROV (SR31) have provenance as their domain area. Given their narrower focus, all of them have a coverage below 50%. But, in the context of Electronic Document and Records Management System (EDRMS), the principle of provenance is of vital importance, as it brings a rationalization of documents and records within organizations.

The PROV Ontology provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts, and can be used in a wide range of applications (W3C, 2011-2013), thus allowing users to use parts of the ontology based on their requirements and the desired level of detail in the provenance information (W3C, 2011-2013). PROV-O is designed for the generalized description of provenance and exchange between different systems, and needs to be merged with other resources to enable the description of preservation metadata used for digital preservation (Li; Sugimoto, 2014). An important area for potential improvement is the enhancement of semantic provenance. This enhancement aims to ensure that the meaning of provenance information is conveyed with clarity and accuracy, thereby making it comprehensible to both human users and automated systems (Li; Sugimoto, 2014).

Open Provenance Model (OPM) is a research result of the International Provenance and Annotation Workshop (IPAW), and the purpose of the OPM OWL Ontology

[2] "The Dublin Core", also known as the *Dublin Core Metadata Element Set*, is a set of fifteen "core"

(SR29) is to capture the concepts of the OPM model and the valid inferences within this model (Moreau *et al.*, 2011). OPMO is designed to assist the interoperability between provenance information on the Semantic Web and to support provenance descriptions for datasets beyond those in the Web of Data (Zhao, 2010). Although the edges in an OPM graph convey a clear informal meaning, they lack formal semantics. Establishing formal semantics for OPM can provide a solid foundation for interpreting provenance data (Kwasnikowska; Moreau; Bussche, 2015). To improve it, it is possible to propose an extended model that allows imprecise edges to be asserted in a graph, propose a formal semantics that allows a rigorous notion of inference, and investigate sound and complete inference in OPM graphs (Kwasnikowska; Moreau; Bussche, 2015).

Provenance, Authoring, and Versioning Ontology (PAV) is a lightweight ontology for capturing descriptions essential for tracking the provenance, authoring, and versioning of web resources (Ciccarese *et al.*, 2013). However, this can also be a limitation in scenarios that require detailed, process-oriented provenance tracking (Ciccarese *et al.*, 2013). To address this, PAV ontology extends the W3C Provenance Ontology (PROV-O) to include specific classes and properties that identify the various roles agents assume when manipulating digital artifacts, such as author, contributor, and curator (Ciccarese *et al.*, 2013). Nevertheless, this limitation is a possible area for improvement in the PAV ontology. By integrating detailed process modeling resources similar to those found in PROV-O, PAV could provide more comprehensive resources for describing the creation, manipulation, and transformation of digital resources. This resource is relevant in the context of digital scientific content, where establishing trust and authenticity is very important (Ciccarese *et al.*, 2013).

The British Broadcasting Corporation (BBC) produces a plethora of rich and diverse content on subjects that interest its audience, so the BBCPROV represents a significant step towards structured data management and transparency in BBC products content provenance (BBC, 2012). Designed to support data management and auditing tasks within BBC's Linked Data Platform, it is used to define the different types of named graphs used in the store and enables their association with metadata that allows to manage, validate, and expose the data to BBC services (BBC, 2012). But it is important to note that the ontologies produced by the BBC are limited to BBC products and, as a result, their applicability may be constrained when used outside the context of the BBC, as the terminologies and structures may not align perfectly with the requirements and practices of other organizations. To enable the adoption of the BBC Provenance Ontology, which would be to extend its applicability beyond the specific content of the BBC, one possibility would be to promote the development of more generalized structures and terminologies that can be easily adapted and integrated by other organizations.

elements (properties) for describing resources. For more details, see (Dublin Core Metadata Initiative, 2023).

To conclude this overview of each resource listed in Table 5, there is OAI-ORE (SR34) conducted under the umbrella of the Open Archives Initiative (OAI). The work of OAI is to promote broad access to digital resources for eScholarship, eLearning, and eScience (OAI, 2001). In the physical world, aggregations of things are continuously created, used, and referenced, and these aggregations are frequently tangible. However, abstract entities are also aggregated (OAI, 2001). This practice of aggregating extends to the Web, where we accumulate Uniform Resource Locators (URLs) in bookmarks or favorites lists in our browsers, organize photos into sets on popular sites, navigate through multi-page documents that are linked together, and discuss websites as if they possess a tangible existence beyond the pages of which they consist (OAI, 2001). The OAI-ORE provides a standard for the description and exchange of aggregations of web resources, allowing the combination of distributed resources with multiple media types and exposing rich content in aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation (OAI, 2001). The Open Archives Initiative is no longer working on new projects, but the project's website <https://www.openarchives.org/> and the specifications associated with OAI-ORE and other projects of this initiative continue to be maintained (OAI, 2001).

The analysis of all of the structured resources presented here aims to provide an overview of the features of each one. However, it is important to note that this is a basic overview and does not claim to exhaust all the information available on the subject. In order to provide a more succinct overview of the contributions of each structured resource to the ontology under development, Table 6 summarizes the possible contributions; it details the points at which each structured resource complements specific aspects of the archival domain. Its aim is to provide an organized basis for the contributions, ensuring that the ontology is built coherently and in line with existing reality.

Table 6 – Potential contributions from Structured Resource

Identifier	**Structured Resource**	**Contribution**
SR06	ArCo	Provide ontology standards to connect people and places about cultural heritage events.
SR02	ArDO	How to describe the hierarchical nature of archival multimedia data, as well as its application.
SR07	ARKIVO	Provide classes to model the structure of archives, as well as historical events.
SR31	BBCPROV	How to track the provenance of digital media artifacts.

Continued on the next page

Identifier	**Structured Resource**	**Contribution**
SR20	BDA	In organizing digital archival records that were originally created in digital format.
SR03	BibO	Help to describe bibliographic entities, such as citations and documents.
SR01	CIDOC-CRM	Helping to describe related collections from various sources, such as museums, libraries, and archives.
SR04	FaBiO	How to describe entities that are published or potentially publishable.
SR08	GAO	How to improve the representation and semantic interoperability of digital records.
SR41	HERO	In capturing different perspectives and aspects of historical events, including thematic and social roles, organizations, scenarios, and representations of time.
SR34	OAI-ORE	How to enable interoperability, digital preservation, discovery, access, and reuse of complex aggregations of digital resources through open standards.
SR29	OPMO	How to capture and describe data provenance.
SR30	PAV	How to identify and support the clear differentiation of authors, curators, and versions of records.
SR15	PERICLES	How to facilitate the representation and interoperability of historical data.
SR26	PREMIS OWL	How to capture and describe metadata related to digital preservation.
SR28	PROV-O	How to model and describe provenance.
SR05	RiC-O	How to describe archival record resources and their relationships.
SR21	SBA	How to capture and describe the characteristics and provenance of digital artifacts.
SR22	TIMBUS	How to model and describe the context of business processes and services.

End of table

3.7 Comparative analysis of CIDOC-CRM and RiC-O

Within the realm of cultural heritage and archival description, two prominent resources have emerged as foundational pillars: RIC-O and CIDOC-CRM, offering comprehensive frameworks for representing complex information structures within their respective domains.

Both structured resources aim to facilitate semantic interoperability in their respective application domains. CIDOC-CRM focuses on modeling cultural heritage information and RiC-O on archival description, but they provide structures for representing historical objects, events, agents, and their contextual relationships, supporting the documentation and preservation of cultural heritage and archival materials. Each offers advanced descriptive features, allowing for detailed representations of entities and their attributes, thus facilitating comprehensive documentation and analysis in their respective domains. These structured resources reflect the interdisciplinary contribution of professionals, thus ensuring that they meet the multiple needs of their fields of application. They are resources built by organizations with a strong reputation worldwide and considered to be authorities in their fields. CIDOC-CRM is a resource of the International Committee for Documentation (CIDOC), an organization within the International Council of Museums (ICOM), and RiC-O is a resource of the International Council on Archives (ICA) developed by the Expert Group on Archival Description (EGAD). Both resources aim to carry out broader standardization efforts to promote best practices in their domains.

Despite their similarities, these two resources have characteristics that distinguish them. CIDOC-CRM provides a common semantic structure for information on cultural heritage, allowing data from various sources to be integrated. On the other hand, RiC-O focuses specifically on the description of archives and the intellectual representation of records, and is close to the archival community, while CIDOC can be used by various specialists who work in museums, libraries, and archives. CIDOC-CRM has been in continuous development for the longest period. Its initial version was released in 2004, and it was officially recognized as an ISO standard in 2006. In contrast, RiC-O is still in its developmental stage and is a more recent standard. The first beta version of RiC-O was developed in 2017.

CIDOC-CRM and RiC-O can evolve by simplifying their structures for basic use cases, improving documentation with practical guides, and developing supportive tools such as mapping and visualization tools. Additionally, increasing community engagement through workshops and training, and improving integration with other metadata standards such as Dublin Core and PREMIS, can foster broader adoption. Raising awareness, providing comprehensive documentation, and encouraging active participation in model refinement can further enhance their usability and relevance, ultimately facilitating better integration, preservation, and access to cultural heritage and archival information.

Both structured resources make a significant contribution in each area of expertise, providing standardization of terminologies and metadata, conceptual modeling, and support for semantic technologies. It is clear from the efforts of ICA and ICOM that they are concerned with the effective documentation, preservation, and access to cultural heritage and archival records in an increasingly digital and interconnected world. Exploring all the resources made available by these two structured resources is out of the scope of this research and the interested reader can refer to their full documentation. The comprehensive documentation for CIDOC-CRM can be accessed via <https://cidoc-crm.org/>, while documentation for RiC-O is available through <https://www.ica.org/resource/records-in-contexts-ontology/>.

3.8 Related Work

This section presents works closely related to the application of the CLeAR approach in this dissertation, focusing on those that use systematic mapping studies in the archival domain or closely related areas.

The earliest study we have identified traces back to 1993, when Wallace (1993) surveyed the literature on metadata for electronic records and studied four leading electronic records programs in Canada and the United States. The author concluded in good foresight that "a metadata systems approach to the management of electronic records will become the basic strategy for archivists who must manage these systems." The author discusses the role of "data dictionary metadata" as "describing what the data is and what it means". Albeit with different terminology, it is clear the author was concerned with what we encompass here under the umbrella of "structure resources". The metadata listed (in general) in the study covers some of the aspects of our domain coverage analysis ('creator', 'title of the file', 'date of creation', 'date of modification', 'restriction on use', etc.).

Also predating any of the publications we have identified previously, a report in the context of the CURL Exemplars in Digital ARchiveS (CEDARS) Project (Day, 1998) aimed to "describe the state of the art in the area of metadata for digital preservation". The report includes a (non-systematic) review of metadata formats and initiatives in the specific area of digital preservation that were available (or under development) in 1998, namely, Digital Rosetta Stone (DRS), Dublin Core, Preserving and Accessing Networked Documentary Resources of Australia (PANDORA), the ISO Open Archival Information System (OAIS) reference model and others.

More recently, a number of authors have surveyed the literature with systematic approaches with a focus on specific application domains such as electronic healthcare. For example, Mello *et al.* (2022) presented a comprehensive systematic review of the literature on semantic interoperability focusing specifically on electronic health records.

Based on the protocol proposed by Kitchenham & Brereton (2013), the authors searched seven scientific databases, from which 6,032 studies were selected and through inclusion, exclusion, and quality criteria, they arrived at 28 articles. Also in the same domain, Adel *et al.* (2017) surveyed standards for the semantic interoperability of the electronic health records (EHR). As part of their research protocol, the authors adopted a methodology that includes defining research questions, describing the research strategy, presenting the results of the literature review, and analyzing the identified semantic interoperability frameworks. Initially, 8,274 documents were obtained, but after applying the inclusion and exclusion criteria, no results were obtained, which demonstrates that there are no works or articles that rely on a fuzzy ontology to solve the main problems of the target (Adel *et al.*, 2017). Given the focus on administrative records and the many domain-specific concerns that apply only to health records, we have chosen to explicitly exclude them from this research; therefore, these other works can be seen as complementary, respecting their corresponding domains.

A key aspect of the work was that it aimed not only to chart the literature in this domain, but also to identify a set of structured resources (reference models, representation schemes, data exchange formats, metadata standards, vocabularies, and thesauri) relevant to the surveyed domain, thus creating a knowledge base related to the archival records domain that can be used to develop new structured resources.

3.9 Summary

The systematic mapping used in this chapter follows a methodical approach to searching and analyzing structured features in the domain of archival records. Applying the CLeAR approach to the domain of archival records resulted in a set of 19 structured resources relevant to this domain. These structured resources were analyzed for their coverage of the domain, which reflects their ability to address the characteristics of fixity, organic nature, naturalness, uniqueness, authenticity, and impartiality in archival records. The analysis of the selected structured resources revealed that most of them are ontology-based approaches. Ontologies have become the prevailing formalism for representing knowledge and facilitating semantic interoperability. The Web Ontology Language (OWL) was the most frequently used formalism among the selected resources, highlighting its widespread adoption as the “lingua franca” for ontologies.

Analysis of domain coverage using a matrix-based approach provided valuable insights into the strengths and weaknesses of each structured resource. CIDOC-CRM (SR01) had 100% coverage in relation to aspects of the domain. It has been in continuous development for a significant period, making it a well-established and widely recognized resource. Other structured resources, such as PREMIS (SR26), ArDO (SR02), BibO (SR03), and FaBiO (SR04), also showed promising domain coverage. The structured resource RiC-O (SR05) requires special attention because it is part of the ongoing "Records in Contexts" initiative, led by the International Council on Archives (ICA), and aims to present a structured resource for archival description that allows for the representation of not only individual records, but also the relationships between records, their creators, and the activities they record.

4 Ontological Representation

This chapter is dedicated to describing the development process of the Archival Records Ontology (ARO), proposed to represent the archival record in a structured and semantic way. This chapter also covers the fundamentals of ARO, with explanations of its classes, stereotypes, and relationships, demonstrating how each component contributes to a better understanding of an archival record. In addition, the UFO concepts used in the modeling are discussed, offering the ontological basis used to choose the stereotypes, seeking to ensure semantic consistency and adherence to the characteristics of archival records. In addition, it explores the alignment of the ARO with the structured resources identified through the CLeAR approach and compares the ARO with the RiC-O and CIDOC-CRM models, highlighting their differences.

4.1 Building the Archival Record Ontology (ARO)

On a working day, civil servants create, send, receive, and use many records to help them carry out their activities. It can serve to support the various activities of an organization by recording decisions, processes, and transactions, ensuring transparency, accountability, and compliance with regulations, and providing a historical record for future reference and analysis. However, it is necessary that the concept of an archival record that flows between the various information systems of an organization be well-defined and standardized to ensure proper management, digital preservation, and effective interoperability. To meet this challenge, an ontological representation of an archival record will be developed aiming to contribute to the field of Archival Science and Software Engineering.

Building an ontology in the archival domain requires a systematic approach to ensure consistency, semantic expressiveness, and adequacy to the needs of the domain. According to Guizzardi (2005), a well-structured ontology must correctly capture the ontological commitments of the modeled domain, ensuring that the entities and relationships represented correspond accurately to real phenomena.

Figure 10 provides an overview of how the methodological steps were carried out and how different tools contributed to the development and verification. This visualization highlights the technologies used in each of the five phases of the SABiO methodology chosen to guide the ontology development process. SABiO, as already explained in Subsection 2.4.1, introduced by Falbo (2014), is a method for ontology construction, comprising five main phases: (i) Purpose Identification and Requirements Elicitation, (ii) Capture and Formalization, (iii) Design, (iv) Implementation, and (v) Testing.

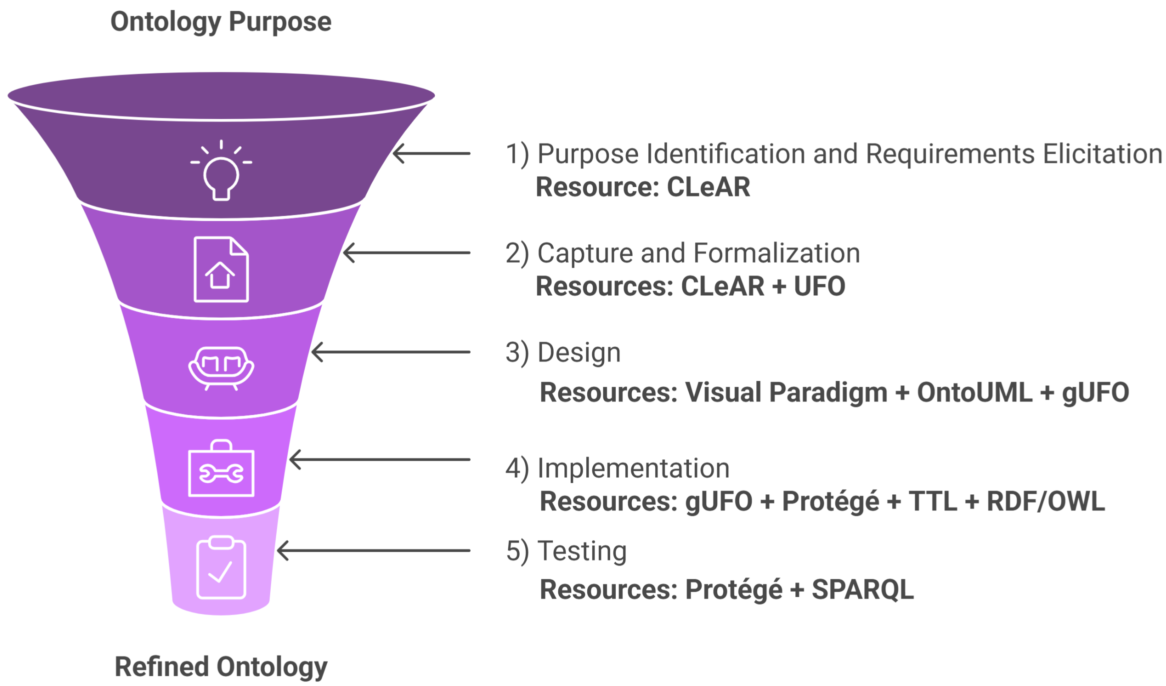

Figure 10 – Phases of the SABiO methodology and Technological Resources used for the development of the ARO Ontology.

It is important to clarify that not all phases are reported in this chapter; in particular, the results of the 'Purpose Identification and Requirements Elicitation' phase were reported already in Chapter 3, using the CLeAR approach to define the integration questions and the scope of the ontology. The results of the 'Capture and Formalization', 'Design' and 'Implementation' phases are reported here. The execution of the 'Testing' phase is reported in Chapter 5, in which the developed ontology will be subjected to a verification process to ensure that it maintains internal coherence and aligns with the requirements defined in the context in which it will be applied.

4.2 Describing the Archival Record Ontology (ARO)

The ontology will be introduced in a segmented way, allowing for a detailed and progressive analysis of its elements, which favors an understanding of each class and its relationships. This gradual approach is expected to provide a clearer view of the ontology's underlying structure and logic. In the end, the ontology will be presented in a complete form, so it will be possible to consolidate the understanding of the concepts, relationships and principles that underpin it, allowing a complete view of the model developed. In order to guarantee coherence, consistency, establish scope and limits, as well as restrictions related to cardinality and existence, the ontology description will also present axioms that help define the precise relationships between concepts, ensuring that the model represents the domain of interest. The purpose of the axioms is to reinforce the logical integrity of the ontology, allowing valid inferences and preventing ambiguities or contradictions.

The first class to be presented is **Record**, shown in Figure 11. It represents a «kind» of **Information Object** that can serve as official evidence of actions carried out by an institution. This class encompasses archival records such as corporate contracts, management reports, institutional correspondence, and other types of records that hold legal, administrative, or historical value. To represent this class, the «kind» stereotype was chosen, which provides the conceptual foundation for the ***Record*** class as a substantial sortal offering a rigid identity principle for its instances, meaning the entity retains its essential identity throughout its entire existence (Guizzardi, 2005). This class is considered an essential asset in the activities of an institution, thus necessitating a unique identity. This allows for consistent identification, irrespective of its role in events or alterations over time.

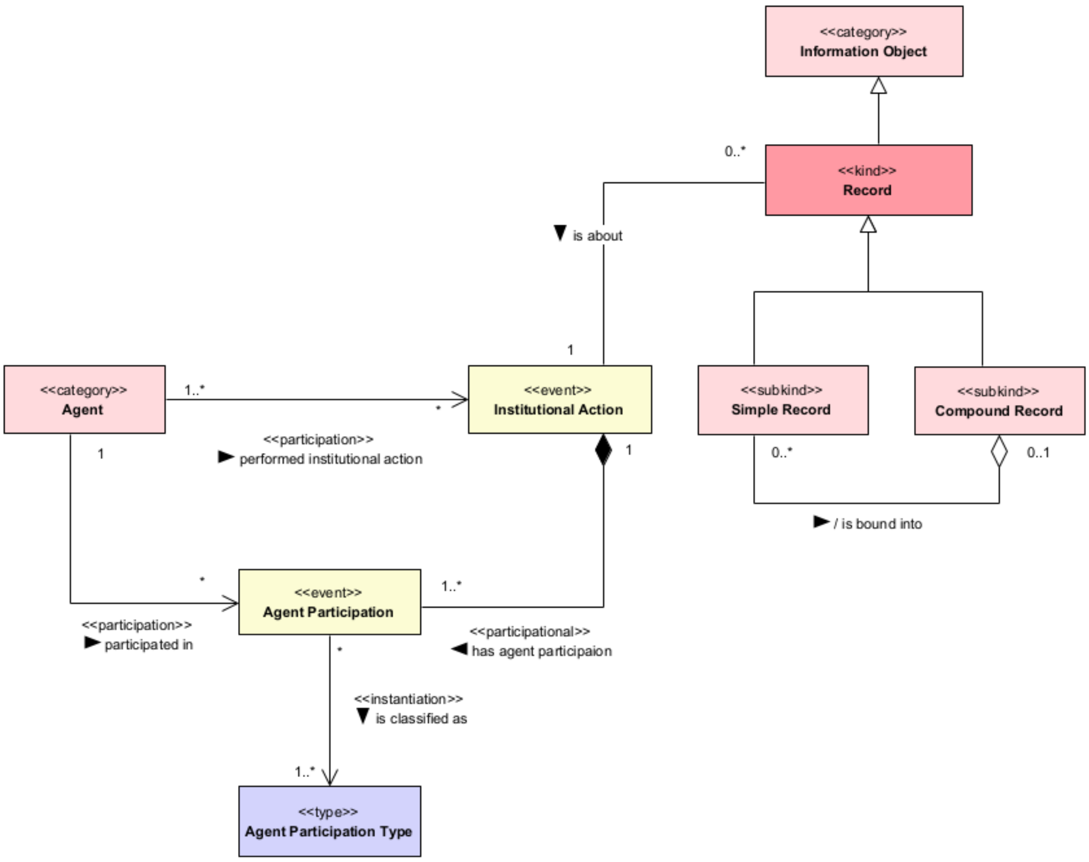

Figure 11 – Fragment of the Archival Record Ontology (ARO).

