UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO CONHECIMENTO

JAIME ANDRADE PINTO

GOVERNMENT INSTITUTIONS AND THEIR DOCUMENTS: AN ONTOLOGICAL APPROACH

Belo Horizonte
2020

JAIME ANDRADE PINTO

GOVERNMENT INSTITUTIONS AND THEIR DOCUMENTS: AN ONTOLOGICAL APPROACH

Tese apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Doutor, área de concentração Ciência da Informação.
Linha de Pesquisa: Gestão e Organização do Conhecimento
Orientador: Maurício Barcellos Almeida

BELO HORIZONTE

2020

P659g Pinto, Jaime Andrade

Government institutions and their documents [recurso eletrônico]: an ontological approach / Jaime Andrade Pinto. - 2020.

1 recurso eletrônico (103 f. : il., color): pdf.

Orientador: Maurício Barcellos Almeida.
Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.
Referências: f. 84-99.
Anexos: f. 100-103.
Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da Informação – Teses. 2. Representação do conhecimento (Teoria da informação) – Teses. 3.Ontologias (Recuperação da informação) - Teses. I. Título. II. Almeida, Maurício Barcellos. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU:025.4.03

Ficha catalográfica: Maianna Giselle de Paula CRB:2642
Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG.

	UNIVERSIDADE FEDERAL DE MINAS GERAIS **PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO**	

FOLHA DE APROVAÇÃO

Government Institutions and their Documents: an Ontological Approach

JAIME ANDRADE PINTO

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Gestão e Tecnologia.

Aprovada em 30 de setembro de 2020, pela banca constituída pelos membros:

Prof(a). Maurício Barcellos Almeida (Orientador)
ECI/UFMG [por videoconferência]

Daniela Lucas da Silva Lemos

Prof(a). Daniela Lucas da Silva Lemos
UFES [por videoconferência]

Fabrício M Mendonça

Prof(a). Fabrício Martins Mendonça
UFJF [por videoconferência]

Livia Marangon Duffles Teixeira

Dr(a). Livia Marangon Duffles Teixeira
Fundação CPqD [por videoconferência]

Sérgio Mariano Dias

Dr(a). Sérgio Mariano Dias
SERPRO [por videoconferência]

Prof(a). Carlos Renato Storck
CEFET-MG [por videoconferência]

Prof(a). João Eduardo de Rezende Dantas
Aposentado/UFMG [por videoconferência]

Belo Horizonte, 30 de setembro de 2020.

ATA DA DEFESA DE TESE DO ALUNO

JAIME ANDRADE PINTO

Realizou-se, no dia 30 de setembro de 2020, às 08:30 horas, Videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *Government Institutions and their Documents: an Ontological Approach*, apresentada por JAIME ANDRADE PINTO, por videoconferência, número de registro 2016720527, graduado no curso de ADMINISTRACAO/NOTURNO, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Maurício Barcellos Almeida - ECI/UFMG [por videoconferência] (Orientador), Prof(a). Daniela Lucas da Silva Lemos - UFES [por videoconferência], Prof(a). Fabrício Martins Mendonça - UFJF [por videoconferência], Dr(a). Livia Marangon Duffles Teixeira - Fundação CPqD [por videoconferência], Dr(a). Sérgio Mariano Dias - SERPRO [por videoconferência], Prof(a). Carlos Renato Storck CEFET-MG [por videoconferência], João Eduardo de Rezende Dantas - Aposentado/UFMG [por videoconferência].

A Comissão considerou a tese:

(**X**) Aprovada

(   ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 30 de setembro de 2020.

Prof(a). Maurício Barcellos Almeida

Prof(a). Daniela Lucas da Silva Lemos

Prof(a). Fabrício Martins Mendonça

Dr(a). Livia Marangon Duffles Teixeira

Dr(a). Sérgio Mariano Dias

Prof(a). Carlos Renato Storck

Prof(a). João Eduardo de Rezende Dantas

Dedico este trabalho a minha amada esposa, Fátima, pelo seu apoio e paciência com as minhas muitas ausências. E também aos meus filhos, Luiza, André e Isabela, pelo interesse e incentivo, tantas vezes demonstrado.

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Maurício, pelo apoio e pelos valiosos ensinamentos.

Agradeço ao SERPRO, meu empregador, pelo apoio institucional, com liberação de horas para estudo.

Agradeço aos Gerentes do SERPRO que me apoiaram e permitiram que esta pesquisa pudesse ser realizada: Antônio Villaschi, Robson Dias e, especialmente, Lourene Nunes, por acreditar em mim desde o começo.

Faço um agradecimento especial ao amigo Célio Faria, não só pelo seu apoio gerencial no SERPRO, mas também pelas inúmeras horas de discussão sobre a pesquisa e como aplicar este conhecimento no nosso trabalho de desenvolvimento de sistemas.

Um apoio essencial, a quem devo muito pelo suporte durante todo o período de estudos, foi dado pela Alessandra Berbert Ferreira, da Universidade Corporativa do SERPRO.

A todos, ofereço a minha gratidão.

“What is needed is not a new technology, but new metaphors: a metalanguage for describing the world that complex systems have wrought. A new shorthand is required, one that simultaneously acknowledges and addresses the reality of a world in which people, politics, culture and technology are utterly enmeshed.”