For this class, the axiom **Record** $\sqsubseteq$ **InformationObject** implies that each ***Record***, besides being an archival record with its own characteristics—like authenticity and integrity—is also considered an ***Information Object***.

The next class is the ***Information Object***, in Figure 11, which shows the informational objects that may include reports, contracts, or any other types of entities containing formal or structured information. The CCSDS (2012) defines an ***Information***

Object as a combination of a Data Object and associated Representation Information, because this union is necessary to transform data into meaningful information. This concept was adopted in this research as it aligns with the principles of digital preservation. By incorporating both data and representation information, an ***Information Object*** ensures that future users can correctly interpret its content, even as technology and context changes (CCSDS, 2012). A digital audio file (the Data Object) serves as an example. Alone, it does not convey all necessary information until it is accompanied by representation details, such as encoding format or sampling rate, that allow a system to accurately reproduce the sound.

The «category» stereotype provides the conceptual foundation for ***Information Object*** class. The goal of «category» stereotype is to group together different kinds of objects that are used to store and send formal information (examples: contracts, reports, and letters), without altering the identity principles already established by their respective specific types, since that is already done by the ***Record*** class.

Represented in the diagram in Figure 11, the ***Simple Record*** class symbolizes a record type that consists of a single archival record. It serves to represent individual records that do not require grouping with other archival records. The ***Simple Record*** class guarantees the unique modeling and preservation of these records as independent items, having all the properties of a formal ***Record***, yet remaining apart from a larger set.

An "Official Letter" sent and received as a unique and autonomous record could be considered as an example of a ***Simple Record***. It functions as a formal proof of communication between two parties. A second example: a "Birth certificate", maintained as a formal and singular record verifying that a birth has been registered.

The «subkind» stereotype provides the conceptual foundation for ***Simple Record*** class, meaning that the basic properties of its type, which in this case is the ***Record*** class, are kept while other properties that make it different from other entities are added (Guizzardi, 2005).

The axiom **SimpleRecord** $\sqsubseteq$ **Record** implies that ***Simple Record*** is a subclass of ***Record***, inheriting its essential properties. This ensures that ***Simple Record*** retains the fundamental characteristics of an archival record while introducing specific distinctions that differentiate it within the broader category of records.

Depicted in Figure 11, the ***Compound Record*** class illustrates a record composed of multiple ***Simple Records*** which are gathered and maintained as a set, as organizational business processes may require this structure. By modeling the organization of various archival records that, although independently created, are related to one another, these multiple ***Simple Records*** are treated as a cohesive set due to their shared purpose within organizational actions. The «subkind» stereotype is the basis for the conceptual foundation for the ***Compound Record*** class, implying the preservation of the essential properties of its kind, which in this model is the ***Record*** class.

The axiom $\textbf{CompoundRecord} \sqsubseteq \textbf{Record}$ implies that ***Compound Record*** is a subclass of ***Record***, inheriting its essential properties and ensuring that ***Compound Record*** maintains the basic characteristics of an archival record while being preserved as a formal set of records.

Example of a ***Compound Record***: an "Administrative Process". It includes various records such as official letters, reports, and decisions, all gathered into a single grouping. Another example: a "Patient Medical Dossier", comprising tests, reports, and prescriptions, all related to a single patient.

The ***Agent*** class, shown in Figure 11, represents entities that act within the lifecycle context of an archival record. These entities may be either persons or organizations — their role is fundamental to the execution of formal actions within an institutional context. This class models the participation of agents in the creation, signing, sending, addressing, and binding of archival records. Its purpose is to enable traceability and formalization of actions carried out by persons or organizations performing specific functions, ensuring identification of those responsible for each action.

The «category» stereotype provides the conceptual foundation for the ***Agent*** class, grouping entities that share essential characteristics without establishing a specific identity principle. According to Guizzardi (2005), by grouping different types of entities through shared essential characteristics, the «category» stereotype allows for organization without imposing an identity principle or specific instances on the resulting categories.

In addition to the functions and participations defined for the ***Agent*** class, it is essential to point out intrinsic characteristics of archival records, such as organic nature, which reflects the natural organization of records within institutional contexts, and naturalness, which ensures that the record is generated as an authentic product of institutional activities. These characteristics reinforce the importance of the actions and responsibilities associated with the role of the ***Agent***, guaranteeing that archival records remain faithfully connected to their context of creation throughout their lifecycle.

In the diagram presented by Figure 11, the ***Agent Participation*** class illustrates how an ***Agent*** class participates in a phenomenon that needs institutional action. Each ***Agent Participation*** instance is associated with exactly one ***Agent***, in accordance with the cardinality constraints shown in the diagram. This action may create archival records that serve as official proof of the institutional activity. For example, the approval of an employee's leave request exemplifies this process, as it requires a formal institutional action that results in the creation of an official record of the decision. This class highlights the importance of clarifying the underlying phenomena of institutional acts and reinforcing the authenticity of the generated archival records. Another example of the ***Agent Participation*** class is when an institutional department needs to acquire medications due to a shortage of these items in the warehouse. Here, the agent steps in when a confirmed medication shortage prompts an institutional action, such as submitting a purchase request, which in turn creates an official record confirming the action.

The «event» stereotype provides the conceptual foundation for ***Agent Participation***. It permits to represent institutional events in which agents take part in actions that have specific effects on the system, like making an archival record. An «event» is defined as something that occurs over a time interval, leading to changes in the state of the entities involved (Guizzardi; Falbo; Guizzardi, 2008). ***Event*** is very important in ontological modeling, according to Almeida, Falbo & Guizzardi (2019), since it shows changes that happen in entities and their relationships, which lets us describe important dynamics in modeled systems. Because it captures the temporal dimension of agents' participation, this stereotype enables the modeling of events that unfold over time and necessitate the presence of agents at specific moments, thereby recording and tracing each participation as part of an ongoing institutional process.

The axiom $\boldsymbol{AgentParticipation} \sqsubseteq = 1\ \textbf{participatedIn}\ \boldsymbol{Agent}$ establishes that one *Agent* must participate in each instance of *Agent Participation*, ensuring that each instance of participation is associated with a specific agent, which contributes to archival record characteristics such as authenticity by establishing a clear link between institutional actions and the agents responsible for them. The relationship between the ***Agent*** class and the ***Agent Participation*** class highlights the central role of *Agent* as a category capable of operating in many contexts and events, maintaining its intentional function and reinforcing the fundamental role of ***Agent*** in contexts where intentionality and traceability are necessary.

Next, Figure 11 illustrates the conceptual structure of ***Agent Participation Type*** class, highlighting its relationship with the ***Agent Participation*** class, which allows for the classification of different types of participation in phenomena underlying institutional actions. Its purpose is to categorize the real-world phenomena that drive the management of archival records, enabling the capture of the context in which these agents interact and the impact on institutional record production. Thus, this class aims to ensure that the records reflect not only institutional actions but also the phenomena themselves.

The ***Agent Participation Type*** class follows the powertype pattern and is modeled using the «type» stereotype, which consists of universals that group together similar characteristics, enabling the classification of entities into classes with common essential properties (Guizzardi *et al.*, 2021). It also establishes connections between individuals and universals, ensuring semantic consistency and compatibility across various ontological representations (Guizzardi *et al.*, 2021). In this case, the instances of this class are types that classify ***Agent Participation***. This provides a way of classifying various types of agent participation, facilitating a well-defined hierarchy. For example, in a public organization, there may be different types of participation for an agent, such as Participation in a Bidding Process or Participation in a Disciplinary Committee.

Moving on, the ***Institutional Action*** class, in the diagram presented in Figure 11, represents formal actions carried out by an organization within an institutional context. These actions may include the creation, signing, dispatch, or processing of records, among other formal activities. The objective of the ***Institutional Action*** class is to model events that occur within the institutional scope, capturing the formal nature of these actions and their relevance within an organizational process. This class enables the representation of events involving decisions, responsibilities, and administrative changes.

The «event» stereotype provides the conceptual foundation for the ***Institutional Action*** class by capturing its nature as an occurrence that brings about formal changes in the institutional domain. According to Almeida, Falbo & Guizzardi (2019), "events map the world from situations (that activate the dispositions of which they are manifestations) to situations (which are brought about by the occurrence of that event)".

The axiom $\mathbf{InstitutionalAction} \sqsubseteq \exists\ \mathbf{performedBy\ Agent}$ specifies that *Institutional Action* requires the involvement of at least one *Agent*, which ensures that institutional actions engage accountable agents, making these formal activities traceable and enabling the identification of responsible parties.

Continuing, the diagram in Figure 12 shows the ***Organization*** class, which represents the collective entities that act as intentional agents within organizational contexts. Such entities may include companies, government institutions, or any other type of formal organization that plays a role in administrative processes. The objective of the ***Organization*** class is to capture the identity and actions of such entities as participants in institutional actions, whether as producers or accumulators of records. Additionally, this class allows for modeling the continuity and stability of the organization's identity, even when the individuals who work within the organization are replaced.

The «kind» stereotype provides the conceptual foundation for the ***Organization*** class, ensuring that its identity is intrinsically defined rather than being determined by the collection of its members. This allows the organization to maintain a stable participation in institutional events despite changes in its membership structure.

The axiom ***Organization*** $\sqsubseteq$ ***Agent*** indicates that an organization is an agent, endowing it with the characteristics of an intentional and accountable entity, similar to individual agents but with a collective identity principle. In this way, the ***Organization*** does not rely on specific members to exist as an entity, reinforcing its status as an organizational unit with its own independent identity.

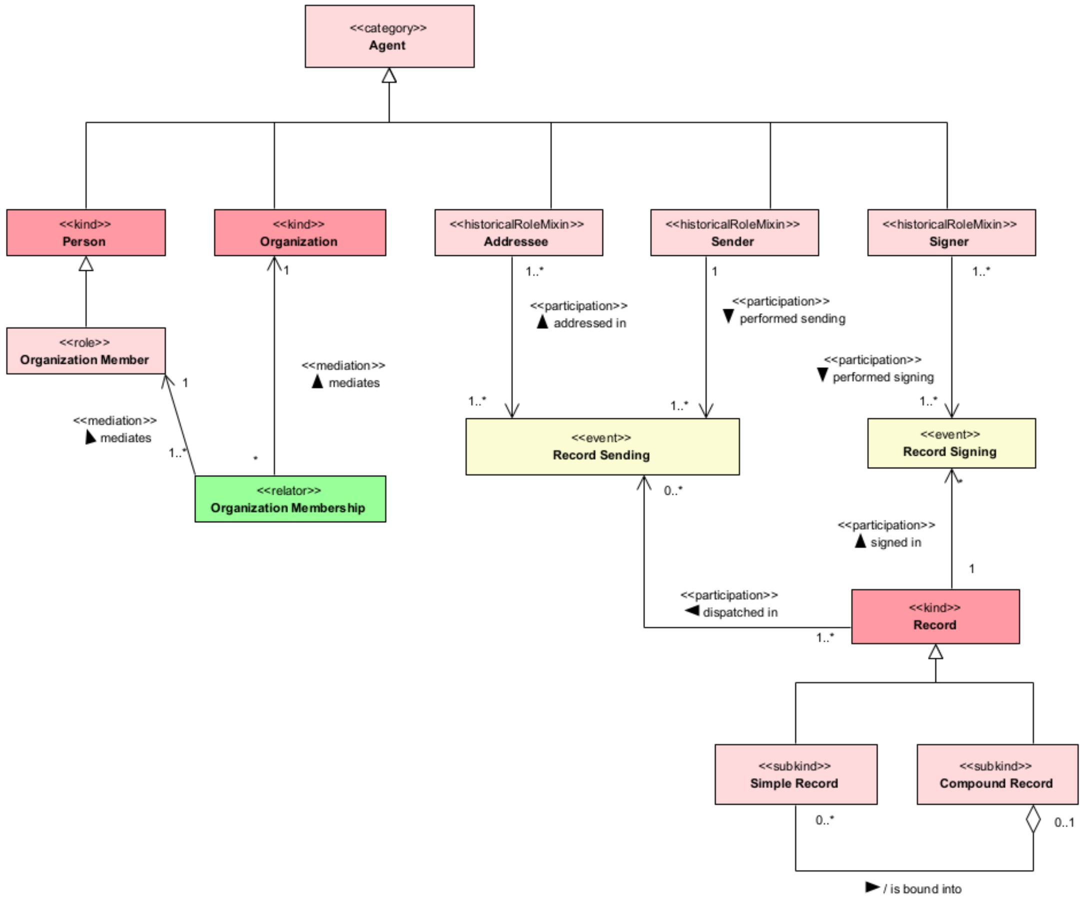

Figure 12 – Fragment of the Archival Record Ontology (ARO).

A practical example of ***Organization*** is a company that, as a formal entity, enters into a contract in partnership with another organization. Even if the employees involved in this process could be changed over time, the company remains the same entity responsible for fulfilling the contract. Another example would be a government agency that sends an official memorandum; even if its employees are substituted, the agency as an organization maintains its identity and formal role in institutional actions, without requiring the continuity of the same individual employees to identify which organization sent the memorandum.

The ***Person*** class in the diagram, Figure 12, represents the participation of individuals in institutional actions, such as signing contracts, sending archival records, or creating new records. A practical example is an employee of a company that could be the person responsible for signing a contract on behalf of the company. Here, the individual assumes the role of "Signer", yet his/her identity as a person remains the same throughout the process. The company consistently recognizes him/her as a person with a fixed identity, even if this person performs different functions, such as "Signer" or "Sender".

The axiom **Person** $\sqsubseteq$ **Agent** establishes that ***Person*** is a subclass of ***Agent***, thus ensuring that people function as agents in institutional actions, being responsible and active participants in formal actions within an organizational context.

Moving on, the next class to be presented is the ***Organization Member*** class. In the diagram in Figure 12, this class characterizes the role an individual assumes when they become a part of an organization, with the ***Organization Membership*** class mediating this affiliation. In this way, it is possible that the member's functional role within the organization may be identified. Such role can change over time without compromising their identity.

The «role» stereotype provides the conceptual foundation for the ***Organization Member*** class referring to an anti-rigid sortal, meaning that entities holding a role are not required to maintain it throughout their entire existence (Guizzardi, 2005). This implies that a ***Person*** can adopt different roles throughout their journey within the organization. The axiom **OrganizationMember** $\sqsubseteq$ **Person** establishes that ***Organization Member*** is a subclass of ***Person***, meaning that only individuals can fulfill the role of member within an organization.

The ***Organization Membership*** class, in the diagram of Figure 12, represents a mediating entity that connects the ***Organization Member*** to the ***Organization*** of which they are a part. The model captures the establishment and maintenance of this connection over time, and its purpose is to ensure the formalization of the relationship between the parties and to reaffirm the member's participation in organizational activities as a formal bond. Example: an employment contract, which mediates the formal relationship between an employee and the company for which they work.

The «relator» stereotype refers to a concept that mediates relationships between other entities, allowing these entities to be connected within a specific relationship (Guizzardi; Falbo; Guizzardi, 2008). The marriage relationship can be cited as an example, because in order for marriage to exist between two people, a marriage event must have taken place, creating a "marriage" «relator» that connects the two parties involved. Ontologically, the relator is not merely an abstraction; it incorporates rights and responsibilities associated with the role each entity plays in the mediated relationship (Guizzardi; Falbo; Guizzardi, 2008).

The axiom **OrganizationMembership** $\sqsubseteq \exists$ **mediates.Organization Member** establishes that an instance of ***Organization Membership*** must be associated with an ***Organization Member***. Similarly, the axiom **OrganizationMembership** $\sqsubseteq \exists$ **mediates.Organization** ensures that ***Organization Membership*** is necessarily linked to an ***Organization***. These axioms reflect that ***Organization Membership*** mediates the relationship between an ***Organization*** and an ***Organization Member***, ensuring that institutional relationships are formally established within the organizational structure. By capturing and guaranteeing this connection along with the chosen stereotype, the class contributes to the authenticity of the archival record, as it establishes auditable evidence of relationships.

The ***Record Sending*** class in the diagram of Figure 12 represents the formal act of transmitting an archival record, denoting the official transfer of a record from one agent to another. This class encapsulates the moment of the archival record's transmission, thereby formalizing the communication among the participating agents. The ***Record Sending*** class ensures the traceability of institutional actions by expressing the transmission of an archival record. The «event» stereotype provides the conceptual foundation for the ***Record Sending*** class, allowing the explicit representation of the temporal occurrence involving agents engaged in the transmission of an archival record. It allows transmission to be modeled as a separate action traceable over time, thereby helping to create a chronology of the archival record's journey within the organization, as well as allowing the agents involved to be identified.

The ***Addressee*** class, in the diagram of Figure 12, represents the role performed by an entity as the recipient of an archival record since it captures the function assumed by an ***Agent*** when formally receiving an archival record. The «HistoricalRoleMixin» stereotype provides the conceptual foundation for the ***Addressee*** class, as it is defined as "a stereotype used to capture roles that are historically played by entities following different identity principles, yet sharing a common role within a specific event" (Sales; Fonseca; Barcelos, 2022).

In this class, the axiom **Addressee** $\sqsubseteq$ **Agent** defines ***Addressee*** as a subclass of ***Agent***, characterizing the addressee as an entity capable of intentionality or responsibility in the process of receiving archival records. In addition, the axiom **Addressee** $\sqsubseteq$ $\exists$ **participatedIn.RecordSending** determines that ***Addressee*** is related to record-sending events. This indicates that the addressee role is specifically assumed during the transfer of an archival record from one point to another, guaranteeing that an agent instantiates ***Addressee*** only if there is an archival record being received.

On the other hand, the ***Sender*** class in the diagram of Figure 12 represents a historical role assumed by an ***Agent*** in the act of formally sending an archival record. This means that ***Sender*** will only exist when an ***Agent*** participates in a RecordSending event. The «HistoricalRoleMixin» stereotype provides the conceptual foundation for the ***Sender*** class. The axiom **Sender** $\sqsubseteq$ **Agent** defines ***Sender*** as a subclass of ***Agent***, implying that the sender is an entity with the capacity for intentionality or responsibility in the process of sending archival records. Additionally, the axiom **Sender** $\sqsubseteq$ $\exists$ **participatedIn.RecordSending** indicates the ***Sender*** class is associated with archival record-sending events, showing that the sender role is specifically assumed during the transfer of an archival record from one point to another.

The ***Record Signing*** class in the diagram of Figure 12 denotes the formal event of signing an archival record, leading to the formalization of an archival record by the signature of one or more persons, which, upon execution, can support the presumption of authenticity to the archival record. This class intends to capture the exact moment of signature execution during an institutional activity. The «event» stereotype provides the conceptual foundation for the ***Record Signing*** class, allowing the signature event to be modeled as a separate action traceable over time, thus helping to create a chronology of the signature and allowing the agents involved to be identified.

Moving on, the ***Signer*** class in the diagram of Figure 12 represents the historical role assumed by an entity when signing an archival record. This action is an important validation process for the archival record, as the signature confirms a party's acceptance or agreement with the content recorded in the archival record. The ***Signer*** class is used to model the participation of agents in the signing act, highlighting the nature of this role, which only exists when the archival record is signed. The «HistoricalRoleMixin» stereotype provides the conceptual foundation for the ***Signer*** class.

The axiom **Signer** $\sqsubseteq$ **Agent** establishes that ***Signer*** is a subclass of ***Agent***, implying that the signer is an entity with responsibility and intention when signing the archival record. Besides this, the axiom **Signer** $\sqsubseteq \exists$ **participatedIn.RecordSigning** indicates that ***Signer*** is directly associated with archival record-signing events, establishing that the role of ***Signer*** is specifically performed during the act of signing. The signature is one of the core elements to ensure the authenticity of an archival record, as it serves as evidence of the approval and validation of the content by the involved agents. Such bond allows for the traceability of authorship and responsibility, which are fundamental aspects of the legal validity of archival records.

Depicted in the diagram in Figure 13, the ***Organization Member Actor*** class represents the role an ***Organization Member*** plays during an institutional action. Example: an individual acting on behalf of the organization performs this function in formal actions, such as creating an archival record or grouping them into compound records.

The «historicalRole» stereotype provides the conceptual foundation for modeling the ***Organization Member Actor*** class, referring to the roles assumed by endurants during specific events that occurred in the past (Almeida; Falbo; Guizzardi, 2019). This idea is summed up by the «historicalRole» stereotype, which is linked to an event through a «participation» relation. It means that an individual must have taken part in an event in order to play the role, which shows the historical context of its participation (Almeida; Falbo; Guizzardi, 2019). For instance, the «historicalRole» of a "Composer" is linked to the "Act of Composition" event through a «participation» relation, signifying that a person's participation in this event specifically qualifies them as a composer (Almeida; Falbo; Guizzardi, 2019).

The axiom **OrganizationMemberActor** $\sqsubseteq$ **OrganizationMember** establishes

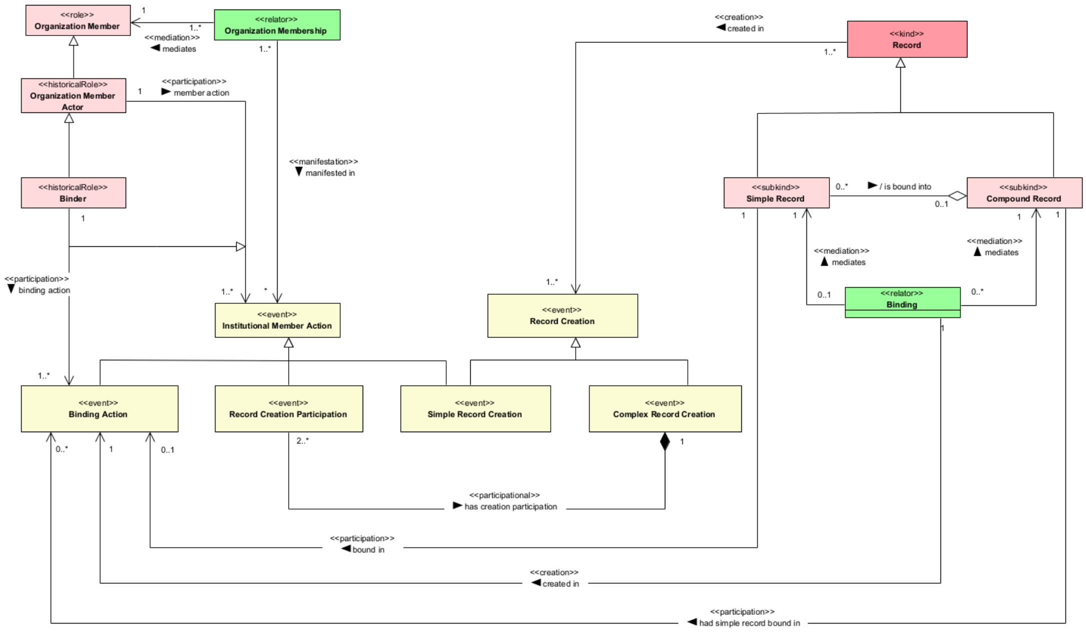

Figure 13 – Fragment of the Archival Record Ontology (ARO).

a specialization relationship, specifying that every ***Organization Member Actor*** is an ***Organization Member***, but not every ***Organization Member*** is necessarily an ***Organization Member Actor***. This implies that ***Organization Member Actor*** represents a specific subset of organization members who perform formal actions in the institutional context.