James BRIDLE. **New Dark Age: Technology and the End of the Future**

ABSTRACT

E-Government studies are of growing importance in our times. Two factors contribute to this: first, the need for better services to citizens and, second, the enormous volume, diversity, and variety of government data. This scenario characterizes e-Government initiatives as a socio-technical system in the current big-data environment. To face these modern challenges, system and service designers need techniques and tools that contemplate solutions for two fundamental problems, namely, the complexity of socio-technical processes required to government functioning, and the obligation to coexist with large legacy databases. One approach that has proven valuable to address these challenges is conceptual modeling supported by ontologies. This modeling mode allows new systems to incorporate two essential functionalities: knowledge representation and heterogeneous data integration. Knowledge Representation is a field of Artificial Intelligence that enables the codification of knowledge about the world in order that a computer system can use it to solve complex problems. In this context, heterogeneous data integration is known as Ontology-Based Data Access: a technique that has proven worthwhile, especially for data stored in relational databases. This work presents the essentials of this theoretical field and a practical experiment in data integration. The study is based on the Theory of Social Acts and is implemented with the ontologies available on the OBO Foundry. Specifically, we demonstrate the principles of the Document Acts Theory and its corresponding ontology, the D-Acts ontology. The philosophical foundations on the nature of public corporations are also explored, and a Systematic Review of e-Government Literature is presented. We focus on government systems, and, for that purpose, a government process and its actual data, both described in the Brazilian Government Open Data Portal, are used as an instance of research. According to considerations of scientific research methodology, the present work was classified as applied and qualitative research, with some exploratory characteristics. The essential thesis contribution is to propose a method to combine two public and generic components able to build an ontological model as well as a piece of an integration system that works to access heterogeneous relational databases. We claim that this construction may be generalized for other government processes and even for a generic business process.

RESUMO

Os estudos de governo eletrônico são de importância crescente em nossos tempos. Dois fatores contribuem para isso: primeiro, a necessidade de melhores serviços aos cidadãos e, segundo, o enorme volume, diversidade e variedade de dados governamentais. Este cenário caracteriza as iniciativas de governo eletrônico como um sistema sociotécnico no atual ambiente de bigdata. Para enfrentar esses modernos desafios os projetistas de sistemas e serviços precisam de técnicas e ferramentas que contemplem soluções para dois problemas fundamentais: a complexidade dos processos sociotécnicos necessários para o funcionamento dos governos e a obrigação de coexistir com grandes bases de dados legadas. Uma abordagem que tem se mostrado valiosa para enfrentar estes problemas é a modelagem conceitual apoiada por ontologias. Este método de modelagem permite que os novos sistemas incorporem duas funcionalidades essenciais: representação do conhecimento e integração de dados heterogêneos. A Representação do Conhecimento é um campo da Inteligência Artificial que permite a codificação do conhecimento humano sobre o mundo para que um sistema de computador possa resolver problemas complexos. Neste contexto, a integração heterogênea de dados é conhecida como *Ontology-Based Data Access*: uma técnica que tem se mostrado útil, especialmente para dados armazenados em bancos de dados relacionais. Este trabalho apresenta o essencial deste campo teórico e uma experiência prática na integração de dados. O estudo é baseado na Teoria dos Atos Sociais e implementado com as ontologias da OBO Foundry. Especificamente, demonstramos os princípios da Teoria dos Atos Sociais e sua ontologia correspondente, a Ontologia dos Atos de Documentos. Os fundamentos filosóficos sobre a natureza das corporações públicas também são explorados, e é apresentada uma Revisão Sistemática da Literatura de Governo Eletrônico. Nosso foco é em sistemas de governo e, para isso, um processo governamental e seus dados reais, ambos descritos no Portal de Dados Abertos do Governo Brasileiro, são utilizados como um estudo de caso. De acordo com considerações de metodologia de pesquisa científica, a presente pesquisa foi classificada como pesquisa aplicada e qualitativa, com algumas características exploratórias. A contribuição essencial da tese é propor um método para combinar dois componentes públicos e genéricos capazes de construir um modelo ontológico, como um elemento de um sistema de integração que funciona para acessar bases de dados relacionais heterogêneas. Consideramos que esta construção pode ser generalizada para outros processos governamentais e até mesmo um processo de negócio genérico.

LIST OF FIGURES

LIST OF TABLES

Summary

1. Introduction

E-Government (e-Gov) studies are fundamental themes for all modern society, whether in government, business, or even in individual sphere. As this field of study is very broad and complex, we need to approach it in comprehensive and multidisciplinary ways. E-Government studies have social and human implications that affect every aspect of our daily lives, whether individually or collectively.

The approach here focuses on a very dear aspect of Information Science, namely, the theoretical view of Knowledge Representation (KR). We analyze formal representations of government institutions, their social nature, and their mode of operation. This study is conducted formally, describing the theoretical framework that supports the approach, and providing a practical system experience that demonstrates the validity of the adopted model. In this context, the word "model" is related to the Knowledge Organization System (KOS) because we chose ontology, which is a type of KOS, as the central component of our study. Thus, we have adopted a formal and well-founded structure to propose a prototype of Integration Ontology for a specific governmental process, which is taken as a case for an experiment. This aforementioned structure is the set of ontologies that uses the Open Biological and Biomedical Ontologies (OBO) Foundry. The OBO Foundry consists of a group of people dedicated to the construction and maintenance of public ontologies related to Life Sciences, all of them based on the Basic Formal Ontology (BFO)[1].

The present work has two parts. Part I is theoretical, and Part II is empirical. Part I – Theoretical Framework – begins with a review of the essentials of conceptual modeling of information systems, presenting a theory overview and its historical origins. We analyze how the use of ontologies can enhance conceptual modeling and the connection with KR, a contemporary and critical area of Artificial Intelligence (AI). We also come up with a Systematic Literature Review (SLR) in which we seek to present and synthesize the research on ontologies applied to the Electronic Government. Besides, we offer an overview of the state of the art on this matter. The goal is to identify the current stage of elaboration of ontologies in this domain that are public, open, and collaboratively built.

A fundamental theoretical concept of our work is the Social Ontology framework, a philosophical study representing a new understanding of social entities that encompass human artifacts and social devices like money, property, governments,

[1] In this work, we use BFO on its latest version, BFO 2.0, as available in https://github.com/BFO-ontology/BFO.

documents, and institutions, to mention but a few. This theoretical ground comes from the middle of the 20th century, with philosophers like Reinach and Searle. Currently, Smith, Almeida, and Brochhausen formalize parts of the social ontology as ontology. To provide a more comprehensive view of the richness and vast application of Social Ontology, we use its concepts and rationale to debate the ontological nature of government corporations, by asking whether they seem to be kinds of public or private institutions. In the sequel, we briefly present the BFO's spatio-temporal basis and the two branches of Social Ontology, namely, the theories of Speech Acts and Document Acts.

Section 4.3 presents a critical point of research on the technological structure. There are two available public technologies that we use to deliver a technical solution. One is the Ontology-Based Data Access (OBDA) technique for accessing relational databases using ontologies. The other component is the Ontology of Document Acts (D-Acts), an intermediate ontology that we use to model government data.

Finally, at the end of Part I, we show how we build an ontology for integration using D-Acts, and we explain the proposed model applied to a real process we took as the case. Besides, we elaborate on D-Acts, adding to it five new classes, all of which were applicable to this case. We claim that the empirical method we use here could be replicated for other business processes. To allow this reproduction, we present in Section 6 the fundamentals and the steps of the experimentation methodology used in the experiment.

In the practical part of this thesis, Part II, there are two fundamental components: a technical methodology with its software tools and a public ontology. The technical part is the application of OBDA techniques using open tools and Internet standards. The ontology used is D-Acts, which is a mid-level ontology based on a theory of documentary acts that describes what people can do with documents. We outline, in practical experience, a method to solve the emerging problem of interoperability in large government data environments. Part II is divided into two sections: Section 7 presents the preliminary methodological phases of the experiment, and Section 8 shows all the results and pieces of evidence.

The scope of our research is the theoretical and applied studies on public ontologies in the area of e-Gov. The theoretical study limits are restricted to Information Sciences. However, our research project involves, in a broad way, socio-technical systems in governmental environments considered together with other areas of study, such as Software Engineering and Database Modeling. In the exact specification of our study focus, we are interested in using the concepts, techniques, and methods of KR to provide a mechanism for shared access to information, seeking to make accessible and intelligible a sophisticated technical vocabulary typically found in government institutions.

The research question is: is it possible to integrate government data using OBDA and public ontologies?

Part I – THEORETICAL FRAMEWORK

In the first part of this work, we introduce the theoretical basis of our work. This research has focused on the application of ontologies in conceptual modeling in the domain of e-Government.

Section 2 deals with Information Systems Conceptual Modeling theories and is divided into three subsections. Section 2. 1 − Conceptual Modeling: Origins and Overview − presents a historical overview of the theory, its fundamentals, and an overview of the state of the art. The main objective of Section 2.2 − Ontological conceptual modeling and knowledge representation − is to show the evolution of conceptual modeling incorporating the use of ontologies and knowledge representation. Thus, we have made an effort to clarify these concepts, providing some well-founded definitions. In Section 2.3 − Ontological approaches to Government and e-Government − we seek to give the reader a view on our way to think about ontological approaches to e-Gov. We intend to demonstrate the reasoning we follow to adopt the vision that is presented here and no other alternative that could be considered.

Section 3 is a Systematic Literature Review (SLR) on the use of public ontologies in the field of e-Government. In Subsection 3.2, we explore the concept of e-Government on its fundamentals and main definitions. The remainder of Section 3 follows the methodological steps and procedures, as stated by Barbara Kitcheham, in her book “Procedures for Conducting the Systematic Review” (KITCHENHAM, 2004). Besides, we use a valuable tool to support the comprehensive literature reviews, called StArt, “State of the Art Through Systematic Review” (FABBRI et al., 2012, 2016; ZAMBONI et al., 2010).

Continuing with the presentation of our theoretical basis, in Section 4 − Socio-technical approaches for e-Government − we explore aspects related to socio-technical approaches to e-Gov. This approach is the complement to the previous two sections because it is the insertion of the facts and objects of e-Gov reality into an ontological conceptual model. We use the philosophical foundations of the BFO, our high-level ontology, to analyze some issues about corporations in general. Moreover, in sequence, we present the Social Ontology as the supporting building block of the D-Acts ontology. In Section 4.1 we start asking about the real nature of government, seeking a shared vision between public and private corporations. We present some interdisciplinary concepts to compare the corporations in an ontological and, at the same time, socio-technical vision. Two fundamental components of our study are also presented: Spatio-temporal ontology (BFO) and Social Ontology with two branches, Speech Acts and Document Acts.