In addition, the ***Institutional Member Action*** class in the diagram of Figure 13 represents formal actions carried out by an ***Organization Member Actor***, aiming to manifest actions generated through an ***Organization Membership***, whose purpose is to ensure that record production occurs within an organizational context, enabling traceability and contributing to the authenticity of archival records.

The «event» stereotype provides the conceptual foundation for the ***Institutional Member Action*** class, recording formal actions taken by an organization member during an activity. This guarantees that these actions are documented and formally recognized as part of the organization's processes. Institutional.

The axiom $\mathbf{OrganizationMemberActor} \sqsubseteq \exists\ \mathbf{memberAction.InstitutionalMemberAction}$ states that every ***Organization Member Actor*** is associated with at least one ***Institutional Member Action*** through the specific object property **memberAction**, which is a specialization of the general property **participatedIn**. This «participation» relation reflects the nature of the ***Organization Member Actor*** role and this relation performs a formal role within a specific institutional action. The axiom $\mathbf{InstitutionalMemberAction} \sqsubseteq \exists\ \mathbf{manifestedIn}^{-}\ \mathbf{OrganizationMembership}$ ensures that institutional actions carried out by ***Organization Member Actor*** are only valid if anchored in an organizational membership that legitimizes them. Meanwhile, the axiom **InstitutionalMemberAction** $\sqsubseteq =$ **participatedIn OrganizationMember-Actor** warrants that each formal action within the organizational context is performed by a single ***Organization Member Actor***. Practically, this imposes a structure in the ontology where each institutional action has a clearly defined responsible party, in order to remove ambiguities regarding who participated in or executed the action. Reflecting the main attributes of archival records, these links are essential for traceability and accountability in record management, as they ensure that each institutional action has an assigned agent, which is a key element in supporting the authenticity of archival records in an organizational environment.

Moving on, the next class ***Record Creation***, in the diagram of Figure 13, represents the formal event in which an archival record is created, establishing the start of its existence in the organization. Modeling the temporal occurrence in which an archival record is created allows the start of its cycle in an organizational context to be captured.

The stereotype «event» provides the conceptual foundation for the class ***Record Creation***, since the action of creating an archival record has a specific duration (beginning and end), characterizing it as a single event delimited in time. The precise recording of the creation date of an archival record also contributes to characteristics such as authenticity and fixity by demonstrating a temporal record with rigorous formality.

The axiom **Record** $\sqsubseteq \exists$ **CreatedIn.RecordCreation** specifies that each ***Record*** is associated with a ***Record Creation*** event, guaranteeing that every archival record has a registered origin, thereby guaranteeing the traceability of its creation and authenticity. This means that each ***Record*** needs a specific creation event (***Record Creation***) to establish its existence. This axiom also contributes to the intrinsic characteristics of the archival record, such as its organic nature, by establishing a connection that validates the record's reflection of an institutional action.

Additionally, the ***Simple Record Creation*** class, in the diagram of Figure 13, represents a specialization of ***Record Creation***, depicting the event in which a simple record is created. Examples of simple records in an organization include circulars, memos, statements, letters, certificates, attestations, and receipts. They are archival records that have a simple structure and are created to serve specific transactions or communications, without the need for elaborate construction processes or input from multiple agents.

By following the «event» stereotype, the class allows the moment of creation of simple archival records to be recorded with the same level of detail and formality as any other institutional action. This provides a temporal traceability that is fundamental to the authenticity of archival records.

The axiom ***SimpleRecordCreation*** $\sqsubseteq$ ***RecordCreation*** states that the ***Simple Record Creation*** class is a subclass of ***Record Creation***. This means that every instance of ***Simple Record Creation*** is also an instance of ***Record Creation***, inheriting all the characteristics and restrictions applicable to record creation. In addition, the axiom ***SimpleRecordCreation*** $\sqsubseteq$ ***InstitutionalMemberAction*** defines that the ***SimpleRecord Creation*** class is also a subclass of ***Institutional Member Action***, inheriting all the properties and characteristics attributed to institutional actions carried out by members of the organization. It declares that the creation of simple records is not only a creation event, but it is a formal institutional action as well, carried out by members of the organization within the scope of their functions.

In addition, the ***Complex Record Creation*** class, in the diagram in Figure 13, is also a specialization of ***Record Creation***, representing the event in which a complex archival record is created. Examples in an organization include: administrative processes, dossiers, patient records, and strategic plans. These are records that require a more elaborate structure, are usually created to meet complex institutional processes, and often require input from multiple agents or review stages.

By following the «event» stereotype, the class allows the moment of creation of complex archival records to be recorded with the same level of detail and formality as any other institutional action. This provides temporal traceability, which is fundamental to the authenticity of archival records.

As with the ***Simple Record Creation*** class, there are also axioms to guarantee the consistency of the ***Complex Record Creation*** class. The axiom **ComplexRecordCreation** $\sqsubseteq$ **RecordCreation** establishes that the ***Complex Record Creation*** class is a subclass of ***Record Creation***, defining that every instance of ***Complex Record Creation*** is also an instance of ***Record Creation***, inheriting all the characteristics and restrictions attributed to record creation.

The ***Record Creation Participation*** class in the diagram of Figure 13 represents the specialization of an ***Institutional Member Action*** that takes part in the creation of complex archival records. There are archival records that require an elaborated creation process, such as, for example, the construction of a strategic plan, which requires the collaboration of multiple agents. Thus, through ***Record Creation Participation***, it is possible to record each participation action in the process of constructing the complex archival record. The «event» stereotype provides the conceptual foundation for the ***Record Creation Participation*** class, allowing the temporal recording of the action of participating in the creation of an archival record.

The axiom ***RecordCreationParticipation*** $\sqsubseteq$ ***InstitutionalMemberAction*** establishes that every instance of ***Record Creation Participation*** is a subclass of ***Institutional Member Action***, which ensures that every archival record creation action is also considered an institutional action. In other words, it guarantees that archival record creation does not occur in isolation, but within a context of institutional actions.

Moving on, the next class ***Binding Action*** in the diagram of Figure 13 represents the formal event of grouping archival records, where records are joined together into a coherent set. It captures the intentional action of joining a ***Simple Record*** to a ***Compound Record***, as well as joining a ***Simple Record*** to another ***Simple Record*** to form a new ***Compound Record***.

In the context of ***Binding Action***, the stereotype «event» provides the conceptual foundation for capturing the moment when archival records are joined into a formally recognized cohesive set within institutional actions. Example: the process of making purchases in order to supply a storeroom.

In addition, the ***Binder*** class in the diagram of Figure 13 represents the role played by an ***Organization Member Actor*** who performs the action of binding archival records. Example: in a purchasing process, several ***Simple Records***, such as the purchase order, price survey, senior management approval, and purchase result, are organized into a unified set. Although autonomous, these archival records are the result of the same institutional action.

The stereotype «historicalRole» provides the conceptual foundation for the ***Binder*** class, as it is used to model the role of an entity when performing an action of binding archival records. This role is maintained exclusively during the period in which the event takes place.

The axiom ***Binder*** $\sqsubseteq$ ***OrganizationMemberActor*** indicates that ***Binder*** is a subclass of ***Organization Member Actor***, since only formal members of an organization can assume the role of ***Binder*** in institutional actions. The axiom ***Binder*** $\sqsubseteq \exists$ ***participatedIn BindingAction*** defines that ***Binder*** is related to ***Binding Action*** events; thus, the role of binder is identified when some kind of link between archival records occurs.

Then, the next axiom, ***BindingAction*** $\sqsubseteq$ ***InstitutionalMemberAction*** defines that the ***Binding Action*** class is a subclass of ***Institutional Member Action***, meaning that every ***Binding Action*** is considered a formal action within the institutional context, carried out by official members of the organization. It ensures that the grouping of records is not an isolated action, but an institutional activity with formal and recognized value, which corroborates the guarantee of authenticity of the archival record, as well as being an element that can guarantee fixity.

The final class to be discussed is the ***Binding*** class shown in the diagram of Figure 13, which represents the result of the action of binding archival records, where related archival records are formally linked to form a cohesive set. The purpose of this class is to mediate the binding of ***Simple Records*** to other ***Simple Records***, resulting in a ***Compound Record*** or ***Simple Records*** to a ***Compound Record***.

The stereotype «Relator» provides the conceptual basis for the ***Binding*** class, as this class acts as a mediator in the connection between archival records, which allows the bond between archival records to be formally stated.

The axiom ***Binding*** $\sqsubseteq$ ***Relator*** defines that the ***Binding*** class is a subclass of ***Relator***, it means that ***Binding*** acts as a mediator, establishing and sustaining the relationships between archival records in the set.

The elements shown in Figure 11, 12 and 13 form the basis of the ARO ontology, developed to represent the archival record. Each class and relationship described contributes to the semantic structure of the ontology, also seeking to guarantee intrinsic characteristics of the archival record, such as fixity, organicity, naturalness, uniqueness, authenticity and impartiality. Figure 14 shows a complete visualization of this ontology, consolidating the relationships and specializations between the components described.

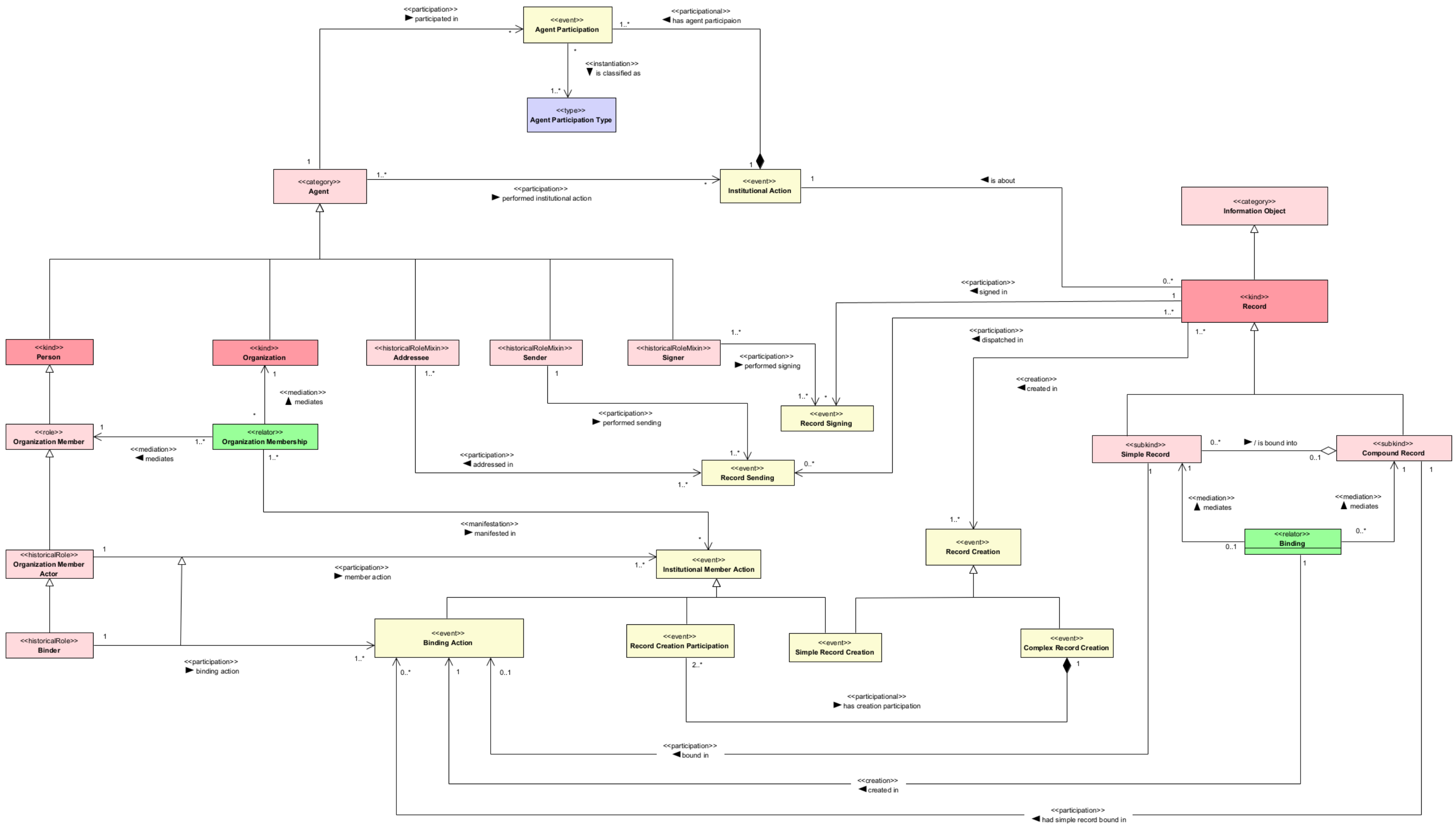

Figure 14 – Archival Record Ontology (ARO).

Additionally, Table 7 provides a summary of the ontology's classes along with their respective definitions, offering a detailed description of each element to enhance clarity and facilitate understanding of the structure and its elements.

Table 7 – Summary of ARO ontology classes and their definition

Class	**Stereotype**	**Definition**
Addressee	HistoricalRoleMixin	Represents the role assumed by an entity as the recipient of an archival record, capturing the function of receiving the record.
Agent	Category	An entity acting within the context of the archival record lifecycle, which can be a person or organization, participating in institutional actions.
Agent Participation	Event	The participation of an agent in institutional events, such as the creation, sending, or signing of archival records.
Agent Participation Type	Type	Classifies the different types of agent participation in phenomena underlying institutional actions.
Binder	HistoricalRole	A role performed by an organization member responsible for the action of grouping archival records.
Binding	Relator	The result of the action of grouping archival records, forming a cohesive and formally linked set.
Binding Action	Event	A formal event capturing the action of grouping archival records, joining simple records into compound records.
Complex Record Creation	Event	Represents the event in which a complex archival record is created.
Compound Record	Subkind	A record composed of multiple related simple records, organized into a cohesive set.
Information Object	Category	A combination of a data object and its associated representation information, allowing its correct interpretation and long-term preservation.

Continued on the next page

Class	Stereotype	Stereotype
Institutional Action	Event	A formal action carried out by an agent, such as the creation, signing, or processing of records.
Institutional Member Action	Event	An action executed by an organization member, reflecting formal activities contributing to record production.
Organization	Kind	A collective entity acting as an intentional agent in organizational contexts, participating in institutional processes.
Organization Member	Role	A role assumed by a person as a member of an organization, with specific functions in the context of documentary activities.
Organization Member Actor	HistoricalRoleMixin	A role of an organization member during an institutional action, formally representing the organization.
Organization MemberShip	Relator	A mediating entity connecting the organization member to the organization, formalizing this relationship.
Person	Kind	A class representing individuals involved in record-related activities, acting as agents in institutional actions.
Record	Kind	An archival record produced or received during an activity, serving as evidence of the actions performed.
Record Creation	Event	An event marking the formal creation of an archival record, establishing the beginning of its existence.
Record Creation Participation	Event	The participation of agents in record creation, documenting their contributions to the process.
Record Sending	Event	The formal act of transmitting an archival record from one agent to another.
Record Signing	Event	An event of signing an archival record.
Sender	HistoricalRoleMixin	The role of an agent who performs the sending of an archival record.

Continued on the next page

Class	Stereotype	Stereotype
Signer	HistoricalRoleMixin	The role of an agent who signs an archival record.
Simple Record	Subkind	A simple record, consisting of a single informational unit, such as a letter or memorandum.
Simple Record Creation	Event	An event of creating a simple record, documenting its moment of origin.

End of table

4.3 ARO and the structured resources identified with CLeAR

The modeling of the Archival Record Ontology (ARO) was influenced by the structured resources identified through the CLeAR approach. The application of CLeAR made it possible to identify structured resources such as ontologies, vocabularies and conceptual models that provided support for modeling archival records and their contextual relationships. The aim of this section is to discuss how ARO relates to these structured resources and what their main influences have been. Thus, to facilitate the analysis of the relationship between ARO and the structured resources identified, these have been grouped based on their similarity of contribution.

Life Cycle and Contextual Relationships of Archival Records

CIDOC-CRM contributes by presenting an event-based structure to model the trajectory of documents linked to cultural heritage over time. RiC-O reinforces the need to structure the relationships between archival records, agents and administrative functions, promoting a systematic organization of archival records. In addition, BDA introduces the need to structure archival records that were born digital, with special emphasis on their life cycle. GAO, designed to describe government archives in China, is concerned with the contextual and hierarchical relationships between archival records and the public organization in which they exist. Based on the CIDOC-CRM's event-based structure, RiC-O's systematic organization of archival relationships, BDA's focus on born-digital records, and GAO's contextualization within public organizations, ARO seeks to consolidate a structure for archival records concerned with their relationships and their life cycle, by recording the administrative events that took place during their trajectory.

Digital Preservation and Interoperability

PREMIS OWL is a structured resource whose principles align with ARO's objectives, especially in the domain of digital preservation. Although it has not been fully incorporated, its approach to describing fixity, authenticity and provenance reinforces the concerns modeled implicitly in ARO's classes, stereotypes, and axioms. Future work could explore closer integration of PREMIS OWL to improve ARO's representation of digital preservation metadata. The contribution of OAI-ORE is to show that an archival record needs to be reusable, because the same archival record can be present in different contexts without losing its identity. It also reinforced ARO's concern with digital preservation and interoperability between repositories. GAO helped to strengthen the commitment with semantic interoperability by emphasizing the need for an ontological model that allows archival records to be integrated into different systems and contexts. GAO should be highlighted because, like CLeAR, it is concerned with structuring information related to archival records based on the 5W1H method (Who, What, When, Where, Why, How). This structuring makes it possible to capture who produced the record, what it represents, when it was created, where it was produced, why it was generated and how it fits into the organizational context, in order to aid semantic interoperability between Chinese government archive systems. ARO used this way of structuring information, from the application of the CLeAR approach, to answer the integration questions. Table 5 presented how the aspects related to the integration questions are grouped using the 5W1H method.

Events

Although ARO already integrates concepts from established ontologies such as RiC-O, CIDOC-CRM and PREMIS, its modeling has been enriched with contributions from structured resources aimed at representing events and historical context, such as ArCo, ARKIVO and HERO. These models provide complementary approaches to describing historical, cultural and administrative events, allowing ARO to more accurately align archival records with their future historical relevance. Event modeling is a central element in representing the lifecycle of archival records for ARO, allowing it to capture administrative actions, institutional interactions and transformations of archival records over time. For this reason, ARO is based on UFO-B, which provides the theoretical basis for this modeling by adopting the concept of perdurants, making it possible to represent changes in the state of archival records over time.

Provenance

According to SAA (2005-2023b), provenance is a “fundamental principle of archives, referring to the individual, family, or organization that created or received the items in a collection”. Thus, “the principle of provenance or the *respect des fonds* dictates that records of different origins (provenance) be kept separate to preserve their context” (SAA, 2005-2023b). The ARO ontology incorporates provenance as one of the central axes of its modeling, ensuring that organizational actions, interactions between agents and modifications to archival records are represented in a structured way. In the context of ARO, provenance modeling is essential to guarantee the authenticity and fixity of archival records. This is why ARO seeks the contributions of the structured resources PROV-O, BBCPROV, OPMO and PAV. In particular, PROV-O's contribution to the ARO development process should be highlighted. PROV-O is an ontological model developed by the World Wide Web Consortium (W3C) to represent provenance on the Semantic Web. As an example of how PROV-O concepts contributed to the development of ARO, the class ***Archival Record*** in ARO corresponds to `prov:Entity`, administrative events affecting archival records, represented in ARO by the ***Record Creation***, align with `prov:Activity`, and the agents responsible for the creation or transmission of records, modeled in ARO bye ***Agent*** class, correspond to `prov:Agent`. With regard to the structured resources BBCPROV, OPMO, PAV, it was identified that: (i) the BBCPROV ontology reinforces the need for the organization to guarantee the provenance of its digital collections, (ii) the OPMO ontology is useful because it is concerned with representing provenance within institutional environments, enabling provenance information to be exchanged between systems, and (iii) the PAV ontology provides a solution for versioning digital objects for the ARO, so if ARO evolves and starts to control the stages of archival record creation, from draft to finished record, it can look to the PAV for contributions on how to version the archival record, guaranteeing provenance and authorship.

The analysis of the structured resources identified by CLeAR has played an important role in the development of the ARO, as it has enabled the identification of points of convergence, supported the conceptual decisions of the ARO, and provided perspectives for its evolution. The identification and study of structured resources highlighted the importance of representing events, agents and activities, increasing its capacity to represent the life cycle of archival records in a semantic, interoperable way and in line with international standards.

4.4 Differences between ARO, RiC-O and CIDOC-CRM

Different approaches have been developed, or are under development, to represent "documents" or "archival records" and their contextual relationships, including CIDOC-CRM, which focuses on cultural heritage, and RiC-O, developed by ICA to structure the description of archival records. However, these models do not fully meet the specific needs of digital archival records in public administration systems. So, in response to this gap, the Archival Records Ontology (ARO) was proposed to represent digital archival records and their life cycle, guaranteeing their semantic interoperability in government information systems.

ARO is a new model with the potential to be adopted in the management of electronic archival records especially in environments that seek to work with modern approaches such as semantic interoperability. RiC-O also has significant potential in this context and has been progressively adopted in archives as a promising model for interoperability between archival repositories. However, as it is relatively new, its adoption is still expanding. CIDOC-CRM, on the other hand, has been widely adopted in the museum sector, as its primary focus is on cultural heritage. While it can be applied to archival records, such use requires adaptations, since its original scope does not cover the management of records within public administration systems.