In Section 5, we bring together all the theories and models and propose a prototype model to represent a governmental process, the Public Budgeting. We model this process as a "social document act," which is an essential class on the ontology of D-Acts. From this point on, we present some contributions to D-Act and propose this model as a template or an example of how to model other business and governmental processes using BFO and D-Acts, always from a socio-technical point of view.

Finally, in Section 6, we base our methodological choices and explain the context of the practical part of the research. The fundamental theories and empirical studies on the Cross Standard Process for Data Mining (CRISP-DM) are discussed and evaluated for use in this context.

2. Conceptual Modeling of Information Systems

Conceptual modeling involves picking up various aspects of the real world and describing them in the way of a model that can be used for communication among different stakeholders involved in developing information systems, such as domain experts and system designers (MYLOPOULOS, 1992). The result of a conceptual modeling effort can be a human-readable diagram or model that can be translated into an understandable computer format, usually through a specialized language (STOREY; TRUJILLO; LIDDLE, 2015). More precisely, Wand says that conceptual modeling is about "capturing and representing certain aspects of human perceptions of the real world" in such a manner that we can integrate it into an information system (WAND; STOREY; WEBER, 1999, p. 495).

Besides, some authors claim that designers are modeling a reality, not undertaking conceptual modeling. What we are doing is modeling the reality, be it our present reality, a previous reality, or a future reality. For example, one might be using UML to draw a software artifact. By doing this, the designer is glimpsing a future reality where the software system exists. The models allow us to comprehend aspects of the future system. Ontologies and conceptual modeling are entirely appropriate for the assessment of the appropriateness of a modeling language to perform modeling tasks from a given point of view (MILTON; KAZMIERCZAK, 2006).

At this introductory text, we need a more precise definition of information systems (IS). The IS concept began to be accessible after the 1960s, and today it can be considered a well-known and widely used concept. However, it is not so easy to define because it encompasses three different and complementary perspectives: (i) the contribution they provide, (ii) their structure and behavior, and (iii) the functions they perform. For our purposes, the third approach is the most useful because it abstracts from why and how the designed system performs its functions (OLIVÉ, 2007). At a top-level view, an IS has three primary functions. An IS stores, retrieves, and manipulates information about a specific portion of the world. We can accept this oversimplification for both its simplicity and generality.

In the context of the present work, “socio-technical systems” are considered to be systems that include, in addition to technological artifacts, social institutions. Technological artifacts are, fundamentally, computer equipment and programs. Social institutions that are involved in information systems are, in general, knowledge, capital, human work, and cultural meaning. Socio-technical systems do not work autonomously, but rather as a result of the activities of human actors, who are engaged in some business process (GEELS, 2004; SAVAGET et al., 2019; SOMMERVILLE, 2016).

At this point, we need one more constraint to apply this definition to conceptual modeling studies. What kind of information does the system perform its functions on? It is not a piece of general information, but specific information that describes a part of the world here called the knowledge domain of the system. As Borgida says, "this description can be viewed as a model of that world, or more accurately, the user's conceptualization of the world" (BORGIDA, 1986, p. 461).

The rest of this section is organized as follows: Section 2.1 addresses Conceptual Modeling, its origins as theory and practice and gives an overview of the state of the art; Section 2.2 explores how Conceptual Modeling can be improved using ontologies and become one of the fundamentals of KR; in Section 2.3, we go straight to our point by presenting an ontological approach to e-Gov studies; and finally Section 2.4 seeks to situate this approach in the field of applied ontologies and the context of LIS.

2.1. Conceptual modeling: origins and overview

The field of conceptual modeling has its foundations in classical areas from Computer Science, like database techniques, programming language design and, more recently, and artificial intelligence. The need to share ideas and methods between these distinct and related fields led to the creation of the term conceptual model and its valuable and productive research area.

The first data modeling initiatives to represent reality were carried out to create data models that met the requirements of computer data models. We can report a long history from the old sequential files that contain and store data in chronological order to modern Database Management Systems (DBMS)[2]. Throughout this evolutionary process, the modeling task presented, in a broader sense, the same challenges: how to harmonize the characteristics and needs of computer programs and systems with the user's comprehension of the domain of knowledge. The model's evolution follows, approximately, the following steps: data models, semantic models, and conceptual models (ALMEIDA; RODRIGUES BARBOSA, 2009).

In this work, the operational definition of a domain will be applied to the e-Gov as a concept, according to Smiraglia's definition. This definition is derived from various empirical works and is based on the following characteristics to define what a domain is: i) an ontological base with an underlying teleology; ii) a set of common hypotheses; iii) an epistemological consensus; iv) a reasonable consensual semantics (SMIRAGLIA, 2016).

One of the first semantic modeling languages was Chen's Entity-Relationship (ER) model (CHEN, 1976). Modern system designers widely use the ER model for being straightforward, concise, and easy to understand. The ER model is the constructive basis of relational databases, so commonly used in large corporate and government databases around the world. Moreover, it was using the ER model that system designers developed the ability to model data by separating the program logic from the physical design of the database (STOREY; TRUJILLO; LIDDLE, 2015).

We can consider that database technology today is, in a way, stabilized on its applications. Relational databases dominate the market, and schools continually educate professionals with the knowledge of these methods and languages. However, a much more essential and pervasive problem of conceptual modeling still remains. As Smith says, "the first years of conceptual database modeling were mostly marked by *ad hoc* and inconsistent modeling, leading to the many practical database integration problems we face today" (SMITH; WELTY, 2001, p. 4).