The main difference between ARO, RiC-O and CIDOC-CRM lies in their ontological foundation. ARO uses UFO, which provides greater semantic rigor in defining the concepts of the domain being modeled. In contrast, RiC-O uses modeling based on semantic networks and the description of relationships between archival records and agents, without using a formal foundational ontology. Its development was influenced by several well-known ontologies and conceptual models, such as CIDOC-CRM, PROV-O, FOAF, Dublin Core, among others. Its main objective is to structure the contextual relationships between records and their producers, guaranteeing interoperability between different archival repositories, but not between public administration business systems. CIDOC-CRM, in turn, does not have an explicit ontology as its foundation. Adopting a conceptual model, it uses an event-oriented conceptual structure to describe relationships in the cultural heritage domain and their interactions over time. Although it is used in the museum sector, the CIDOC-CRM can also be used in archives, but it does not explain the main characteristics of an archival record, such as fixity, organic nature, naturalness, uniqueness, authenticity, and impartiality. On the other hand, the ARO was developed with a specific focus on these characteristics, aiming to incorporate them into its ontological structure.

ARO aims to contribute to a detailed modeling of the archival record's life cycle, which takes place within an organization, allowing it to be represented from its creation, transmission, handling to its final destination, by capturing the administrative acts that impact it, as well as representing the participation of agents and their actions. However, RiC-O incorporates elements of the document's life cycle, but does not detail the administrative events associated with its processing and preservation, as its focus is on describing archival records and their contextual relationships, without a formal structure to represent administrative processes existing in organizations. In contrast, the CIDOC-CRM deals indirectly with the record life cycle, emphasizing the description of historical events of cultural objects, without covering document management in the archival context.

Digital preservation is an essential aspect of archival record management, especially given the challenges imposed by technological obsolescence and the need to guarantee the authenticity and integrity of archival records. In this context, both ARO and RiC-O can be aligned with standards such as PREMIS and Dublin Core to support digital preservation. However, due to their different areas of application, the extent and manner of this alignment may vary. On the other hand, CIDOC-CRM was not originally developed for this purpose, requiring extensions to deal specifically with digital preservation.

The comparison between ARO, RiC-O and CIDOC-CRM highlights that each structured resource was developed to meet different needs in the field of document[1] management. While RiC-O focuses on the description of archival records and their contextual relationships, and CIDOC-CRM on the modelling of cultural heritage, ARO sets itself apart by offering a semantically structured approach that enhances the clarity and consistency in the modeling of archival records, adapting to the needs of archival records management in an organizational environment. ARO can complement and expand the resources of RiC-O, taking advantage of its structure to represent archival relationships and incorporating essential concepts for managing archival records whose life cycle has not yet reached its final destination phase. In reference to CIDOC-CRM, ARO can act as a specialized model for archival records, adapting its event-oriented approach to represent not only the provenance and contextual relationships of archival records, but also the administrative processes and formal requirements of archival records management from an archival perspective.

4.5 Summary

Chapter 4 presented the ontological model developed to represent archival records, known as the Archival Records Ontology (ARO). It began by describing the development process of the ontology, followed by a detailed description of its classes, stereotypes, and axioms. Each component was analyzed in terms of its role within the model, ensuring a structured and semantically consistent representation of archival records. The chapter also examined how the ontology captures key archival principles, such as provenance, authenticity, and fixity, reinforcing its applicability in records management and digital preservation. Additionally, the logical constraints embedded in the ontology were discussed, demonstrating how axioms and relationships contribute to maintaining the integrity of the model. In addition, the ontology was analyzed in order to demonstrate how ARO relates to the structured resources identified in the CLeAR approach, and a comparative analysis was carried out between ARO, RiC-O and CIDOC-CRM, highlighting their conceptual and structural differences, as well as their different purposes and applications. In addition, the ontology was analyzed to demonstrate how ARO relates to the structured resources identified through the CLeAR approach. Finishing with a comparative analysis of ARO, RiC-O and CIDOC-CRM, highlighting their conceptual and structural differences, as well as their distinct purposes and applications.

[1] The term "document" is adopted in this context to cover the different approaches of the resources analyzed here. While ARO uses the term "archival record", CIDOC-CRM uses "document" in a broader sense, related to cultural heritage and museum documentation. RiC-O, however, structures the description of archival records using specific terms such as "Record", "RecordSet", and "RecordResource".

5 Ontology Verification

This chapter presents the ontology verification process, which used Protégé software to enable interactive testing of ontologies. Verification was carried out using SPARQL queries developed to assess the ontology's ability to correctly answer integration questions. Before running the SPARQL queries, the verification environment had to be prepared, including the instantiation of the ontological concepts and the insertion of representative data.

5.1 From Conceptual Modeling to Operational Ontology

The SABiO methodology conducts the development of two ontology types: reference ontology and operational ontology (Falbo, 2014). "By domain reference ontology we mean a domain ontology that is built with the goal of making the best possible description of the domain" (Falbo, 2014). "Contrary to reference ontologies, operational ontologies are designed with the focus on guaranteeing desirable computational properties" (Falbo, 2014).

Chapter 4 presented Archival Records Ontology (ARO), a reference ontology, which constitutes a specific type of conceptual model (Falbo, 2014) and aims to provide a well-founded conceptual representation of the archival record domain. However, the reference ontology needs to be transformed into a specific machine-readable ontology language (e.g. OWL) (Falbo, 2014).

It is important to note that during the development of the conceptual model, continuous validations were carried out to ensure the consistency and ontological conformity of the reference ontology. This process included systematic verification of the modeled structures and relationships, using the plugin's "Check Diagram" functionality to identify possible inconsistencies. Figure 15 illustrates the resources used for the checking process, highlighting the "Check Diagram" functionality and the window showing the result obtained after the check. The model must be checked to ensure there are no errors in the combination of OntoUML stereotypes.

The next step was to export the ontology to an operational format. To do this, the OntoUML plugin to Visual Paradigm was used, with its "Export to gUFO" functionality. The plugin automatically generates a file in Turtle (TTL) format, enabling it to be used in tools such as Protégé for automated queries and inferences. Figure 16 shows the resources used, highlighting the "Export from gUFO" functionality and the confirmation of the successful export.

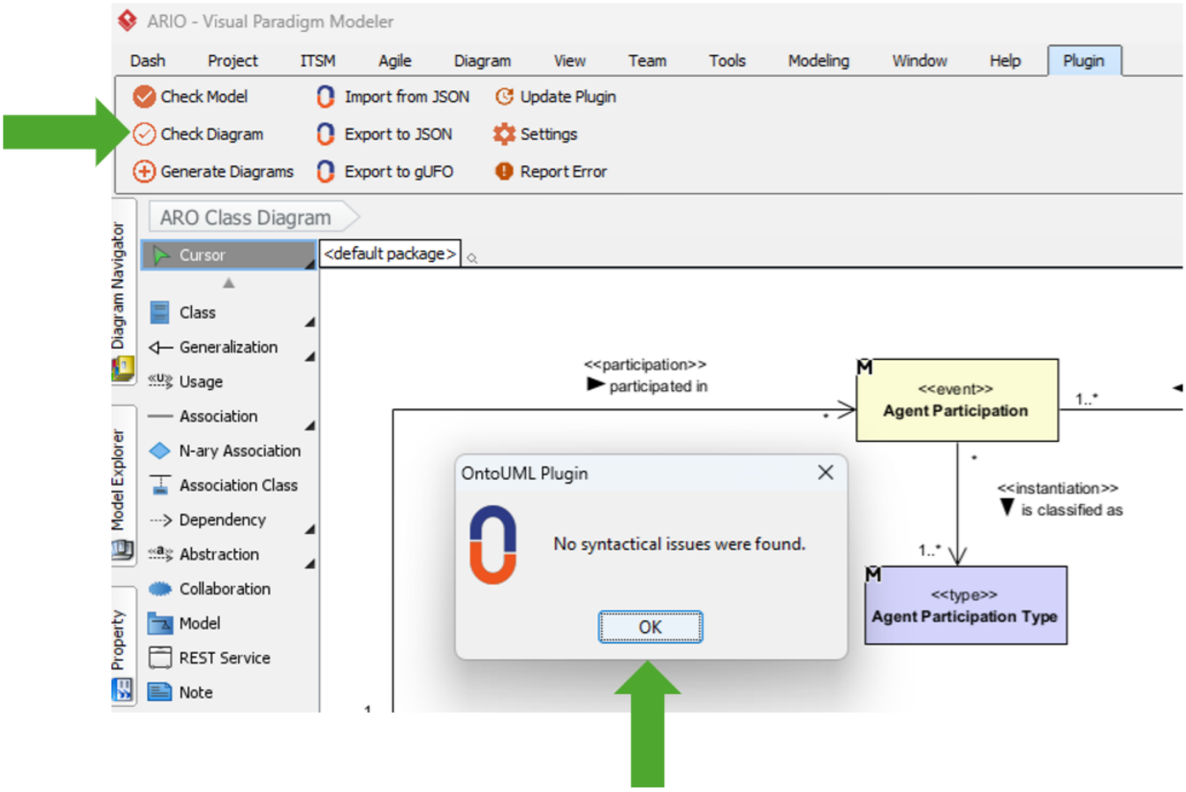

Figure 15 – Diagram checking process in Visual Paradigm.

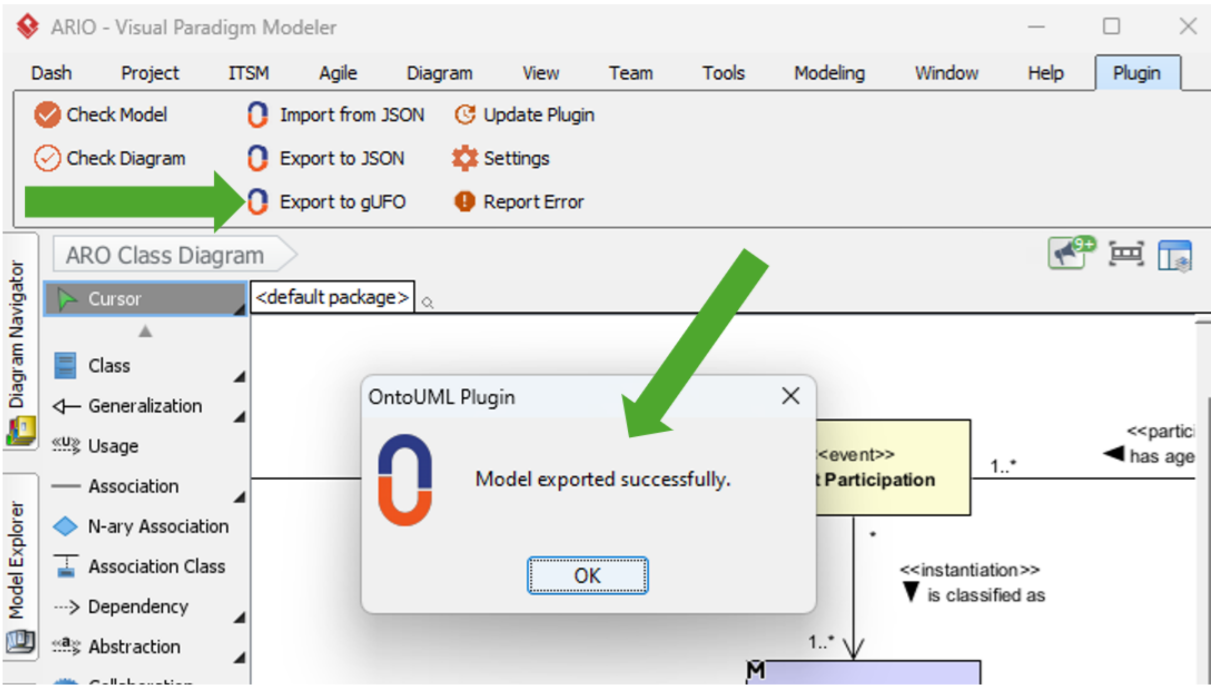

Figure 16 – Conversion of the reference ontology into an operational ontology using Visual Paradigm's gUFO export feature.

Figure 17 shows the correspondence between conceptual modeling and operational ontology, illustrating an excerpt from the OntoUML diagram and its equivalent representation in the Turtle (TTL) file. The left side, which shows part of the OntoUML diagram, shows the hierarchical structure of the ***Record*** class, with its ***Simple Record*** and ***Compound Record*** subclasses. The right side shows part of the content of the TTL file, which uses a formal textual language to present this same hierarchy using the rdfs:subClassOf property, ensuring that `:SimpleRecord` and `:CompoundRecord` are subclasses of `:Record`, while `:Record` is a subclass of `:InformationObject`. The arrows indicate the correspondence between the elements of the conceptual diagram and their representations in the TTL file, demonstrating how the structure defined in the OntoUML model is preserved in the operational ontology.

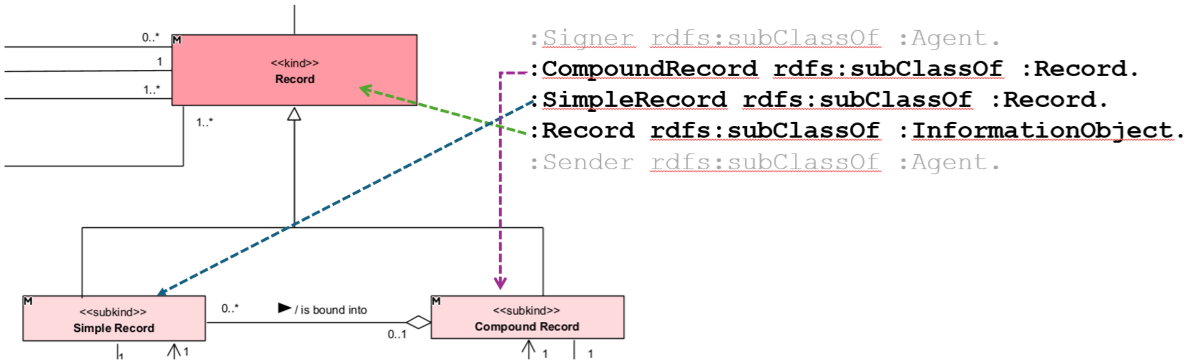

Figure 17 – Correspondence between a part of the OntoUML diagram and its representation in the Turtle file.

In addition to the classes and relationships modeled in the reference ontology, ARO incorporates a set of essential properties for characterizing archival records to assist in the verification process. Some of these properties come from recognized standards, such as Dublin Core, which is used in the context of digital preservation of archival records; other properties can also be incorporated to address the need to respond to integration questions identified in the modeling process. Thus, the transition from reference ontology to operational ontology involves not only exporting the conceptual structure, but also incorporating resources such as individuals, properties, axioms, in order to assist the verification process.

Although the conceptual model was developed in Visual Paradigm using OntoUML and exported using the gUFO plug-in, it is important to highlight that the **data properties** in the ontology — such as **recordID**, **title**, **creationDate**, among others — were defined directly within the Protégé environment and were not automatically derived from attributes in the OntoUML diagram. Although Visual Paradigm supports the definition of conceptual attributes in OntoUML classes and their subsequent export as **data properties**, this work opted to define such properties directly during the computational implementation phase. This methodological decision aimed to keep the conceptual modeling focused on the structural and relational elements essential to the domain, leaving the detailed specification of data properties to a later stage, where greater control over types, constraints, and annotations could be exercised within Protégé.

5.2 Preparing for Answering the Integration Questions

In the SABiO method, "the ontology test refers to dynamic verification and validation of the behavior of the operational ontology on a finite set of test cases, against the expected behavior regarding the competency questions" (Falbo, 2014). This process ensures that the ontology is able to correctly answer the questions that define its scope and usefulness in the domain, allowing a detailed assessment of its functionality and accuracy (Falbo, 2014). Verification is conducted by means of queries that represent real cases in the domain, ensuring that the ontological elements and structures implemented are logically consistent and sufficient to support the expected objectives (Falbo, 2014). Thus, the verification phase in the SABiO plays an essential role in guaranteeing that the ontology developed maintains internal coherence and aligns with the defined requirements in the context in which it will be applied (Falbo, 2014).

Campos (2019) describes that the Integration Questions (IQs), defined in the CLeAR approach, are analyzed and refined into Competency Questions (CQs) in the context of ontology engineering methodologies, such as SABiO. The process involves using the IQs to identify data sources and domain aspects, which are then decomposed into CQs to define the scope of the ontology and evaluate the ontology during its development (Campos, 2019; Campos *et al.*, 2020). This approach shows how the integration requirements addressed by the IQs are directly correlated with the ontological requirements captured by the CQs. Although the SABiO methodology uses competency questions to define the scope and functional requirements of the ontology, in the context of this research this methodology was adapted to answer integration questions specific to the modeled domain. The integration questions will serve as guidelines for verifying the ontology, ensuring that it is able to correctly answer the questions defined in Section 3.1.

The ontology verification process was carried out using the Protégé tool[1], which allows ontologies to be tested interactively. Within this environment, SPARQL[2] queries were implemented to verify the ontology's ability to correctly answer integration questions. The following description of the verification process adheres to the order of integration questions outlined in Table 1, with the complete source code for all SPARQL queries provided in Appendix B.

However, before starting the verification of the ontology through SPARQL queries,

[1] *Protégé* is an open-source ontology editor used for modeling ontologies based on OWL. Available at: <https://protege.stanford.edu/>.

[2] *SPARQL* is a query language and protocol for querying and manipulating data stored in Resource Description Framework (RDF) format. More information available at: <https://www.w3.org/TR/sparql12-query/>.

it was necessary to prepare the environment for this task. This process involved the instantiation of ontological concepts, ensuring that the ontology was properly populated with representative data. To guarantee the completeness of the verification, instances were created for all the classes necessary to answer the integration questions, including the definition of attributes, the insertion of data, and the establishment of relationships between the instances, assuring the coherence of the ontological model and its adherence to the domain of archival records.

The instantiation was performed in the Protégé software, as illustrated in Figure 18, which shows the **Record1** instance, belonging to the ***Record*** class. As can be seen in the figure, the instance has several associated properties, including identifier (ID), file format, title, conditions, and access restrictions, as well as semantic relationships with other individuals, such as record creation, sending, and signing events.

In the process of verifying the ontology through SPARQL queries, the Dublin Core Metadata Initiative standard was used, represented by the prefix dc: to enrich the information retrieval. Although Dublin Core has not been incorporated into ontology modeling, it has been adopted in SPARQL queries to ensure semantic interoperability in information retrieval. The use of properties such as dc:title, dc:creator, dc:date and dc:format allowed to structure the queries in line with widely recognized standards for metadata description, which are used in digital preservation systems and metadata description.

The data entered were designed to simulate the structure and characteristics of a real system, guaranteeing that the ontology could be tested in a representative manner. However, no real data was used, considering information security and sensitive data protection guidelines, so the adoption of fictitious data, but structurally like that of a real environment, assures that the verification of the ontology occurs without compromising the confidentiality of institutional information.

5.3 Answering the Integration Questions

Once the environment preparation activities had been completed, including ontology instantiation and insertion of the necessary data, construction of the SPARQL queries began. The formulation of the queries followed a systematic approach, ensuring that each integration question was represented by exploring its classes, properties and semantic relationships. The SPARQL queries developed for the verification are presented below, along with their respective explanations and interpretations, following the order of the integration questions defined in Chapter 3.

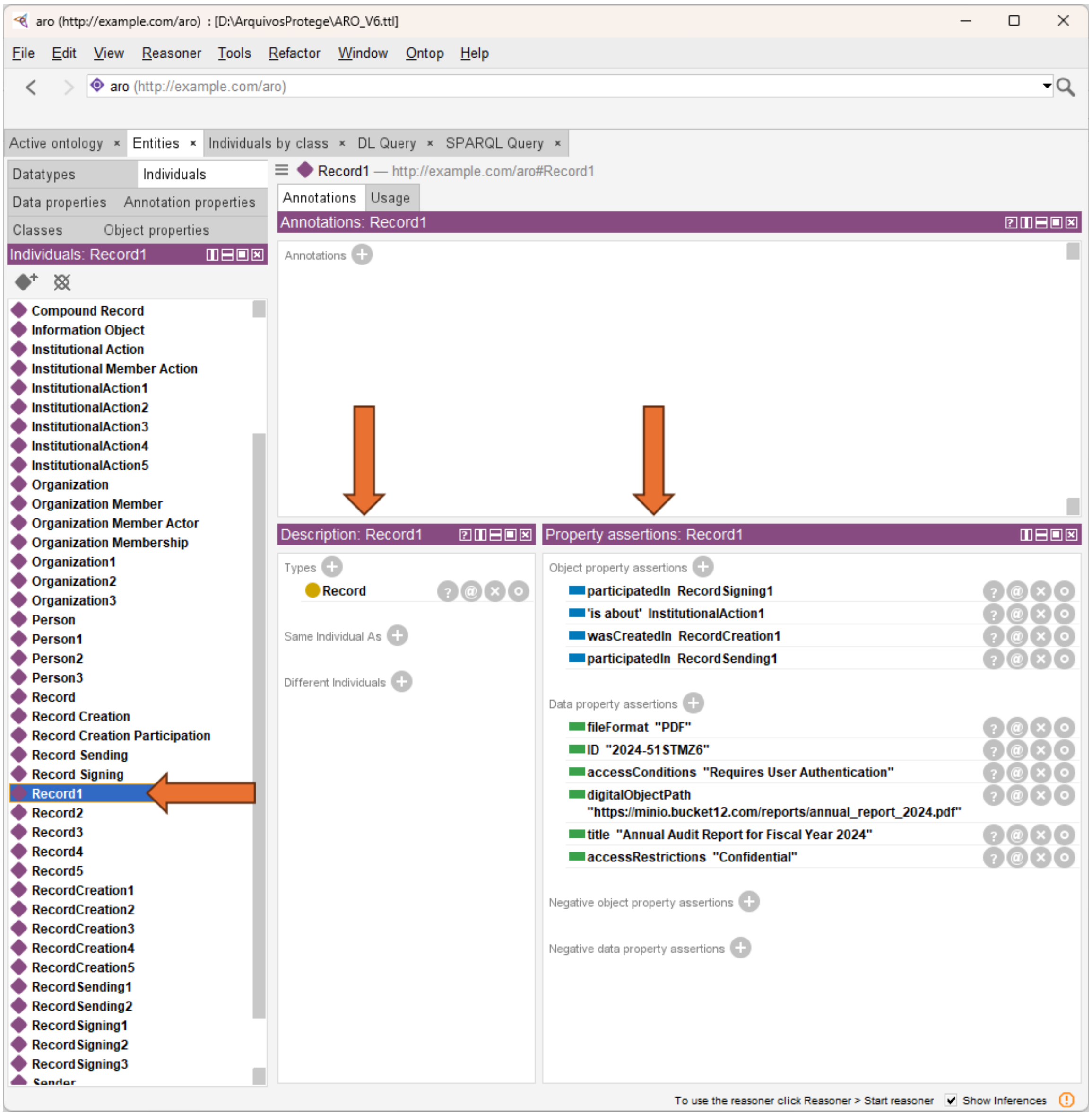

Figure 18 – Instantiation of the ***Record*** class in Protégé, demonstrating its attributes and relationships.

IQ01 - Who is the producer of the record?