Research in programming languages in the 1990s led to the successful development of object-oriented (OO) languages. It quickly became popular among software engineers and is one of the most well-known and widely used approaches today. Using OO, the designer's team has to develop the skill to divide the whole system into small components that combine both data and process. The processes represent dynamic behavior, like business processes and use cases, and it is modeled using algorithm project techniques. The static aspects of the domain are represented by formal structures like classes and components (RUMBAUGH et al., 1991).

---

[2] DBMS is software system that enables users to define, create, maintain and control access to a database.

One can consider that OO has sufficient built-in meaning to classify it as a semantic model. Alternatively, at least it could be a separate category from the data models. Despite the similarities, like objects vs. entities or attributes vs. properties, the languages that support OO do not have the expressivity power required by the conceptual modeling (ALMEIDA; RODRIGUES BARBOSA, 2009).

The OO framework leads the system designer to compose for himself a beneficial and complete mindset to design and build information systems. It is full of support tools, enabling resources, and integration at all levels. We have operating systems, advanced DBMS, high-level languages with their respective Integrated Development Environment (IDE). However, as Smith notes, the crucial and fundamental point remains the same over decades of development.

> […] the field of artificial intelligence was marked by debates between the so-called proceduralists and declarativists. What is the relative significance of process and content (or of procedures and data) in the project of constructing intelligent machines? Proceduralists believed that the way to create intelligent machines was by instilling into a system as much knowledge how as, via ever more sophisticated programs. Declarativists, on the other side, believed that intelligent machines would best be arrived at by instilling into a system a maximum amount of content, of knowledge that – knowledge in the form of representations (SMITH, 2002, p. 23).

2.2. Ontological conceptual modeling and knowledge representation

As seen in the previous section, conceptual modeling is an essential task in the system design process, and its products are crucial to modern information solutions and services. Nowadays, there are new challenges in modeling; the most significant ones are related to the evolution of the web on the Internet to the Semantic Web. A full discussion of Semantic Web is beyond the scope of this thesis, but see the seminal work on this subject, “The Semantic Web” (BERNERS-LEE; HENDLER; LASSILA, 2001).

One problem to address in the modern modeling research is what Smith calls the Database Tower of Babel Problem, as explained in the following citation.

> Different groups of data and knowledge-base system designers have for historical and cultural and linguistic reasons their own idiosyncratic terms and concepts by means of which they build frameworks for information representation. Different databases may use identical labels but with different meanings; alternatively the same meaning may be expressed via different names. As ever more diverse groups are involved in sharing and translating ever more diverse varieties of information, the problems standing in the way of putting such information together within a larger system increase geometrically. (SMITH, 2002, p. 33)

Another problem that has arisen mainly in the field of artificial intelligence is the Closed World Assumption. Of course, a database designer should include in the system data model only the entities and relationships that were mentioned in the system requirements. Another designer does some interviews, or makes direct process observations, possibly collecting these requirements with clients. For practical reasons, it is not possible to include in the database all the facts related to objects of the system domain. Suppose an Internet user makes a query to a public database about something that is not recorded, but that the referred relationship exists in the data model. If the data, as a predicate of the relationship, is not present, the only possible answer from the DBMS will be “data not found.” Sometimes the answer may be in the Boolean format, as false or null. When programming this response, the system designer is making a simplifying assumption in the following format: if the formula is not true in the database, it is therefore false. This simplification precisely reflects the Closed World Assumption. We can summarize this presupposition by saying that the database reflects a particular subset of reality, and there is nothing outside the database. This assumption is both insufficient and inadequate for modern intelligent systems hosted in the cloud (GANDON, 2018).

Furthermore, this assumption is based on the idea that the set (program + database) contains all positive information about the domain, assuming that the database contains not only all the existing instances but also all the possible attributes that an entity can have. Systems built on this premise are much simpler from a programming point of view than others that take as base the representation of the real world (SMITH, 2002). However, this excessive simplification of the real world is an obvious problem when related to the semantic web or artificial intelligence environments, which need another assumption: The Open World.

Table 1 - Synoptic table summarizing the views about ontologies

Distinction	Field	What is it?	Purpose
Ontology as a discipline	Philosophy	Ontology as a system of categories	Understand reality, things that exist, and their characteristics
Ontology as an artifact	Computer Science	Ontology as a theory (logic-based)	Understand a domain and reduce it to models
		Ontology as a software artifact	Create vocabularies for representation in systems and generate inferences
	Information Science	Ontology as a theory (informal)	Understand a domain and classify terms
		Ontology as an informal conceptual system	Create controlled vocabularies for information retrieval from documents

Source: (ALMEIDA, 2013)

> The open world assumption means that we cannot assume something doesn’t exist until it is explicitly stated that it does not exist. In other words, because something hasn’t been stated to be true, it cannot be assumed to be false — it is assumed that ‘the knowledge just hasn’t been added to the knowledge base’ (HORRIDGE et al., 2004, p. 69).

With these two points, the different representations of the same object and the problem Open-world vs. Closed-world, we introduce the need for more powerful instruments to conceptual modeling. By examining these two problems described here, the researchers point out two leading causes, which are shared by both. The first cause is the lack of semantics in conceptual modeling, and the second is the lack of formalism in registering the model and its concepts. Researchers and system engineers have been studying this situation for several decades, and the most promising solution found is the use of ontological conceptual modeling. As we see, every information system already has its ontology, in the format of a conceptual model, because it assigns meaning to the symbols used by a particular view of the domain. What we are discussing here are the benefits of conceptual modeling using domain ontologies in order to avoid problems diagnosed. A meaningful and modern discussion is about the so-called ontology-driven information system, that argues in favor of architectural perspectives where ontologies play a central role in systems and services design (GUARINO, 1998).