Question IQ01 aims to identify the organization or person responsible for creating or accumulating records; this question seeks to ensure the traceability and authenticity of the Archive Record. This not only reinforces the fundamental principle of Provenance in Archival Science, but also ensures the precise identification and coherent organization of records of the same origin into "fonds[3]".

Using SPARQL, a query was developed to retrieve archival records (`:Record`) and

[3] The term "fonds" (or "archival fonds") refers to the entire body of records created or accumulated by an individual, family, or corporate body as a result of its functions or activities, representing the highest-level aggregation in archival organization and encompassing all records organically created, collected, and used by the creator in conducting their affairs (IP2, 2002-2007).

agents (`:Agent`) involved in institutional actions (`:InstitutionalAction`), extracting the titles of both records and agents and identifying each agent as either a person (`:Person`) or an organization (`:Organization`). To achieve this, the query uses the `dc:title` property to capture titles and `gufo:participatedIn` to link agents to institutional actions related to the records. The `STR` [4] function converts titles into strings, while the `OPTIONAL` and `BIND` clauses conditionally assign the agent type as "**Person**" or "**Organization**", ensuring that agents are accurately categorized without omitting incomplete data. The SPARQL query for IQ01 uses the prefixes `:`, `dc:`, and `gufo:` to reference, respectively, terms from the ARO ontology at URI `http://example.com/ARO#`, Dublin Core elements at URI (`http://purl.org/dc/elements/1.1/`), and the Lightweight Implementation of the UFO at URI (`http://purl.org/nemo/gufo#`). Figure 19 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

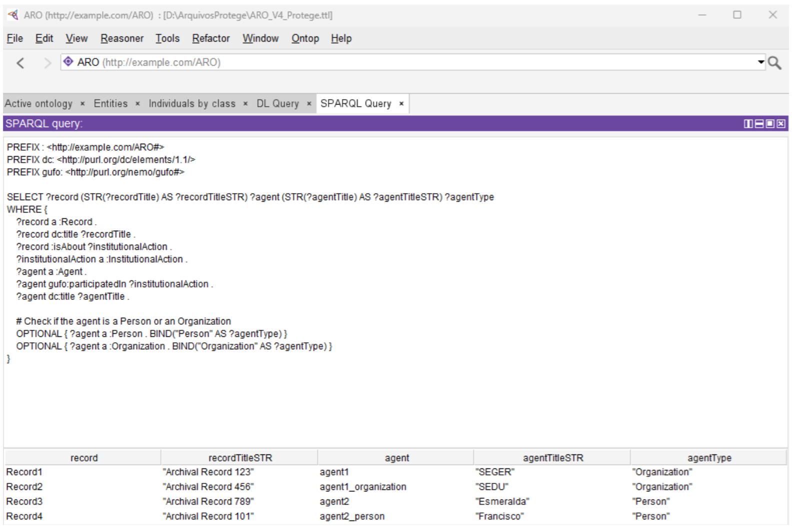

Figure 19 – SPARQL query for IQ01 verification.

[4] The `STR` function in SPARQL converts an IRI or a typed literal into a simple string, facilitating comparisons, filtering, and textual operations. This ensures compatibility in queries involving different RDF representations and data types (W3C, 2013).

IQ02 What is the date and place of production of the record?

Question IQ02 aims to identify the creation date of the archival record, the organizational system in which it was produced, and characteristics that contribute to its fixity and authenticity.

Using SPARQL, a query was developed to retrieve archival records (`:Record`) and their associated creation events (`:RecordCreation`), obtaining the record's identifier, creation date, and location. To accomplish this, the query uses the `:ID` property to capture the identifier, `gufo:wasCreatedIn` to associate the record with the creation event, and the properties `:creationDate` and `:creationLocation` to extract the date and location of creation. The `STR` function converts these values into strings, ensuring that the date and location information is consistently displayed in the query result.

The SPARQL query for IQ02 uses the prefixes `:` and `gufo:` to reference, respectively, terms from the ontology at URI `http://example.com/aro#` and the Lightweight Implementation of the UFO at URI (`http://purl.org/nemo/gufo#`). Figure 20 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

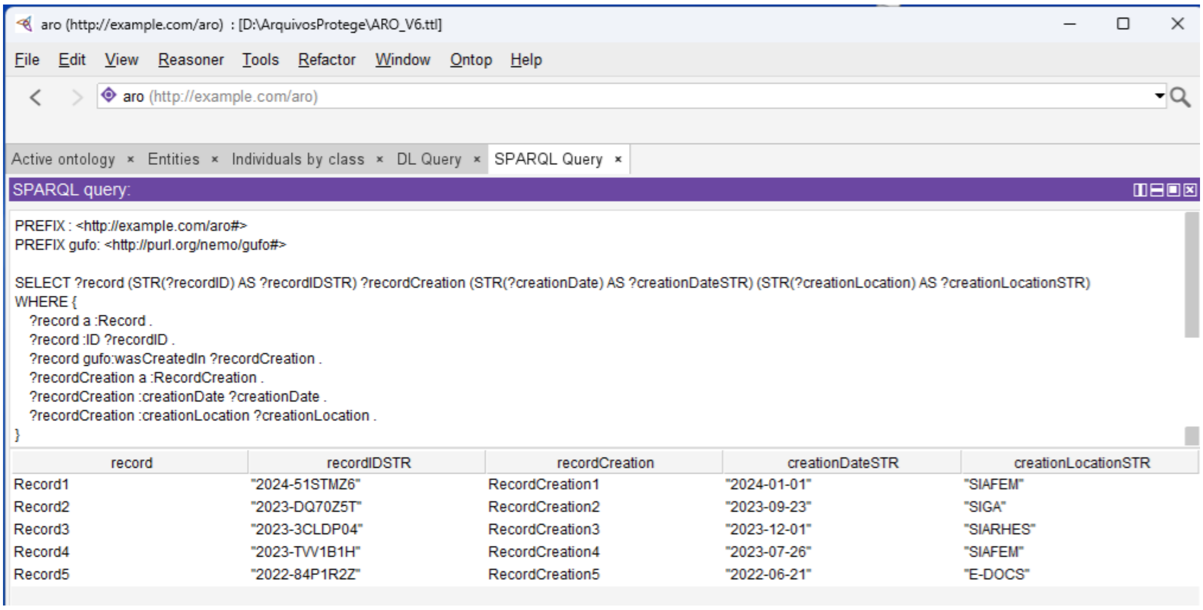

Figure 20 – SPARQL query for IQ02 verification.

IQ03 What is the administrative motivation for producing the record?, **IQ11 What is the classification scheme assigned to the record?**, and **IQ14 What activity gave origin to the record?**

Throughout the development of the ontology, it was observed that integration questions IQ03, IQ11 and IQ14 are closely related, as all these questions are connected to the institutional action of the organization responsible for producing the archival record. In the verfication carried out, the SPARQL query returns the institutional action linked to the archival record, which directly represents the answer to these questions, since the institutional action encapsulates the elements of administrative motivation, classification and source activity, offering a unified view of the context in which the archival record was created.

Using SPARQL, a query was developed to retrieve archival records (`:Record`) and their associated institutional actions (`:InstitutionalAction`), obtaining the record's identifier and the motivation for the institutional action. To achieve this, the query uses the `:ID` property to capture the unique identifier of the record, as well as the `:isAbout` and `:motivation` properties to associate the record with the institutional action and extract its motivation. The `STR` function converts the identifier and motivation to strings, ensuring consistent data presentation in the query result. The SPARQL query uses the prefix `:` to reference, from the ARO ontology at URI `http://example.com/aro#`. Figure 21 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

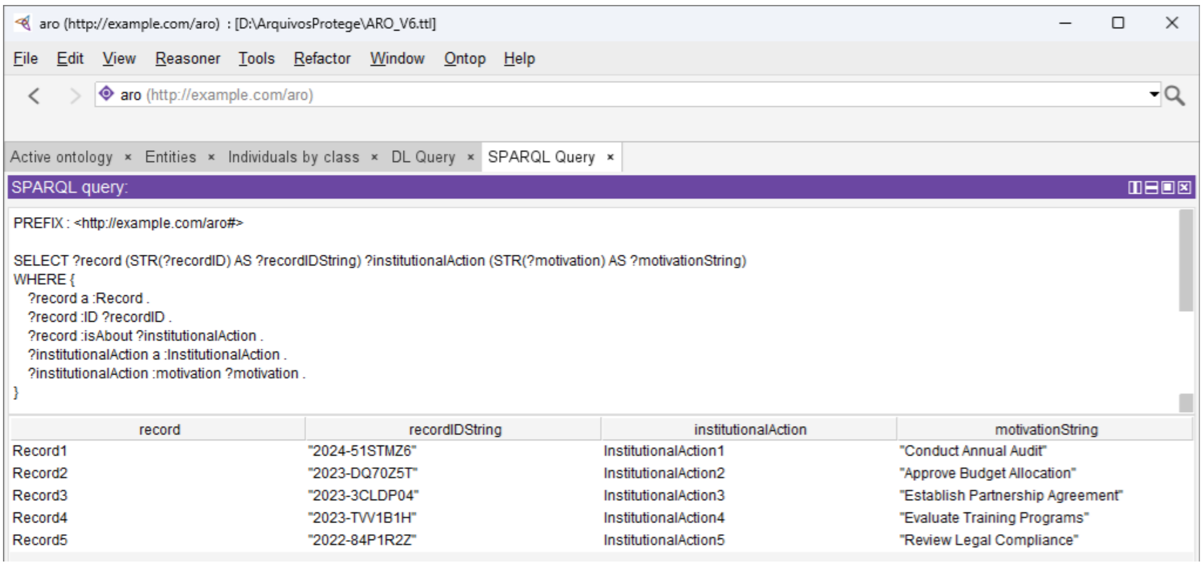

Figure 21 – SPARQL query for IQ03, IQ11 e IQ14 verification.

IQ04 What is the unique identifier of the record?

Question IQ04 addresses the existence of a unique identifier for each archival record; the presence of a unique identifier is indispensable in archival contexts, contributing in particular to the retrieval of the archival record.

Using SPARQL, a query was developed to retrieve archival records (`:Record`) and their respective unique identifiers. To achieve this, the query uses the `:ID` property to capture each record's identifier and the `STR` function to convert it into a string, ensuring consistency in data presentation in the query result. The SPARQL query for IQ04 uses the prefix `:` to reference, from the ARO ontology at URI `http://example.com/aro#`. Figure 22 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

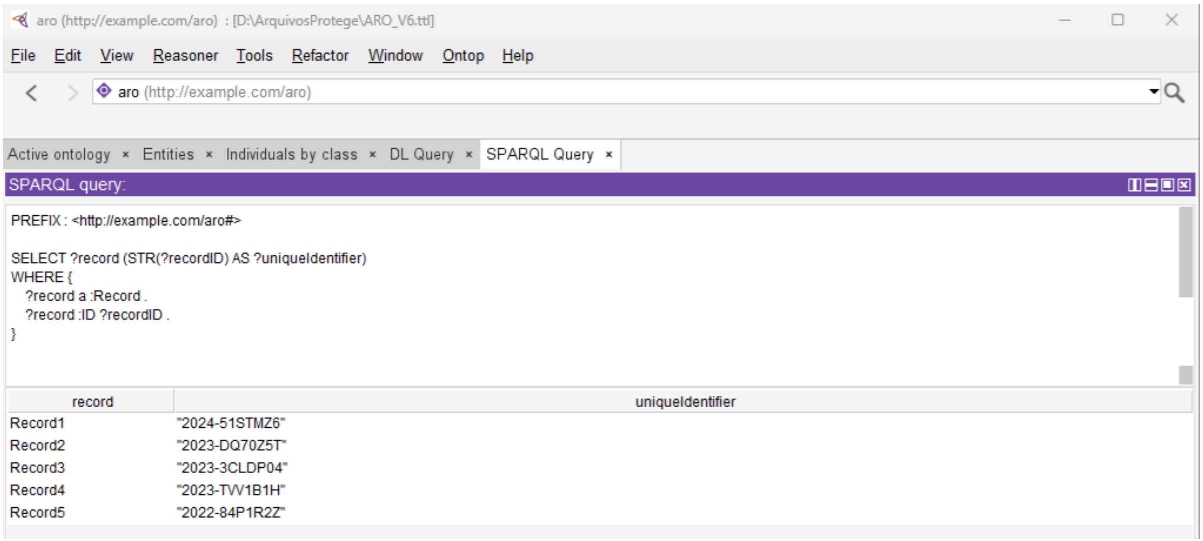

Figure 22 – SPARQL query for IQ04 verification.

IQ05 What is the title of the record?

Question IQ05 aims to identify whether the archival record has a title, as this information provides a short, identifiable description that also makes it easier to identify the archival record, helping with its retrieval.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), obtaining their unique identifier and title. To accomplish this, the query uses the `:ID` property to capture the identifier and `:title` to retrieve the record's title. The `STR` function converts both values to strings, ensuring consistency in data presentation within the query result. The SPARQL query for IQ05 uses the prefix `:` to reference, from the ARO ontology at URI `http://example.com/aro#`. Figure 23 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

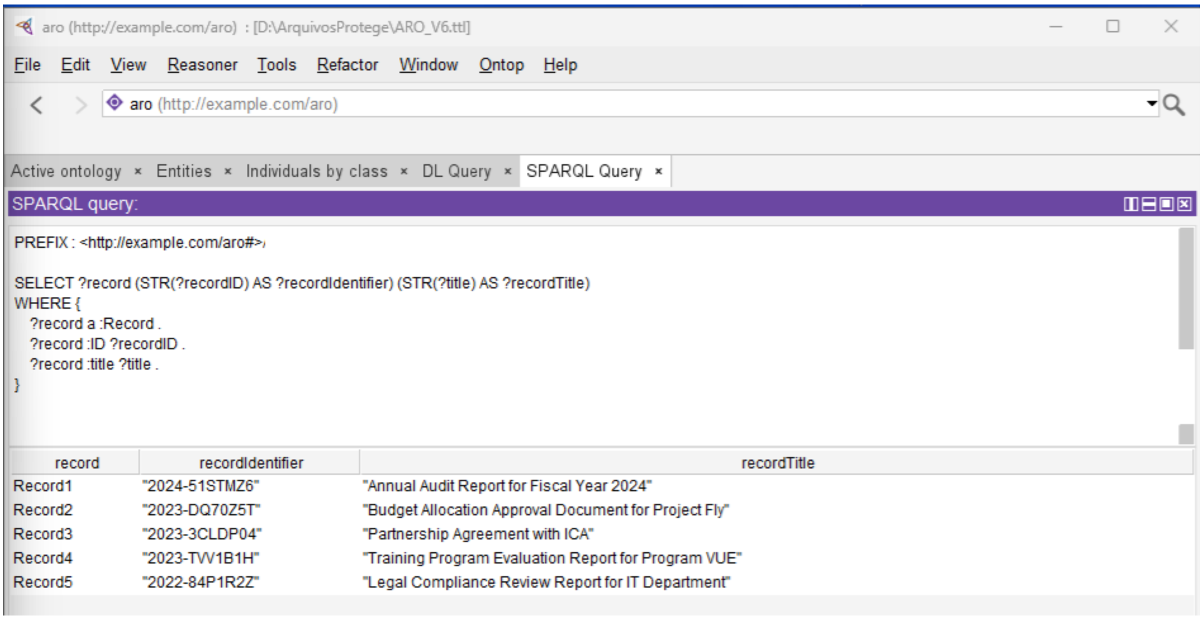

Figure 23 – SPARQL query for IQ05 verification.

IQ06 Who are the parties concerned?

Question IQ06 seeks to identify the agents (people or organizations) involved in the institutional action that led to the specific archival record. This information is crucial in an archival and administrative context, particularly when dealing with issues related to the General Data Protection Act.

Using SPARQL, a query was developed to retrieve archival records (`:Record`) and agents (`:Agent`) involved in institutional actions (`:InstitutionalAction`), as well as the record's identifier and title, the agent's name, and the motivation behind the institutional action. To accomplish this, the query uses the properties `:ID` to capture the record's identifier, `:name` for the agent's name, and `:motivation` for the action's motivation. The `STR` function converts these values into strings, ensuring consistent data presentation in the query result. The SPARQL query for IQ06 uses the prefixes `:` and `gufo:` to reference, respectively, terms from the ARO ontology at URI `http://example.com/aro#` and the Lightweight Implementation of UFO at URI (`http://purl.org/nemo/gufo#`). Figure 24 shows the Protégé IDE interface where the SPARQL query was executed, along with the result obtained.

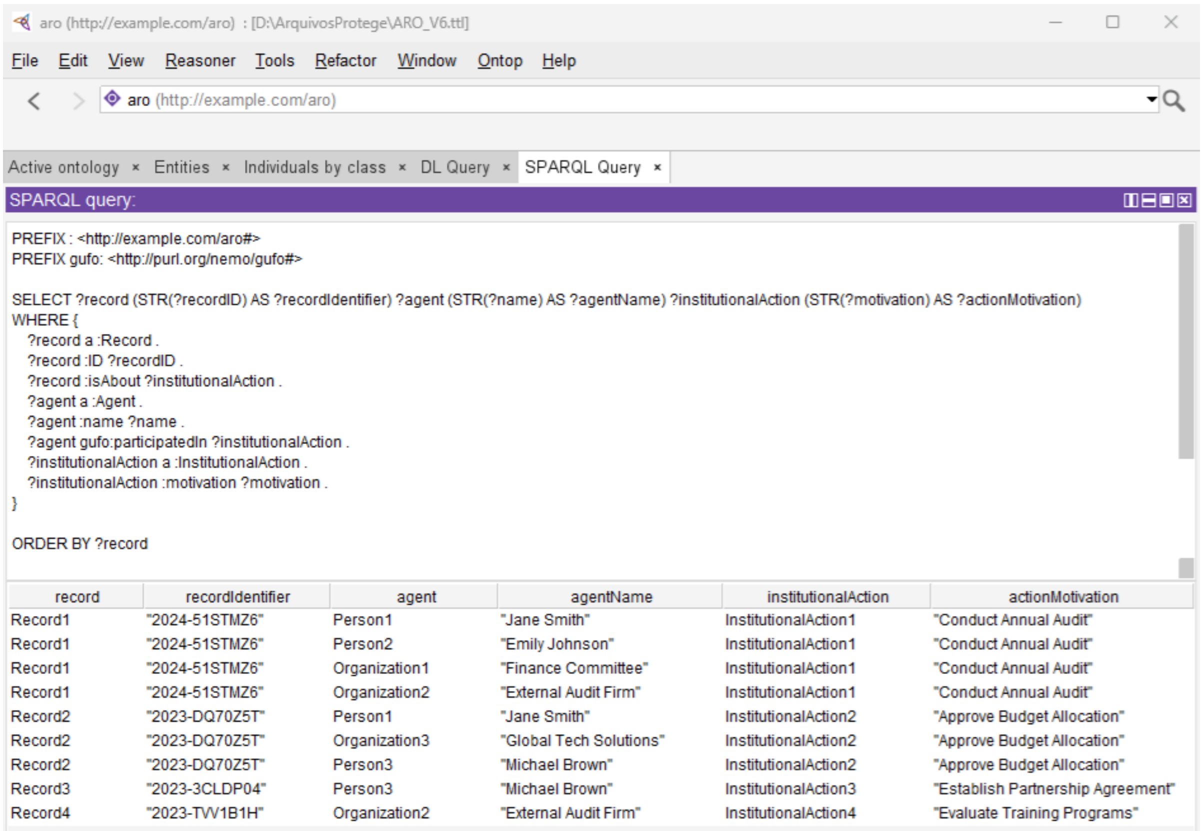

Figure 24 – SPARQL query for IQ06 verification.

IQ07 Is the record subject to signature? What type of signature? What is the method of signature? Who signs it? When was it signed?

Question IQ07 aims to verify whether the ontology is capable of adequately representing information related to the signing of archival records. The signature is an essential element for authenticity and fixity, and involves identifying various details, such as the type of signature, the method used, the identity of the signatory and the date on which the signature was made.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), events related to signing (`:RecordSigning`), and signing agents (`:Signer`), capturing detailed information about the type, method, and date of each signature. To achieve this, the query uses the `:ID` property to obtain the unique identifier of the record, and the properties `:signatureType`, `:signatureMethod`, and `:signatureDate` to capture the attributes of the signing events. The `:name` property retrieves the name of the signing agent. The `STR` function converts these values to strings, ensuring consistent data presentation in the query result. The `ORDER BY` clause organizes the results by record and signing event, facilitating a structured view of the data. The SPARQL query for IQ07 uses the prefixes `:` and `gufo:` to reference, respectively, terms from the ARO ontology at URI `http://example.com/aro#` and the Lightweight Implementation of UFO at URI (`http://purl.org/nemo/gufo#`). Figure 25 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

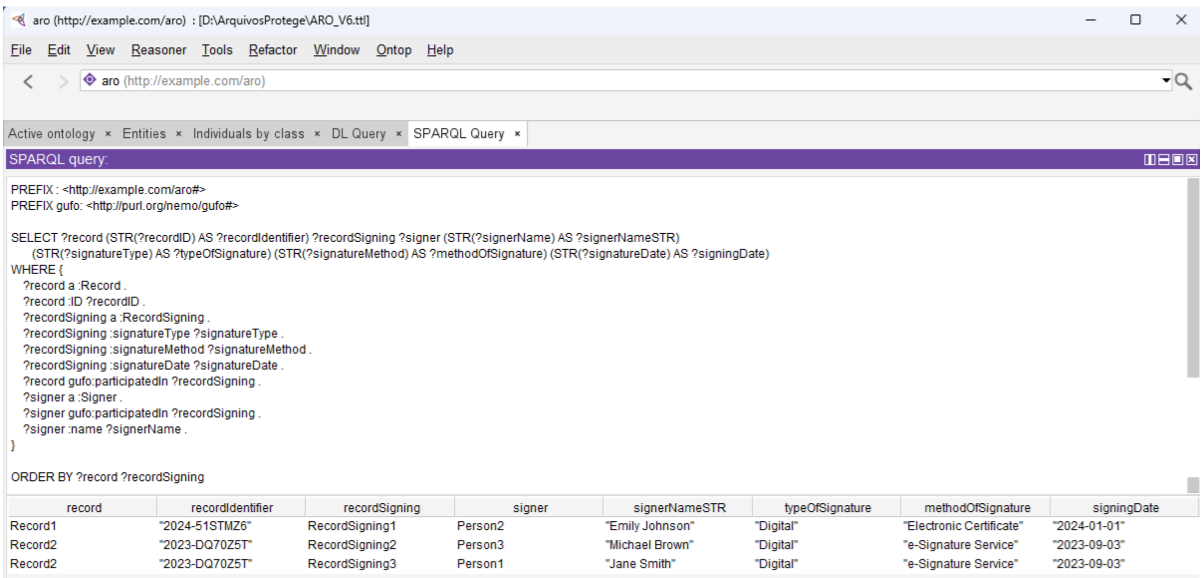

Figure 25 – SPARQL query for IQ07 verification.

IQ08 Does the content of the record have access restrictions?

Question IQ08 verifies whether the ontology is able to represent the access restrictions associated with the content of an archival record. In a record management context within organizations, access restrictions are essential to protect sensitive information and to ensure that access to the archival record is made in accordance with organizational or legal policies and standards.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), obtaining information on access restrictions, unique identifier, title, and file format for each record. To achieve this, the query uses the properties `:accessRestrictions` to capture access limitations, `:ID` to retrieve the unique identifier, `:title` to obtain the record title, and `:fileFormat` to specify the file format. The `STR` function converts these values to strings, ensuring consistency in data presentation in the query result. The `ORDER BY` clause organizes the result by record identifier, facilitating structured data analysis. The SPARQL query for IQ08 uses the prefix `:` to reference terms from the ARO ontology at URI `http://example.com/aro#`. Figure 26 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

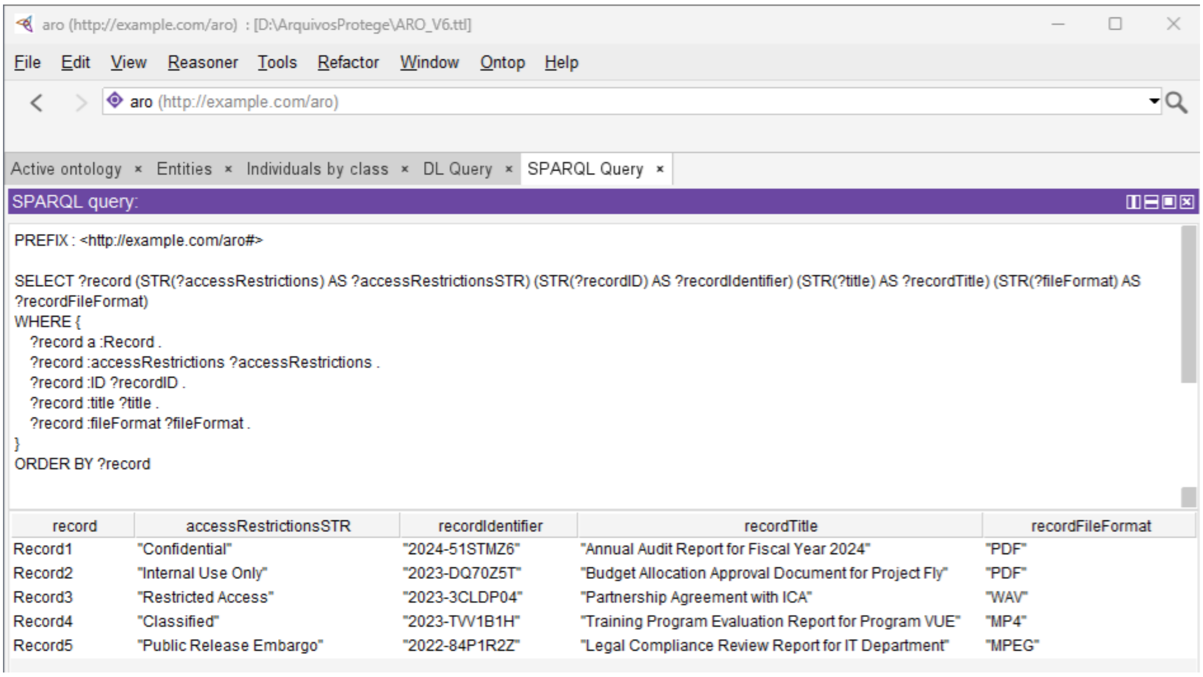

Figure 26 – SPARQL query for IQ08 verification.

IQ09 What is the file format of the record?

Question IQ09 aims to identify whether the ontology can display the file format of each archival record. The file format is an indispensable attribute for digital preservation actions, since it directly influences accessibility over time.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), obtaining their unique identifier, title, and file format. To achieve this, the query uses the properties `:ID` to capture the identifier, `:title` to retrieve the record's title, and `:fileFormat` to specify the file format. The `STR` function converts these values to strings, ensuring consistency in data presentation within the query result. The `ORDER BY` clause organizes the result by record identifier, facilitating structured data analysis. The SPARQL query for IQ09 uses the prefix `:` to reference terms from the ARO ontology at URI `http://example.com/aro#`. Figure 27 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

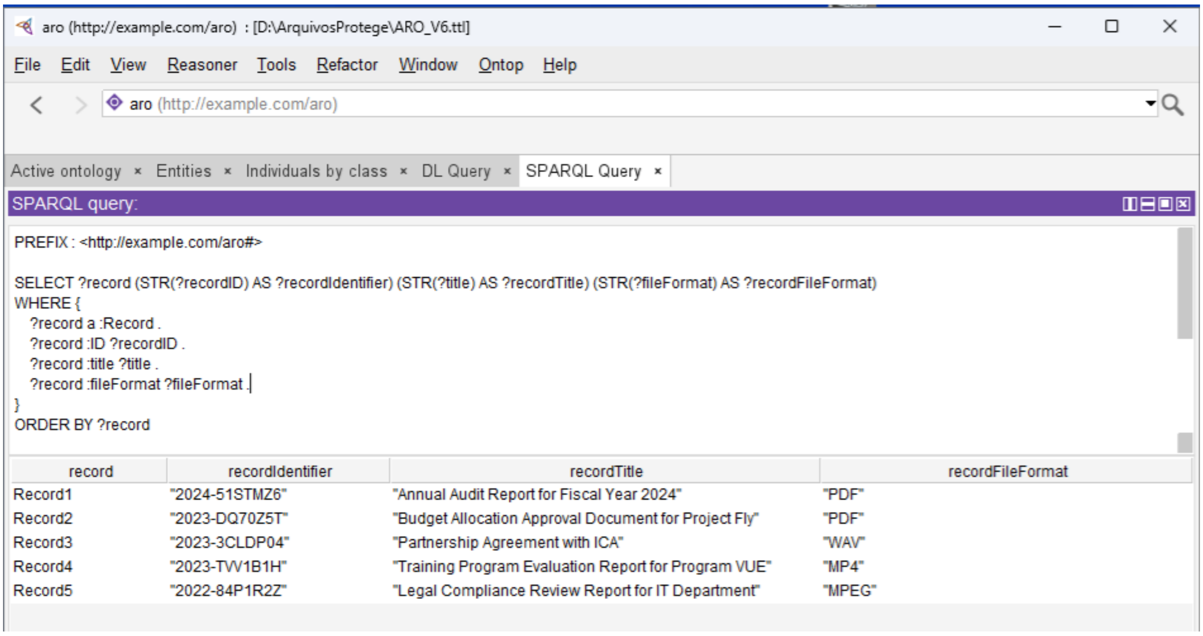

Figure 27 – SPARQL query for IQ09 verification.

IQ10 Does the record have any special access conditions?

Question IQ10 aims to verify the ontology's ability to represent specific conditions that may be required for access to an archival record, such as payment of fees, prior authorization or the need for physical presence at a specific location.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), obtaining information on access conditions, unique identifier, title, and file format for each record. To achieve this, the query uses the properties `:accessConditions` to capture access conditions, `:ID` to retrieve the unique identifier, `:title` for the record's title, and `:fileFormat` to specify the file format. The `STR` function converts these values to strings, ensuring consistency in data presentation within the query result. The `ORDER BY` clause organizes the result by record identifier, facilitating structured analysis. The SPARQL query for IQ10 uses the prefix `:` to reference terms from the ARO ontology at `http://example.com/aro#`. Figure 28 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

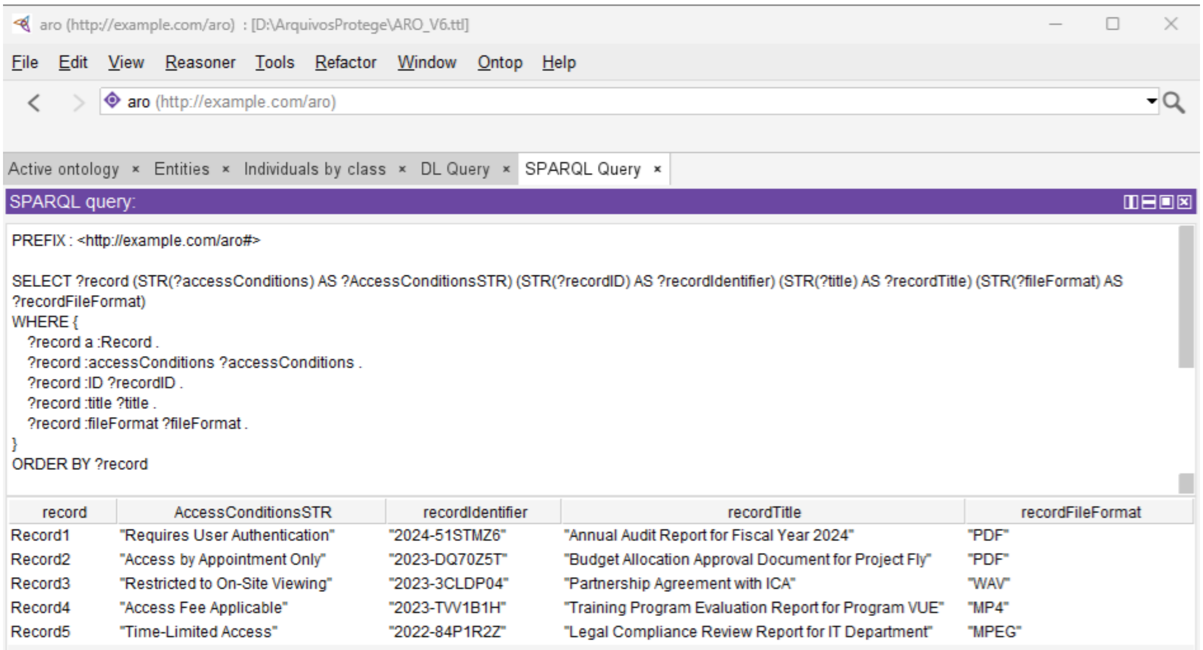

Figure 28 – SPARQL query for IQ10 verification.

IQ12 Which administrative body or authority is the record addressed to?

Question IQ12 seeks to verify whether the ontology is capable of presenting the institutional recipient or administrative authority to which the archival records were addressed. This feature makes it possible to trace the record flow and understand the institutional relationships involved, facilitating the management and retrieval of archival records according to organizational needs.

Using SPARQL, a query was developed to retrieve archival records (`:Record`), record sending events (`:RecordSending`), and addressees (`:Addressee`), obtaining the record identifier, the sending event, and the addressee's name. To achieve this, the query uses the `:ID` property to capture the record's identifier, `gufo:participatedIn` to associate the record and addressee with the sending event, and `:name` to retrieve the addressee's name. The `ORDER BY` clause organizes the result by record identifier, facilitating structured data analysis. The SPARQL query for IQ12 uses the prefixes `:` and `gufo:` to reference, respectively, terms from the ARO ontology at URI `http://example.com/aro#` and the Lightweight Implementation of the UFO at URI (`http://purl.org/nemo/gufo#`). Figure 29 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

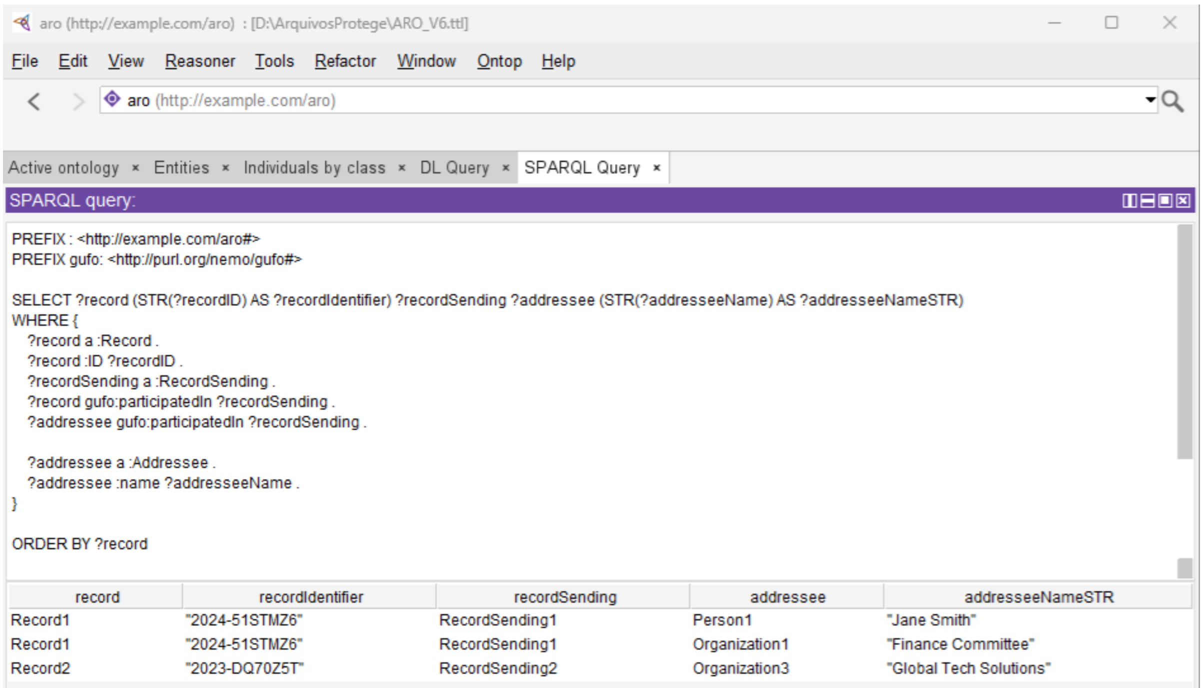

Figure 29 – SPARQL query for IQ12 verification.

IQ13 Where is the digital object located?

Question IQ13 aims to verify whether the ontology is capable of showing the fisical digital location of an archival record. In electronic records management systems, it is essential to know where the archival record is stored, thus guaranteeing its accessibility.

Using SPARQL, a query was developed to retrieve archival record (`:Record`), obtaining information on each record's unique identifier, file format, and digital object path. To achieve this, the query uses the properties `:ID` to capture the record's identifier, `:fileFormat` to specify the file format, and `:digitalObjectPath` to retrieve the digital object path. The `STR` function converts these values to strings, ensuring consistency in data presentation within the query result. The `ORDER BY` clause organizes the result by record identifier, facilitating structured data analysis. The SPARQL query for IQ13 uses the prefix `:` to reference terms from the ARO ontology at URI `http://example.com/aro#`. Figure 30 shows the Protégé IDE interface in which the SPARQL query was executed, along with the result obtained.

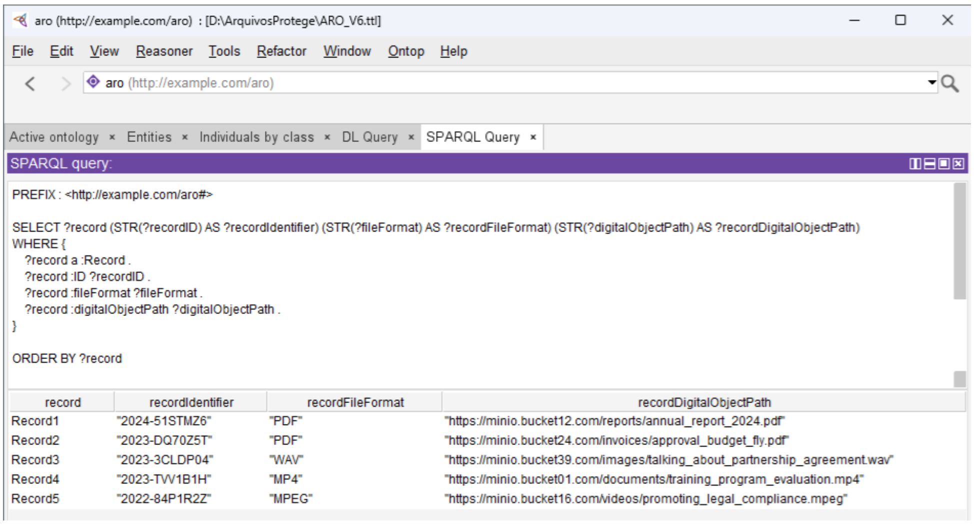

Figure 30 – SPARQL query for IQ13 verification.

5.4 Summary

Chapter 5 presented the ontology verification process, demonstrating how SPARQL queries were used to validate the ontology's ability to correctly answer integration questions. The test environment was structured with representative instances, ensuring that the ontological elements were properly populated to allow an evaluation close to reality. The SPARQL queries prepared covered all the integration questions defined in Section 3.1, thus allowing the analysis of the adequacy of the modeled concepts and their relationship with the needs of the archival domain. Thus, the ontological verification showed that the model developed has met the established requirements, consolidating the validity of the proposed model.

6 Final Considerations

This chapter presents the final considerations and conclusions regarding the research carried out in this dissertation. Initially, a brief summary is presented, highlighting the main aspects covered throughout the work, followed by a description of the contributions made. The difficulties and limitations of the work are also discussed, ending with suggestions for future work.

6.1 Summary of the Work

This research presented an ontology developed for the domain of archival records, called the Archival Record Ontology (ARO), with the aim of improving interoperability and digital preservation of archival records in administrative systems, especially in e-government contexts. The work began with a theoretical and methodological foundation, which presented the essential characteristics of the archival record, such as fixity, organicity, naturalness, uniqueness, authenticity, and impartiality of the records, as well as the choice of approaches and tools for building the ontology. The main objective of this research, which was to develop the ontology, was divided into three specific objectives, each of which was achieved, as described below.

The first specific objective, which aimed to identify structured resources, was achieved through systematic mapping using the CLeAR approach. To build the ARO ontology, the CLeAR (Conducting Literature Search for Artifact Reuse) approach was initially adopted, which guided a systematic mapping of existing structured resources in the archival records domain. This mapping resulted in the identification of 19 relevant structured resources, the analysis of which provided input to help develop the ontology. It was observed that most of these resources use ontologies as their main formalism for representing knowledge, with Web Ontology Language (OWL) being the most widely used. This finding highlights the importance of ARO aligning itself with widely used standards.

The second specific objective, related to the development of the ontology, was completed with the application of the SABiO methodology, which guided the process of development, employing a set of tools and resources to ensure accuracy and consistency in the representation of the archival record. For the conceptual basis, the principles of UFO were adopted, and gUFO was used to ensure a lightweight ontology compatible with the Semantic Web. The formal representation was built in OntoUML, using Visual Paradigm software, which enabled the visual representation of the classes, stereotypes, and relationships specific to the archival record.

Finally, the third specific objective, which involved the verification of the ontology, was achieved using SPARQL queries in the Protégé software. These queries were based on the integration questions formulated during the application of the CLeAR approach, with the aim of verifying that the ARO was able to answer them. The execution of these queries confirmed ARO's ability to respond adequately to the domain requirements (IQs) defined during the application of the CLeAR approach.

6.2 Applicability of the Work

The initial use of the ARO ontology will be within the Systemic Digital Preservation (SDP) project of archival records. The general objective of this project is to develop a conceptual model that enables the Systemic Digital Preservation (SDP) of digital archival records managed by E-Docs, the Corporate System for Digital Archival Records Management, used by the agencies and entities of the Executive Branch of the Government of the State of Espírito Santo. This project has a partnership with the Federal University of Espírito Santo, through the Archival Science department and the Computer Science Graduate Program.

Additionally, the ARO ontology can be applied to initiatives aimed at enhancing semantic interoperability and facilitating the integration of digital archival records across different document management systems. This would enable consistent representation and seamless exchange of archival records among various governmental agencies, supporting the effective management and long-term preservation of these records in the context of electronic government services.

6.3 Contribution

This work aims to contributes to the field of Archival Science by proposing the construction of an ontology based on the Unified Foundational Ontology (UFO) to represent archival records, with a focus on semantic interoperability and data integration between different systems. According to Guizzardi (2005), "UFO is designed to be extensible and adaptable to different domains", which makes it especially suitable for modeling areas such as archival records, where clearly defined representation is essential to guarantee semantic interoperability and data integration between different systems. It is expected that this ontology will not only enhance the production and preservation of archival records but also provide an effective solution to the challenges of interoperability, ensuring that archival records can be consistently exchanged and understood across different contexts.

In addition to the ontology developed, this study also contributes by demonstrating the application of the CLeAR approach in the archival domain. The systematic mapping conducted using this approach made possible to identify and analyze 19 structured resources relevant to the representation of archival records, highlighting their conceptual coverage and their main applications. The methodology adopted ensured a systematic and iterative survey of available resources, enabling an in-depth analysis of their characteristics and identifying gaps and opportunities for new approaches.

Finally, this dissertation also contributes to the discussion on the adoption of ontologies in the field of Archival Science, demonstrating how these conceptual models can improve the organization, retrieval, and preservation of archival records, as well as enhancing interoperability between government information systems.

6.4 Limitations and Difficulties

In the ARO ontology diagram, the relationship between the classes “Information object” and “Institutional action” does not have a defined stereotype. During the modeling process, the options available in OntoUML, such as “creation” and “participation”, were analyzed and discussed by the parties involved, but it was concluded that none of them adequately captured the nature of the association between these classes:

The «creation» stereotype implies that the ***Information Object*** is always the result of an ***Institutional Action***, however this is not a universal rule, as not all institutional actions lead to the creation of an information object. While the «participation» stereotype represents the involvement of an entity in an event, the relationship in question is more complex and goes beyond mere participation.

Given these considerations, it was decided to leave the relationship without a specific stereotype, thus allowing for greater interpretative flexibility. However, it was recognized the possibility of moving towards defining a new type of stereotype in the future, one that could better capture the specificity of this association. Furthermore, as this is the first version of the ontology, further refinements are expected in a future work, including the possibility of defining a new stereotype that is more appropriate for capturing the specific nature of this relationship in the context of archival record management.

Another limitation of this research is the lack of explicit modeling of other types of agents, such as families, mechanisms, or artificial intelligences. Although it is recognized that there are all kinds of agents, including families, mechanisms, or artificial intelligences, which are potential producers or accumulators of records, these agents have not been explicitly modelled. Adding these types of agents would require generally complex characteristics and relationships that are not within the scope or focus of this research, which is primarily concerned with the actors directly involved in the management and preservation of archival records in the context of public administration. This was to allow work with the depth in the key aspects of the modeled domain without losing either the coherence or clarity on the proposed ontology. Nevertheless, it is acknowledged that such agents can be considered in future work to extend the ontology.

Another difficulty encountered was the lack of documentation or examples of use of the structured resources under analysis. The presence of adequate documentation and practical examples is fundamental for carrying out activities to verify the coverage of the domain, understand the knowledge resources, and ensure proper alignment with the ontology under development. In the absence of these elements, carrying out these activities becomes significantly more complex and time-consuming, potentially leading to the non-use of the structured resource.