Ontology is a broad theme examined in different areas of research and several fields of knowledge. A complete explanation of its origins is outside of this thesis scope. We present, in Table 1, a synoptic table summarizing the views about ontology (ALMEIDA, 2013).

The view that interests us is the use of ontologies as artifacts. Using ontologies as logic-based theory allows us to increase the formalism of design and its requirements. Moreover, using ontologies as an formal conceptual system, we reduce the effects of the Database Tower of Babel Problem, providing the interoperability capabilities.

The third term defined to complete this section is Knowledge Representation (KR). It is a very critical area of research in the AI field of study. Ontologies are one of the formalisms of KR, among others, as semantic nets and frames. The new services and applications of AI are advancing rapidly and are spreading to all communities and collective aspects of daily life. Besides, new AI resources are coming into general use without good legislation and guarantees that limit risks to users.

What makes KR so valuable to AI is the need for a shared set of metadata to support knowledge sharing and exchange on a new interdisciplinary approach, Artificial Intelligence of the Knowledge Domain. The W3C Consortium (W3C[3]) maintains a Community and Enterprise Group to study and develop this theme, the Artificial Intelligence Knowledge Representation (AI KR) Community Group[4]. They present their

[3] The W3C is the main international standards organization for the World Wide Web.
[4] Available at www.w3.org/community/aikr/. Accessed on June 18, 2020.

theoretical and practical subject matter as a significant, broad, and poorly defined domain, increasingly relevant to social sciences and various dimensions of human life, such as politics, commerce, privacy, and government. It is outside the scope of this thesis to examine further details of this subject, but there is a vast literature on it (LANDGREBE; SMITH, 2019a, 2019b), (SAMEK; WIEGAND; MULLER, 2014).

Table 2 - Main concepts definitions

Term	Definition
Conceptual modeling	Conceptual modeling is the activity of representing aspects of the physical and social world for communication, learning, and problem-solving among human users (MYLOPOULOS, 1992).
Ontological conceptual modeling	Ontological conceptual modeling (or ontology-driven conceptual modeling) is the utilization of ontological theories, coming from areas such as formal ontology, cognitive science, and philosophical logic, to develop engineering artifacts (e.g., modeling languages, methodologies, design patterns, and simulators) for improving the theory and practice of conceptual modeling (GUIZZARDI, 2012).
Knowledge representation	Knowledge representation (also called knowledge representation and reasoning - KR&R) is the field of artificial intelligence dedicated to the computer-processable representation of information and is traditionally focused on interesting reasoning patterns and how they can be accounted for semantically and computationally (MYLOPOULOS, 1992) (ALMEIDA; SOUZA; FONSECA, 2011).

Sources: indicated at each table line

KR is a controversial concept, and perhaps we can better understand it by examining five distinct roles it plays.

> 1. A knowledge representation (KR) is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it.
> 2. It is a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world?
> 3. It is a fragmentary theory of intelligent reasoning, expressed in terms of three components: (i) the representation's fundamental conception of intelligent reasoning; (ii) the set of inferences the representation sanctions; and (iii) the set of inferences it recommends.
> 4. It is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance a representation provides for organizing information so as to facilitate making the recommended inferences.
> 5. It is a medium of human expression, i.e., a language in which we say things about the world. (DAVIS; SHROBE; SZOLOVITS, 1993, p. 2)

To clarify the terminology, we show in Table 2 the definitions and their sources, all adopted in this work.

2.3. Ontological approaches to government and e-Government

It is a well-known fact that governments produce and manage a large amount of data. This situation is a big issue for all government bodies since they have heterogeneous databases, most of all, as a legacy from ancient systems. Moreover, in general, there is a weak integration between them. Many difficulties in doing this integration come from the lack of registered meaning of each data item. The distinct development methods and techniques used on each system may be the cause of this non-integration. Another possible reason is the natural evolution of the understanding of various terms used on each technology generation (ATTARD et al., 2015; DING et al., 2011; GEIGER; VON LUCKE, 2012; MORRISON; STEIN, 2014; SANTOS et al., 2017)

On the other way, problems in interoperation within systems are hindering the evolution of public services, which must be digital and universal. General infrastructure incompatibilities or complex issues like laws, methods, and work routines can be the root cause of these problems.