6.5 Future Work

Future work on the Archival Record Ontology (ARO) could explore several fronts to broaden its applicability and impact within the archival domain. Some of the possible directions include:

- **Improving documentation and usability**: One of the limitations identified in this research is the challenge of ensuring adequate documentation of the ontology. A future work could investigate good documentation practices that can improve the clarity, completeness, and usability of the ARO ontology. Well-structured documentation would minimize ambiguities and gaps in interpretation, facilitating the reuse of the ontology in different archival and digital preservation contexts;
- **Expansion to other archival subdomains**: Although this dissertation has focused on the representation of archival records within the context of government administrative systems, a future research could explore other archival subdomains, such as: Personal archives, Audiovisual archives, and Scientific archives;
- **Enhancement and integration with complementary ontologies**: Future research should consider refining ARO by developing an ontological network that integrates other structured resources and existing ontologies. This integration could broaden ARO's ability to support wider archival applications, ensuring alignment with emerging interoperability frameworks such as CIDOC-CRM and RiC-O;
- **Support for other types of agents**: The current scope of ARO does not explicitly model collective agents outside the context of formal institutions, such as communities, families, or cultural groups, which can play a significant role in the production and preservation of personal or cultural heritage archives. Future work could extend the ontology to incorporate these types of agents, thus enabling

a broader and more inclusive representation of archival practices throughout the record life cycle;

- **Automated archival record classification**: Classification is an essential activity of archival management, which seeks to determine the context of the archival record within the organization, as well as enabling a logical and structured organization of the archival record. As ARO defines a set of classes, relationships and restrictions that represent the entities involved in the archival record's life cycle, it can be used to structure databases and train machine learning models for automatic classification.

- **Record lifecycle**: Additionally, an important direction for future work is the extension of the ontology to explicitly model aspects related to the appraisal, retention, and disposition of archival records, incorporating concepts associated with the Records Retention and Disposition Schedule.

- **Alignment with UFO-C constructs**: Furthermore, a promising direction for future work is the ontological representation of the intentional and social aspects associated with archival records. In particular, it would be relevant to explore the alignment of the motivation attribute of record creation events with the concepts of the UFO-C module, such as Intentional Moments and Commitments. This alignment could enrich the modeling of institutional and agents' intentions underlying the creation of archival records, strengthening the ontological foundations for representing the provenance and contextual semantics of records.

The additional questions and observations raised during the defense process contributed to broadening the reflection on possible directions for future research. All comments were carefully analyzed, and detailed responses were sent to the supervisor. Although several suggestions were incorporated directly into the revised text, certain contributions—particularly those of an exploratory, expansive, or methodological prospective nature—were more appropriately addressed as avenues for future investigation. In this sense, insights such as the potential use of artificial intelligence to support the CLeAR method, alternative validation strategies, and the practical application of ontology in institutional contexts will be part of the future developments of this research, especially with regard to the extension and application of ARO.

Bibliography

ADEL, E. *et al.* Distributed electronic health record based on semantic interoperability using fuzzy ontology: a survey. *International Journal of Computers and Applications*, v. 40, p. 223 – 241, 2017. Cited on page 64.

ALMEIDA, J. a. P. A.; FALBO, R. A.; GUIZZARDI, G. Events as entities in ontology-driven conceptual modeling. In: *Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2019. p. 469–483. ISBN 978-3-030-33222-8. Available on: <https://doi.org/10.1007/978-3-030-33223-5_39>. Cited 6 on pages 8, 35, 36, 71, 72, and 76.

ALMEIDA, J. P. A. *et al. gUFO: a lightweight implementation of the Unified Foundational Ontology (UFO)*. [S.l.], 2020. Available on: <http://purl.org/nemo/doc/gufo>. Cited 2 on pages 40 and 41.

ARENAS, M.; PÉREZ, J. Querying semantic web data with sparql. In: *Proceedings of the thirtieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. [S.l.: s.n.], 2011. p. 305–316. Cited 2 on pages 20 and 21.

BALDO, A.; GOY, A.; MAGRO, D. *A Pipeline Supporting a Smart Access to Historical Documents based on a Rich Semantic Representation of Their Content: A Case Study on Time Expressions*. 2018. Available on: <https://doi.org/10.5220/0006929601990206>. Cited on page 56.

BARROS, T. H. B.; GOMES, D. L. Classification and knowledge organization systems: ontologies and archival classification: Proceedings of the fifteenth international isko conference 9-11 july 2018 porto, portugal. p. 103–111, 01 2018. Cited 2 on pages 17 and 29.

BARROS, T. H. B.; SOUSA, R. T. B. de. Organização do conhecimento e arquivologia: abordagens metodológicas. *Informação & Informação*, v. 24, n. 2, p. 76–92, 2019. Available on: <https://lume.ufrgs.br/bitstream/handle/10183/257121/001107977.pdf?sequence=1>. Cited on page 18.

BARROS, T. H. B.; SOUSA, R. T. de. Archival science and knowledge organization: Mapping methodological relationships. *Knowledge Organization*, v. 46, p. 493–501, 07 2019. Cited 2 on pages 18 and 29.

BASTOLA, R.; CAMPUS, P. Developing domain ontology for issuing certificate of citizenship of nepal. *Journal of Information Technology*, v. 2, n. 02, p. 73–90, 2020. Cited on page 17.

BBC. *Provenance Ontology — bbc.com*. 2012. [Accessed 10-07-2022]. Available on: <https://www.bbc.com/ontologies/provenance-ontology/>. Cited on page 59.

BIAGETTI, M. T. A comparative analysis and evaluation of bibliographic ontologies. In: *Challenges and Opportunities for Knowledge Organization in the Digital Age*. [S.l.]: Ergon Verlag, 2018. p. 499–510. Cited on page 55.

BOLES, F.; GREENE, M. Et Tu Schellenberg? Thoughts on the Dagger of American Appraisal Theory. *The American Archivist*, v. 59, n. 3, p. 298–310, 07 1996. ISSN 0360-9081. Available on: <https://doi.org/10.17723/aarc.59.3.k3x81g0852825l31>. Cited on page 24.

CAMPOS, P. M. C. *Designing a Network of Reference Ontologies for the Integration of Water Quality Data*. Dissertação (Mestrado) — Federal University of Espírito Santo, 2019. Available on: <https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/designing_a_network_of_reference_ontologies_for_the_integration_of_water_quality_data_2019.pdf>. Cited 6 on pages 8, 19, 25, 26, 53, and 93.

CAMPOS, P. M. C. *et al.* Finding reusable structured resources for the integration of environmental research data. *Environmental Modelling & Software*, Elsevier, v. 133, p. 104813, 2020. Cited 5 on pages 19, 25, 26, 53, and 93.

CARRIERO, V. A. *et al.* Arco: The italian cultural heritage knowledge graph. In: *The Semantic Web – ISWC 2019*. Springer International Publishing, 2019. p. 36–52. Available on: <https://link.springer.com/chapter/10.1007/978-3-030-30796-7_3>. Cited 2 on pages 55 and 56.

CCSDS. *Reference Model for an Open Archival Information System (OAIS)*. [S.l.], 2012. Magenta Book, Issue 2. Available on: <https://public.ccsds.org/pubs/650x0m2.pdf>. Cited 2 on pages 68 and 69.

CICCARESE, P. *et al.* Pav ontology: provenance, authoring and versioning. *Journal of biomedical semantics*, Springer, v. 4, p. 1–22, 2013. Available on: <https://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-37>. Cited on page 59.

CLINE, S. The archivist as translator: Representation and the language of context. *The American Archivist*, 2022. Cited on page 23.

COLAVIZZA, G. *et al.* Archives and ai: An overview of current debates and future perspectives. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, v. 15, p. 1 – 15, 2021. Available on: <https://dl.acm.org/doi/pdf/10.1145/3479010>. Cited on page 16.

COOK, T. Archival science and postmodernism: new formulations for old concepts. *Archival Science*, v. 1, p. 3–24, 2001. Available on: <https://doi.org/10.1007/BF02435636>. Cited on page 17.

COOK, T. Evidence, memory, identity, and community: four shifting archival paradigms. *Archival science*, Springer, v. 13, p. 95–120, 2013. Available on: <https://link.springer.com/article/10.1007/s10502-012-9180-7>. Cited 3 on pages 17, 23, and 24.

CORTESE, C.; MANTEGARI, G. Extending the Digital Archives of Italian Psychology with Semantic Data. In: *Proceedings of the 1st International Workshop on Semantic Digital Archives (SDA 2011)*. CEUR-WS.org, 2011. (CEUR Workshop Proceedings, v. 801), p. 60–71. Available on: <https://ceur-ws.org/Vol-801/paper5.pdf>. Cited on page 52.

DARANYI, S. *et al.* Pericles – digital preservation through management of change in evolving ecosystems. In: . [S.l.: s.n.], 2015. p. 51–74. Cited on page 57.

DAY, M. *CEDARS Project Document AIW01*. 1998. Accessed: 2023-08-25. Available on: <http://www.ukoln.ac.uk/metadata/cedars/AIW01.html#40>. Cited on page 63.

DERRIDA, J. *Archive fever: A Freudian impression*. [S.l.]: University of Chicago Press, 1996. Cited on page 16.

Dublin Core Metadata Initiative. *Dublin Core Metadata Element Set, Version 1.1*. 2023. Accessed: 2025-01-19. Available on: <https://www.dublincore.org/specifications/dublin-core/dces/>. Cited on page 59.

DURANTI, L. *Diplomatics: New Uses for an Old Science*. Lanham: Scarecrow Press, 1997. Available on: <https://www.researchgate.net/publication/225035619_Diplomatics_New_Uses_for_an_Old_Science>. Cited on page 23.

EASTWOOD, T. Nailing a little jelly to the wall of archival studies. *Archivaria*, v. 35, 1992. Available on: <https://archivaria.ca/index.php/archivaria/article/download/11900/12853>. Cited on page 24.

EASTWOOD, T. What is archival theory and why is it important? *Archivaria*, p. 122–130, 1994. Available on: <https://archivaria.ca/index.php/archivaria/article/download/11991/12954/>. Cited on page 23.

ELENA, T. *et al.* Creating an historical archive ontology: Guidelines and evaluation. In: *2006 1st International Conference on Digital Information Management*. [S.l.: s.n.], 2007. p. 102–109. Cited on page 28.

ESWARAIAH, P.; SYED, H. An efficient ontology model with query execution for accurate document content extraction. *Indonesian Journal of Electrical Engineering and Computer Science*, 2023. Cited on page 29.

FALBO, R. de A. Sabio: Systematic approach for building ontologies. In: *ONTO.COM/ODISE@FOIS*. [s.n.], 2014. Available on: <https://api.semanticscholar.org/CorpusID:17310637>. Cited 8 on pages 8, 19, 31, 32, 33, 66, 90, and 93.

FALBO, R. de A. Ontouml quick guide and pattern language. In: . [s.n.], 2017. Available on: <http://inf.ufes.br/~vitorsouza/falbo/OntoUML_Quick_Guide_and_Pattern_Language.pdf#:~:text=OntoUML%20is%20an%20ontology%20modeling%20language%20that%20focuses>. Cited on page 39.

FARINELLI, F.; ALMEIDA, M. B.; MELO, S. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental. In: *14º Encontro Nacional de Pesquisa em Ciência da Informação*. Florianópolis, SC: [s.n.], 2013. Available on: <https://brapci.inf.br/index.php/res/v/185252>. Cited on page 17.

GILLILAND, A. J.; MCKEMMISH, S.; LAU, A. J. *Research in the Archival Multiverse*. Monash University Publishing, 2016. Accessed: 2023-11-02. Available on: <https://library.oapen.org/handle/20.500.12657/31429>. Cited on page 17.

GOY, A. *et al.* Building semantic metadata for historical archives through an ontology-driven user interface. *Journal on Computing and Cultural Heritage (JOCCH)*, ACM, v. 13, n. 3, p. 1–36, 2020. Cited 2 on pages 28 and 52.

GOY, A.; MAGRO, D.; BALDO, A. A semantic web approach to enable a smart route to historical archives. *Journal of Web Engineering*, v. 18, n. 4–6, p. 287–318, 2019. Available on: <https://ieeexplore.ieee.org/abstract/document/10251822>. Cited on page 56.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v. 5, n. 2, p. 199–220, 1993. ISSN 1042-8143. Available on: <https://www.sciencedirect.com/science/article/pii/S1042814383710083>. Cited 2 on pages 18 and 27.

GUARINO, N. The ontological level. In: . [s.n.], 1994. Available on: <https://api.semanticscholar.org/CorpusID:11834869>. Cited 2 on pages 27 and 28.

GUARINO, N. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. [S.l.]: IOS press, 1998. v. 46. Cited on page 27.

GUARINO, N.; OBERLE, D.; STAAB, S. What is an ontology? *Handbook on ontologies*, Springer, p. 1–17, 2009. Available on: <https://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf>. Cited 2 on pages 27 and 28.

GUARINO, N.; WELTY, C. An overview of ontoclean. In: STAAB, S.; STUDER, R. (Ed.). *Handbook on Ontologies*. [S.l.]: Springer, 2009. p. 201–220. Cited on page 24.

GUIZZARDI, G. *Ontological foundations for structural conceptual models*. Tese (PhD Thesis - Research UT, graduation UT) — University of Twente, oct 2005. Available on: <https://research.utwente.nl/files/6042428/thesis_Guizzardi.pdf>. Cited 12 on pages 19, 24, 33, 38, 39, 41, 66, 68, 69, 70, 74, and 109.

GUIZZARDI, G. On ontology, ontologies, conceptualizations, modeling languages, and (meta)models. In: *International Baltic Conference on Databases and Information Systems*. [s.n.], 2007. Available on: <https://api.semanticscholar.org/CorpusID:7704374>. Cited on page 27.

GUIZZARDI, G. *et al.* Ufo: Unified foundational ontology. *Appl. Ontology*, v. 17, p. 167–210, 2021. Available on: <https://api.semanticscholar.org/CorpusID:244875087>. Cited 5 on pages 8, 33, 34, 38, and 71.

GUIZZARDI, G.; FALBO, R. de A.; GUIZZARDI, R. Grounding software domain ontologies in the unified foundational ontology (ufo): The case of the ode software process ontology. In: *Conferencia Iberoamericana de Software Engineering*. [s.n.], 2008. Available on: <https://api.semanticscholar.org/CorpusID:32490124>. Cited 9 on pages 8, 33, 34, 35, 36, 37, 38, 71, and 74.

GUIZZARDI, G. *et al.* Endurant types in ontology-driven conceptual modeling: Towards ontouml 2. In: TRUJILLO, J. C. *et al.* (Ed.). *Conceptual Modeling - 37th International Conference, {ER} 2018, Xi'an, China, October 22-25, 2018, Proceedings*. [S.l.]: Springer, 2018. p. 136–150. Cited 4 on pages 19, 39, 40, and 41.

GUIZZARDI, G. *et al.* Towards ontological foundations for conceptual modeling: The unified foundational ontology (ufo) story. *Applied Ontology*, IOS Press, v. 10, n. 3-4, p. 259 – 271, 2015. ISSN 1570-5838. Cited 2 on pages 19 and 34.

HAWKINS, A. Archives, linked data and the digital humanities: increasing access to digitised and born-digital archives via the semantic web. *Archival Science*, v. 22, n. 3, p. 319–344, 2022. ISSN 1573-7500. Available on: <https://doi.org/10.1007/s10502-021-09381-0>. Cited on page 28.

HEBECHE, L. A. Ontologia. In: . UFSC, 2012. ISBN 978-85-61484-24-8. Available on: <https://hebeche.paginas.ufsc.br/files/2016/03/OntologiaI.pdf>. Cited on page 27.

HEDSTROM, M. Digital preservation: A time bomb for digital libraries. *Computers and the Humanities*, v. 31, p. 14, 05 1997. Available on: <https://api.semanticscholar.org/CorpusID:15327062>. Cited on page 18.

HORROCKS, I.; PATEL-SCHNEIDER, P. F.; HARMELEN, F. V. From shiq and rdf to owl: The making of a web ontology language. *Journal of web semantics*, Elsevier, v. 1, n. 1, p. 7–26, 2003. Available on: <https://www.sciencedirect.com/science/article/abs/pii/S1570826803000027>. Cited on page 31.

IORIO, A. D.; CARON, B. Premis 3.0 ontology: Improving semantic interoperability of preservation metadata. In: *Proceedings of the International Conference on Preservation of Digital Objects (iPRES)*. Bern, Switzerland: [s.n.], 2016. Accessed: 2023-07-07. Available on: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>. Cited on page 58.

IP2, I. T. *The InterPARES 2 Project Dictionary*. InterPARES Trust, 2002–2007. Available on: <http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=37148968&CFTOKEN=41583140>. Cited on page 95.

ISA, W. M. W. *et al.* Digital Preservation of Cultural Heritage: Terengganu Brassware Craft Knowledge Base. *International Journal of Advanced Computer Science and Applications*, Science and Information (SAI) Organization Limited, v. 10, n. 6, 2019. Cited on page 52.

ISO. *15489-1 – Information and documentation – Records management – Part 1: Concepts and principles*. Geneva, Switzerland, 2016. Available on: <https://www.iso.org/standard/62542.html>. Cited on page 16.

KETELAAR, E. Archival temples, archival prisons: Modes of power and protection. *Archival Science*, v. 2, n. 3-4, p. 221–238, 2002. Available on: <https://doi.org/10.1007/BF02435623>. Cited on page 24.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. *Inf. Softw. Technol.*, Butterworth-Heinemann, USA, v. 55, n. 12, p. 2049–2075, dec 2013. ISSN 0950-5849. Cited on page 64.

KONTOPOULOS, E. *et al.* Ontology-based representation of context of use in digital preservation. In: . [s.n.], 2016. Available on: <https://ceur-ws.org/Vol-1608/paper-08.pdf>. Cited on page 57.

KWASNIKOWSKA, N.; MOREAU, L.; BUSSCHE, J. V. D. A formal account of the open provenance model. *ACM Trans. Web*, Association for Computing Machinery, New York, NY, USA, v. 9, n. 2, may 2015. ISSN 1559-1131. Available on: <https://doi.org/10.1145/2734116>. Cited on page 59.

LAGOS, N.; VION-DURY, J.-Y. Digital preservation based on contextualized dependencies. In: *Proc. of the 2016 ACM Symposium on Document Engineering*. [S.l.: s.n.], 2016. p. 35–44. Cited on page 52.

LAGOS, N.; WADDINGTON, S.; VION-DURY, J.-Y. On the preservation of evolving digital content–the continuum approach and relevant metadata models. In: SPRINGER. *Research Conference on Metadata and Semantics Research*. [S.l.]: Springer International Publishing, 2015. p. 15–26. Cited on page 52.

LASSILA, O.; HENDLER, J.; BERNERS-LEE, T. The semantic web. *Scientific American*, v. 284, n. 5, p. 34–43, 2001. Available on: <https://lassila.org/publications/2001/SciAm.pdf>. Cited on page 40.

LEMIEUX, V. L.; TRAPNELL, S. E. *Public Access to Information for Development: A Guide to the Effective Implementation of Right to Information Laws*. Washington, DC: World Bank, 2016. Available on: <https://www.researchgate.net/publication/304541058>. Cited on page 23.

LI, C.; SUGIMOTO, S. Provenance Description of Metadata using PROV with PREMIS for Long-term Use of Metadata. In: *International conference on Dublin Core and metadata applications*. [S.l.: s.n.], 2014. p. 147–156. Cited 2 on pages 52 and 58.

LIYANAGE, H.; KRAUSE, P.; LUSIGNAN, S. D. Using ontologies to improve semantic interoperability in health data. *Journal of innovation in health informatics*, BMJ Specialist Journals, v. 22, n. 2, p. 309–315, 2015. Cited 2 on pages 17 and 18.

Lohmann, Steffen and Haag, Florian and Negru, Stefan. Towards a Visual Notation for OWL: A Brief Summary of VOWL. In: *Proc. of the 12th OWL: Experiences and Directions Workshop*. [S.l.]: Springer International Publishing, 2016. p. 143—-153. ISBN 9783319332444. Cited on page 50.

LÖW, M. M. *Proveniência arquivística e semântica: uso de ontologias de BFO para representação da produção documental e contexto*. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2021. Available on: <https://lume.ufrgs.br/handle/10183/238060>. Cited on page 30.

LUZ, C. d. S. *Ontologia digital arquivística: interoperabilidade e preservação da informação arquivística em sistemas informatizados de arquivos e na web*. Tese (Doutorado) — Universidade de São Paulo, 2016. Available on: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-03022017-154503/pt-br.php>. Cited on page 30.

MANDAL, S. Item Relationships using Dublin Core, BIBO, FOAF, and FRBR for Managing Resources of Cultural Heritage: Designing a Prototype Integrated Framework. *Library Philosophy and Practice*, Library Philosophy and Practice, p. 1–16, 2021. Cited on page 52.

MARKETAKIS, Y.; TZITZIKAS, Y. Dependency management for digital preservation using semantic web technologies. *International Journal on Digital Libraries*, Springer, v. 10, n. 4, p. 159–177, 2009. Cited on page 52.

MATOUSEK, K.; KOUBA, Z. Knowledge exploration in historical artefact record keeping. In: IEEE. *Proc. of the 13th International Workshop on Database and Expert Systems Applications*. [S.l.], 2002. p. 555–558. Cited on page 52.

MAYER, R. *et al.* Using ontologies to capture the semantics of a (business) process for digital preservation. *International Journal on Digital Libraries*, Springer, v. 15, n. 2, p. 129–152, 2015. Cited on page 57.

MAZIKANA, P. The challenges of archiving digital information. *The International Information & Library Review*, v. 29, n. 3, p. 307–317, 1997. ISSN 1057-2317. Available on: <https://www.sciencedirect.com/science/article/pii/S1057231797900746>. Cited on page 16.

MBEMBE, A. The power of the archive and its limits. In: *Refiguring the archive*. Springer, 2002. p. 19–27. Available on: <https://link.springer.com/chapter/10.1007/978-94-010-0570-8_2>. Cited on page 16.

MCKEON, R. *The basic works of Aristotle*. New York: Modern Library, 1941. ISBN 9780307417527. Available on: <https://archive.org/details/basicworksofaris00aris/page/n7/mode/2up>. Cited on page 24.

MEDINA, M. A.; CHAVEZ-ARAGON, A.; CHAVEZ, R. O. Construction, implementation and maintenance of ontologies of records. In: *Proceedings of the Fourth Latin American Web Congress*. USA: IEEE Computer Society, 2006. (LA-WEB '06), p. 67–73. ISBN 0769526934. Available on: <https://doi.org/10.1109/LA-WEB.2006.9>. Cited on page 29.