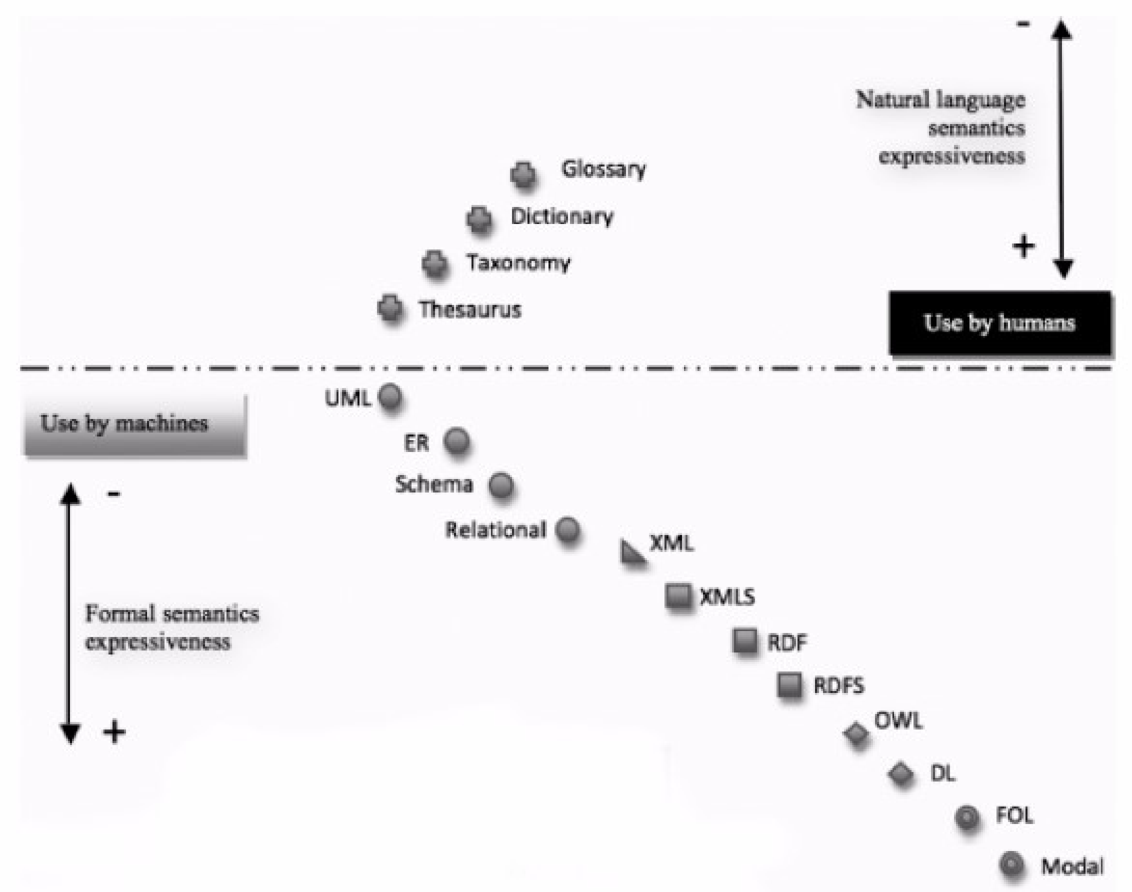
Thus, the studies began using KR approaches to work on two fronts: heterogeneous data integration and system interoperation design. Both solutions are essential to achieve research goals and are an active research area in data integration using ontologies.

The field made substantial progress in this direction in recent years with the first commercial products and tools emerging. It seems that the critical point is to develop intermediate ontologies complete enough to embrace the diversity of domains to integrate. For complete information about this, see Section 4.3.1 − Ontology-Based Data Access.

Therefore, as we are looking for better e-Gov solutions using KR, we need KOS for this area of knowledge. Furthermore, as a premise in our research, we had in mind the purpose of reusing artifacts, publicly available, and not developing a new KOS.

It is outside the scope of this work to discuss the concept of KOS and its broad theoretical foundation. Nevertheless, it is essential to note that KOS has two characteristics that are of importance in this classification of artifacts: semantic expressiveness and ease of computational processing.

Figure 1 - Spectrum of KOS by expressiveness



Source: (SOUZA; TUDHOPE; ALMEIDA, 2012)

Semantic expressiveness is the characteristic that makes it possible to better represent reality in a computer system or language. Moreover, computational processing refers to the greater or lesser ease of understanding, by humans or computers, of the formalized artifact. That is, the more formalization approaches natural language, the easier human understanding gets. Furthermore, possibly, the more difficult computer processing becomes (SOUZA; TUDHOPE; ALMEIDA, 2012).

Considering this reasoning, a two-dimensional scheme of artifacts classification is proposed, reproduced in Figure 1. The "ontology" artifact sought herein is located in the lower right corner of the figure, under the label OWL.

In our initial studies, we thought that the study of the financial aspects of the Governments could be a useful approach since it represents a big part of the transactional activities of their processes and agencies. Then we started focusing on a search for public accessible KOS on the domain of general finance. There are some studies on that, on both academic and commercial proposals (BENNETT, 2013; WANG; FOX, 2016).

Examining these studies, we realized that finance systems are, in general, Online Transactions Processing (OLTP) systems. As all OLTP system, a financial system usually carries on a type of data that is not the most valuable to understand the real world we want to represent. There are pieces of evidence that we need more high-level concepts like government corporations, functions, citizens and their roles, and so on (BENNETT, 2014; EL IDRISSI ESSERHROUCHNI; FRIKH; OUHBI, 2014; KINGSTON; SCHAFER; VANDENBERGHE, 2004; KRAINER, 2012).

Then we change our focus to government KOS, based on the understanding that "government" is a domain of study, by itself. As explained in the SLR in the next section, we found some academic sources on this subject and, in general, they think about the study of government as political science. Government is an academic field of investigation and a vast subject, encompassing political institutions and political theories. There are several sub-disciplines within this area; most of them have a target on another domain of study. To give some examples, one can cite public administration as the human management of government institutions or political economy as the way the government deals with tax issues. Furthermore, there are other similar fields of study, e.g., international relations, political theory, and comparative politics (HENDLER et al., 2012; LAMHARHAR; CHIADMI; BENHLIMA, 2014; LINDERS, 2012; MATHEUS; RIBEIRO; VAZ, 2015; MUSAFIR, 2018; SOUROUNI et al., 2010).