MELLO, B. H. de *et al.* Semantic interoperability in health records standards: a systematic literature review. *Health and Technology*, Springer, p. 1–18, 2022. Available on: <https://doi.org/10.1007/s12553-022-00639-w>. Cited on page 63.

MOREAU, L. *et al.* The open provenance model core specification (v1.1). *Future Generation Computer Systems*, v. 27, n. 6, p. 743–756, 2011. ISSN 0167-739X. [Accessed 10-07-2022]. Available on: <https://www.sciencedirect.com/science/article/pii/S0167739X10001275>. Cited on page 59.

MUNIR, K.; ANJUM, M. S. The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, 2018. Available on: <https://api.semanticscholar.org/CorpusID:56247516>. Cited on page 29.

OAI. *Open Archives Initiative*. 2001. [Accessed 16-08-2023]. Available on: <https://www.openarchives.org/>. Cited on page 60.

OHNO, T. *Toyota Production System: Beyond Large-Scale Production*. 1st. ed. Portland: Productivity Press, 1988. Available on: <https://doi.org/10.4324/9780429273018>. Cited on page 53.

PRAMARTHA, C.; DAVIS, J.; KUAN, K. Digital preservation of cultural heritage: an ontology-based approach. In: *ACIS 2017 Proceedings*. [S.l.: s.n.], 2017. p. 55. Cited on page 52.

PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata, Version 3.0*. 2015. Available on: <http://www.loc.gov/standards/premis/v3/>. Cited on page 23.

PREMIS Editorial Committee. *Guidelines for using the PREMIS Version 3 OWL Ontology*. 2022. Available on: <https://www.loc.gov/standards/premis/ontology/pdf/premis3-owl-guidelines.pdf>. Cited 2 on pages 57 and 58.

RHOADS, J. B. *The role of archives and records management in national information systems: a RAMP study*. [S.l.], 1983. 56 p. Available on: <https://unesdoc.unesco.org/ark:/48223/pf0000084735.locale=en>. Cited on page 17.

ROGERS, C. A literature review of authenticity of records in digital systems: From 'machine-readable' to records in the cloud. *Acervo*, v. 29, p. 16–44, 2016. Available on: <https://api.semanticscholar.org/CorpusID:193853516>. Cited on page 24.

ROUSSEAU, J.-Y.; COUTURE, C. *Os fundamentos da disciplina arquivística*. Lisboa, Portugal: Publicações Dom Quixote, 2004. Cited on page 52.

SAA. *Society of American Archivists. Definition of Archival Record*. 2005–2023. Accessed: 2022-10-05. Available on: <https://dictionary.archivists.org/entry/archival-record.html>. Cited on page 17.

SAA. *Society of American Archivists. Definition of Provenance*. 2005–2023. Accessed: 2022-10-05. Available on: <https://dictionary.archivists.org/entry/provenance.html>. Cited on page 86.

SAA. *Society of American Archivists. Definition authenticity*. 2022. Accessed: 2022-10-05. Available on: <https://dictionary.archivists.org/entry/authenticity.html>. Cited on page 24.

SALES, T. P.; FONSECA, C. M.; BARCELOS, P. P. F. *OntoUML Vocabulary Specification - Version 1.1.0*. 2022. Available on: <https://w3id.org/ontouml/vocabulary/docs>. Cited on page 75.

SANTOS, V. B. D. Preservação de documentos arquivísticos digitais. *Ciência da Informação*, v. 41, n. 1, 2012. Cited 4 on pages 22, 23, 24, and 25.

SCHELLENBERG, T. R. *Modern Archives: Principles and Techniques*. Chicago: University of Chicago Press, 1956. Available on: <https://archive.org/details/modernarchivespr0000sche_e6f7>. Cited on page 23.

SHOTTON, D. *Comparison of BIBO and FaBiO*. 2011. Available on: <https://web.archive.org/web/20220127164713/https://opencitations.wordpress.com/2011/06/29/comparison-of-bibo-and-fabio/>. Cited on page 55.

SMITH, B. Ontology and information systems. In: . [s.n.], 2002. Available on: <https://api.semanticscholar.org/CorpusID:14326521>. Cited on page 27.

TECO. *TIMBUS: Digital Preservation for Timeless Business Processes and Services*. 2024. Accessed: 2024-03-30. Available on: <https://www.teco.edu/research/timbus-digital-preservation-for-timeless-business-processes-and-services/>. Cited on page 57.

TENNIS, J. T.; ROGERS, C. Authenticity metadata and the ipam: progress toward the interpares application profile. In: *International Conference on Dublin Core and Metadata Applications*. [S.l.: s.n.], 2012. p. 38–45. Cited on page 51.

TIKHONOV, A. Preservation of digital images: Question of fixity. *Heritage*, v. 2, n. 2, p. 1160–1165, 2019. ISSN 2571-9408. Available on: <https://www.mdpi.com/2571-9408/2/2/75>. Cited 2 on pages 22 and 24.

TSHERING, Y.; ANUTARIYA, C. Enabling Semantic Interoperability in Bhutan's E-Government: An Ontology-based Framework. In: *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. [S.l.: s.n.], 2022. p. 1–6. Cited on page 17.

UNESCO. *Archives for a Better World*. 2023. Accessed: 2024-03-06. Available on: <https://www.unesco.org/en/archives4betterworld>. Cited on page 17.

UNESCO. *UNESCO Archives: Preserving Our Heritage for Future Generations*. 2023. Accessed: 2024-03-06. Available on: <https://www.unesco.org/en/articles/unesco-archives-preserving-our-heritage-future-generations>. Cited on page 17.

Visual Paradigm. *Visual Paradigm Learning and Training*. 2024. Available on: <https://www.visual-paradigm.com/support/>. Cited on page 39.

VSESVIATSKA, O. *et al.* ArDO: an ontology to describe the dynamics of multimedia archival records. In: *Proc. of the 36th Annual ACM Symposium on Applied Computing*. [S.l.: s.n.], 2021. p. 1855–1863. Cited 2 on pages 52 and 54.

W3C, T. P. W. G. at. *PROV-O: The PROV Ontology — w3.org*. 2011–2013. [Accessed 10-07-2022]. Available on: <https://www.w3.org/TR/prov-o/>. Cited on page 58.

W3C, W. W. W. C. *SPARQL Query Language for RDF*. 2013. Available on: <https://www.w3.org/TR/rdf-sparql-query/>. Cited 2 on pages 20 and 96.

WALLACE, D. A. Metadata and the archival management of electronic records: a review. *Archivaria*, 1993. Cited on page 63.

WANG, Z. *et al.* An ontology for chinese government archives knowledge representation and reasoning. *IEEE Access*, v. 9, p. 130199–130211, 2021. Available on: <https://ieeexplore.ieee.org/document/9535509>. Cited on page 56.

WOHLIN, C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: . New York, NY, USA: Association for Computing Machinery, 2014. (EASE '14). ISBN 9781450324762. Available on: <https://doi.org/10.1145/2601248.2601268>. Cited on page 49.

YEO, G. Telling the difference : Perceptions of uniqueness and identity in archives. In: *XVI Congresso Brasileiro de Arquivologia*. São Paulo, Brazil, August 2010: [s.n.], 2010. Available on: <https://www.ucl.ac.uk/infostudies/geoffrey-yeo/tellingdifference.pdf>. Cited on page 24.

ZARUBIN, A. A.; KOVAL, A.; MOSHKIN, V. S. Ontological model of an electronic archive document. *Proceedings of the 2020 3rd International Conference on Geoinformatics and Data Analysis*, 2020. Available on: <https://api.semanticscholar.org/CorpusID:220281541>. Cited on page 29.

ZHAO, J. Openprovenance model vocabulary specification. 2010. [Accessed 10-07-2022]. Available on: <https://open-biomed.sourceforge.net/opmv/ns.html>. Cited on page 59.

Appendix

APPENDIX A – ARO Turtle Syntax

Note: The content of this appendix was generated using Visual Paradigm's "Export to gUFO" feature.

```
@prefix : <http://example.com/aro#>.
@prefix gufo: <http://purl.org/nemo/gufo#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.

<http://example.com/aro> rdf:type owl:Ontology;
    owl:imports gufo:.
:OrganizationMembership rdf:type owl:Class, gufo:Kind, owl:NamedIndividual;
    rdfs:subClassOf gufo:Relator;
    rdfs:label "Organization Membership"@en.
:RecordSending rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Record Sending"@en.
:Record rdf:type owl:Class, gufo:Kind, owl:NamedIndividual;
    rdfs:label "Record"@en.
:Sender rdf:type owl:Class, gufo:Role, owl:NamedIndividual;
    rdfs:subClassOf gufo:FunctionalComplex;
    rdfs:label "Sender"@en.
:RecordCreation rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Record Creation"@en.
:BindingAction rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:label "Binding Action"@en.
:InstitutionalAction rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Institutional Action"@en.
:ComplexRecordCreation rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:label "Complex Record Creation"@en.
:OrganizationMember rdf:type owl:Class, gufo:Role, owl:NamedIndividual;
    rdfs:label "Organization Member"@en.
:Signer rdf:type owl:Class, gufo:Role, owl:NamedIndividual;
    rdfs:subClassOf gufo:VariableCollection;
    rdfs:label "Signer"@en.
:Organization rdf:type owl:Class, gufo:Kind, owl:NamedIndividual;
    rdfs:subClassOf gufo:FunctionalComplex;
    rdfs:label "Organization"@en.
:Agent rdf:type owl:Class, gufo:Category, owl:NamedIndividual;
    rdfs:subClassOf gufo:Object;
    rdfs:label "Agent"@en.
:InformationObject rdf:type owl:Class, gufo:Category, owl:NamedIndividual;
    rdfs:subClassOf gufo:FunctionalComplex;
    rdfs:label "Information Object"@en.
:AgentParticipation rdf:type owl:Class, gufo:EventType, owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Agent Participation"@en.
```

```
:SimpleRecord rdf:type owl:Class , gufo:SubKind , owl:NamedIndividual;
    rdfs:label "Simple Record"@en.
:CompoundRecord rdf:type owl:Class , gufo:SubKind , owl:NamedIndividual;
    rdfs:label "Compound Record"@en.
:Binding rdf:type owl:Class , gufo:Kind , owl:NamedIndividual;
    rdfs:subClassOf gufo:Relator;
    rdfs:label "Binding"@en.
:InstitutionalMemberAction rdf:type owl:Class , gufo:EventType , owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Institutional Member Action"@en.
:SimpleRecordCreation rdf:type owl:Class , gufo:EventType , owl:NamedIndividual;
    rdfs:label "Simple Record Creation"@en.
:OrganizationMemberActor rdf:type owl:Class , gufo:Role , owl:NamedIndividual;
    rdfs:label "Organization Member Actor"@en.
:RecordCreationParticipation rdf:type owl:Class , gufo:EventType , owl:NamedIndividual;
    rdfs:label "Record Creation Participation"@en.
:RecordSigning rdf:type owl:Class , gufo:EventType , owl:NamedIndividual;
    rdfs:subClassOf gufo:Event;
    rdfs:label "Record Signing"@en.
:Person rdf:type owl:Class , gufo:Kind , owl:NamedIndividual;
    rdfs:subClassOf gufo:FunctionalComplex;
    rdfs:label "Person"@en.
:Addressee rdf:type owl:Class , gufo:Role , owl:NamedIndividual;
    rdfs:subClassOf gufo:VariableCollection;
    rdfs:label "Addressee"@en;
    rdfs:comment "ip."@en.
:Binder rdf:type owl:Class , gufo:Role , owl:NamedIndividual;
    rdfs:label "Binder"@en.
:AgentParticipationType rdf:type owl:NamedIndividual;
    rdfs:subClassOf gufo:ConcreteIndividualType;
    rdfs:label "Agent Participation Type"@en.
:Signer rdfs:subClassOf :Agent.
:CompoundRecord rdfs:subClassOf :Record.
:SimpleRecord rdfs:subClassOf :Record.
:Record rdfs:subClassOf :InformationObject.
:Sender rdfs:subClassOf :Agent.
:RecordCreationParticipation rdfs:subClassOf :InstitutionalMemberAction.
:SimpleRecordCreation rdfs:subClassOf :RecordCreation.
:ComplexRecordCreation rdfs:subClassOf :RecordCreation.
:BindingAction rdfs:subClassOf :InstitutionalMemberAction.
:OrganizationMemberActor rdfs:subClassOf :OrganizationMember.
:OrganizationMember rdfs:subClassOf :Person.
:bindingAction rdfs:subPropertyOf :memberAction.
:Organization rdfs:subClassOf :Agent.
:Person rdfs:subClassOf :Agent.
:Addressee rdfs:subClassOf :Agent.
:SimpleRecordCreation rdfs:subClassOf :InstitutionalMemberAction.
:Binder rdfs:subClassOf :OrganizationMemberActor.
:isAbout rdf:type owl:ObjectProperty;
    rdfs:domain :Record;
    rdfs:range :InstitutionalAction;
    rdfs:label "is about"@en.
:isBoundInto rdf:type owl:ObjectProperty;
    rdfs:domain :CompoundRecord;
    rdfs:range :SimpleRecord;
```

```
    rdfs:label "is bound into"@en.
:AgentParticipationType gufo:categorizes :AgentParticipation.
:OrganizationMembership rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:mediates;
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Organization
].
:InstitutionalMemberAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:manifestedIn ];
  owl:someValuesFrom :OrganizationMembership
].
:OrganizationMembership rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:mediates ];
  owl:someValuesFrom :OrganizationMember
].
:OrganizationMember rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:mediates;
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :OrganizationMembership
].
:RecordSending rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:someValuesFrom :Record
], [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:someValuesFrom :Addressee
].
:Addressee rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:someValuesFrom :RecordSending
].
:RecordSending rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Sender
].
:Sender rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:someValuesFrom :RecordSending
].
:RecordCreation rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:wasCreatedIn ];
  owl:someValuesFrom :Record
].
:Record rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:wasCreatedIn;
```

```
  owl:someValuesFrom :RecordCreation
].
:RecordSigning rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Record
].
:SimpleRecord rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :BindingAction
].
:BindingAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:maxQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :SimpleRecord
].
:Binding rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:mediates;
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :CompoundRecord
].
:CompoundRecord rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :BindingAction
].
:BindingAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Binder
].
:Binder rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:someValuesFrom :BindingAction
].
:BindingAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:wasCreatedIn ];
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Binding
].
:Binding rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:wasCreatedIn;
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :BindingAction
].
:InstitutionalAction rdfs:subClassOf [
  rdf:type owl:Restriction;
```

```
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:someValuesFrom :Agent
].
:AgentParticipation rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:isEventProperPartOf ];
  owl:someValuesFrom :InstitutionalAction
].
:InstitutionalAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:isEventProperPartOf;
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :AgentParticipation
].
:RecordCreationParticipation rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:isEventProperPartOf ];
  owl:minQualifiedCardinality "2"^^xsd:nonNegativeInteger;
  owl:onClass :ComplexRecordCreation
].
:ComplexRecordCreation rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:isEventProperPartOf;
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :RecordCreationParticipation
].
:RecordSigning rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:someValuesFrom :Signer
].
:Signer rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:someValuesFrom :RecordSigning
].
:AgentParticipation rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Agent
].
:Binding rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:mediates ];
  owl:maxQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :SimpleRecord
].
:SimpleRecord rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:mediates;
  owl:minQualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :Binding
].
:InstitutionalMemberAction rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty [ owl:inverseOf gufo:participatedIn ];
```

```
  owl:qualifiedCardinality "1"^^xsd:nonNegativeInteger;
  owl:onClass :OrganizationMemberActor
].
:OrganizationMemberActor rdfs:subClassOf [
  rdf:type owl:Restriction;
  owl:onProperty gufo:participatedIn;
  owl:someValuesFrom :InstitutionalMemberAction
].
```

APPENDIX B – ARO SPARQL Queries

IQ01 Who is the producer of the record?

```
PREFIX : <http://example.com/ARO#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX gufo: <http://purl.org/nemo/gufo#>

SELECT ?record (STR(?recordTitle) AS ?recordTitleSTR) ?agent (STR(?
    agentTitle) AS ?agentTitleSTR) ?agentType
WHERE {
    ?record a :Record .
    ?record dc:title ?recordTitle .
    ?record :isAbout ?institutionalAction .
    ?institutionalAction a :InstitutionalAction .
    ?agent a :Agent .
    ?agent gufo:participatedIn ?institutionalAction .
    ?agent dc:title ?agentTitle .

    # Check if the agent is a Person or an Organization
    OPTIONAL { ?agent a :Person . BIND("Person" AS ?agentType) }
    OPTIONAL { ?agent a :Organization . BIND("Organization" AS ?agentType)
        }
}
```

IQ02 What is the date and place of production of the record?

```
PREFIX : <http://example.com/aro#>
PREFIX gufo: <http://purl.org/nemo/gufo#>

SELECT ?record (STR(?recordID) AS ?recordIDSTR) ?recordCreation (STR(?
    creationDate) AS ?creationDateSTR) (STR(?creationLocation) AS ?
    creationLocationSTR)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record gufo:wasCreatedIn ?recordCreation .
    ?recordCreation a :RecordCreation .
    ?recordCreation :creationDate ?creationDate .
```

```
    ?recordCreation :creationLocation ?creationLocation .
}
```

IQ03 What is the administrative motivation for producing the record?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?recordID) AS ?recordIDString) ?institutionalAction (
    STR(?motivation) AS ?motivationString)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record :isAbout ?institutionalAction .
    ?institutionalAction a :InstitutionalAction .
    ?institutionalAction :motivation ?motivation .
}
```

IQ04 What is the unique identifier of the record?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?recordID) AS ?uniqueIdentifier)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
}
```

IQ05 What is the title of the record?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?recordID) AS ?recordIdentifier) (STR(?title) AS ?
    recordTitle)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record :title ?title .
}
```

IQ06 Who are the parties concerned?

```
PREFIX : <http://example.com/aro#>
PREFIX gufo: <http://purl.org/nemo/gufo#>

```

```
SELECT ?record (STR(?recordID) AS ?recordIdentifier) ?agent (STR(?name) AS ?agentName) ?institutionalAction (STR(?motivation) AS ?actionMotivation)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record :isAbout ?institutionalAction .
    ?agent a :Agent .
    ?agent :name ?name .
    ?agent gufo:participatedIn ?institutionalAction .
    ?institutionalAction a :InstitutionalAction .
    ?institutionalAction :motivation ?motivation .
}

ORDER BY ?record
```

IQ07 Is the record subject to signature? What type of signature? What is the method of signature? Who signs it? When was it signed?

```
PREFIX : <http://example.com/aro#>
PREFIX gufo: <http://purl.org/nemo/gufo#>

SELECT ?record (STR(?recordID) AS ?recordIdentifier) ?recordSigning ?signer (STR(?signerName) AS ?signerNameSTR)
    (STR(?signatureType) AS ?typeOfSignature) (STR(?signatureMethod) AS ?methodOfSignature) (STR(?signatureDate) AS ?signingDate)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?recordSigning a :RecordSigning .
    ?recordSigning :signatureType ?signatureType .
    ?recordSigning :signatureMethod ?signatureMethod .
    ?recordSigning :signatureDate ?signatureDate .
    ?record gufo:participatedIn ?recordSigning .
    ?signer a :Signer .
    ?signer gufo:participatedIn ?recordSigning .
    ?signer :name ?signerName .
}

ORDER BY ?record ?recordSigning
```

IQ08 Does the content of the record have access restrictions?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?accessRestrictions) AS ?accessRestrictionsSTR) (STR
    (?recordID) AS ?recordIdentifier) (STR(?title) AS ?recordTitle) (STR(?
    fileFormat) AS ?recordFileFormat)
WHERE {
    ?record a :Record .
    ?record :accessRestrictions ?accessRestrictions .
    ?record :ID ?recordID .
    ?record :title ?title .
    ?record :fileFormat ?fileFormat .
}
ORDER BY ?record
```

IQ09 What is the file format of the record?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?recordID) AS ?recordIdentifier) (STR(?title) AS ?
    recordTitle) (STR(?fileFormat) AS ?recordFileFormat)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record :title ?title .
    ?record :fileFormat ?fileFormat .
}
ORDER BY ?record
```

IQ10 Does the record have any special access conditions?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?accessConditions) AS ?AccessConditionsSTR) (STR(?
    recordID) AS ?recordIdentifier) (STR(?title) AS ?recordTitle) (STR(?
    fileFormat) AS ?recordFileFormat)
WHERE {
    ?record a :Record .
    ?record :accessConditions ?accessConditions .
    ?record :ID ?recordID .
    ?record :title ?title .
```

```
    ?record :fileFormat ?fileFormat .
}
ORDER BY ?record
```

IQ12 Which administrative body or authority is the record addressed to?

```
PREFIX : <http://example.com/aro#>
PREFIX gufo: <http://purl.org/nemo/gufo#>

SELECT ?record (STR(?recordID) AS ?recordIdentifier) ?recordSending ?
    addressee (STR(?addresseeName) AS ?addresseeNameSTR)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?recordSending a :RecordSending .
    ?record gufo:participatedIn ?recordSending .
    ?addressee gufo:participatedIn ?recordSending .

    ?addressee a :Addressee .
    ?addressee :name ?addresseeName .
}

ORDER BY ?record
```

IQ13 Where is the digital object located?

```
PREFIX : <http://example.com/aro#>

SELECT ?record (STR(?recordID) AS ?recordIdentifier) (STR(?fileFormat) AS
     ?recordFileFormat) (STR(?digitalObjectPath) AS ?
    recordDigitalObjectPath)
WHERE {
    ?record a :Record .
    ?record :ID ?recordID .
    ?record :fileFormat ?fileFormat .
    ?record :digitalObjectPath ?digitalObjectPath .
}

ORDER BY ?record
```

Universidade Federal do Espírito Santo
Pró-Reitoria de Pesquisa e Pós-Graduação
Centro Tecnológico
Programa de Pós-Graduação em Informática (PPGI)

DECLARAÇÃO DE VERSÃO FINAL

Eu, Jussara Teixeira, então discente do Programa de Pós-Graduação em Informática (PPGI) da Universidade Federal do Espírito Santo (UFES) e autor desta dissertação, e o meu orientador, Prof. Dr. Vítor E. Silva Souza, declaramos que a versão aqui apresentada incorpora as correções exigidas pelos membros examinadores da banca no dia da defesa.

gov.br
Documento assinado digitalmente
JUSSARA TEIXEIRA
Data: 02/06/2025 11:55:35-0300
Verifique em https://validar.iti.gov.br

Assinatura do autor

gov.br
Documento assinado digitalmente
VITOR ESTEVAO SILVA SOUZA
Data: 02/06/2025 15:20:09-0300
Verifique em https://validar.iti.gov.br

Assinatura do orientador