As the research evolved, we were able to realize that this approach is not what we needed to understand government systems and services from an ontological point of view. We were looking for something like "electronic government." Moving forward with the research and the comprehension of the theoretical models involved, we find several attempts to represent e-Gov as domain ontologies and others to represent systems and services as ontologies or taxonomies. Our focus is on how to link these fields to a useful KR perspective.

2.4. Impacts on the research

Studies on e-Gov are fundamental to modern society, given the importance of public services and the possibility of providing better care to citizens in terms of necessary facilities. E-Gov is a broad and multidisciplinary field of study, as there are collective and individual implications in most aspects involving such valuable services.

There are several possible and relevant approaches, from various branches of science to the subject of such an impact, be it sociological, anthropological, economic, or philosophical studies, to name but a few. The approach here is interdisciplinary, with emphasis on Library and Information Science (LIS), which is an applied social science that teaches how to organize the world into categories. An example of this is ontology research, a field of Metaphysics, and its recent application to information systems. An approach called Applied Ontology (MUNN; SMITH, 2013) has emerged as an alternative for knowledge representation in the digital realm, which reveals problems that are different from those faced by librarians and information professionals 50 years ago. In fact, in the late 1990s, research in Applied Ontology conducted by philosophers was widely referenced and brought new possibilities to LIS practices by connecting them to Semantic Web representation languages (COCCHIARELLA, 1980; JOHANSSON, 1989; SMITH, 1995).

Applied Ontology brings together the two uses of the term "ontology": ontology as discipline and ontology as an artifact. While ontology as a discipline refers to a branch of Metaphysics that studies what exists, ontology as an artifact aims to represent the knowledge of a domain (ALMEIDA; TEIXEIRA, 2020). Applied Ontology, as it encompasses both conceptions, has proved to be a convenient theoretical-practical framework for representation purposes, maintaining several intersections with LIS.

From a theoretical point of view, there is much debate when it comes to choosing a theory to guide the procedures for scientific research. The adoption of this or that theory may vary. However, three modes of approach must be present: epistemological, ontological, and methodological (GUBA, 1990). From a practical standpoint, ontologies as artifacts are alternatives for the representation of knowledge in the digital context, which makes use of philosophical principles of ontology as a discipline, using formal statements in computational artifacts.

The original work of Hjørland (2003) connects ontologies with artifacts to LIS classification systems by explaining that these systems can be seen as "types of ontologies." In order for ontologies to be represented in the representation languages available today, they lack theories that allow knowledge to be formalized for inference purposes. Automatic inferences are performed by computers and are the only consistent way to deal with digital vocabularies, common today, that easily surpass one hundred thousand terms[5].

3. Systematic Literature Review (SLR)

We also seek to present and synthesize research on Electronic Government Ontologies to provide an overview of the state of the art. The objective is to identify the current stage of ontology constructions in this domain, which is public, open, and built collaboratively. We believe that identifying these efforts throughout the academic and

[5] See for example SNOMED, a medical vocabulary available on the Internet at www.snomed.org. Accessed on July 30, 2020.

business world is essential for the scientific community to build a shared understanding of the challenges met in this theme.

This SLR is based on the recommendations from Kitchenham (2004) and is organized into six parts:

1. Introduction: Brief presentation of the research and the general objective of the SLR.
2. E-Government: Basic principles of research, delimitations, and motivation for the study of e-Government.
3. Review Planning: The research overview, scope, and delimitations.
4. Performance of the Review: How we carry out the research, the primary sources consulted, and the general methodology employed by the researcher.
5. Report of the Review: Which were the findings, how they were classified, the general selection criteria, and the analysis of the works and work fronts considered most relevant.
6. SLR Conclusions: Summary of the findings, the author's conclusions about the SLR.

The remaining of this section is based on the article "*Ontologias Públicas Sobre Governo Eletrônico: uma Revisão Sistemática da Literatura*" (PINTO; ALMEIDA, 2020a). It is organized as a roadmap to conduct a formal SLR: Presentation, Planning, Conduction, Reporting, and Conclusions. Furthermore, in Section 3.2, we explore the foundations of the e-Gov concept and some of its related definitions.

3.1. SLR Presentation

We are seeking to present and synthesize research on Electronic Government Ontologies. Besides this goal, we will relate an overview of state of the art on this matter. The objective is to identify the current stage of elaboration of ontologies of this domain that are public, open, and built collaboratively. We believe that identifying these efforts throughout the academic and business world is essential for the scientific community to build a shared understanding of the challenges that await it in this theme.

This work adopts the following definition of SLR: "a systematic, explicit, comprehensive, and reproducible method for identifying, evaluating, and synthesizing the existing body of completed and recorded work produced by researchers, scholars, and practitioners" (OKOLI; SCHABRAM, 2010, p. 1).

SLR, as a methodological procedure, has its origin in the medical area, and, more recently, adaptations have been made for other fields of research. Regardless of the area, its objectives are the same. In short, SLR seeks to present a complete and impartial summary of the research topic. This summary is built following previously known, tested and accepted procedures in the area (FABBRI et al., 2012).

There are many reasons to perform an SLR. In particular, three reasons guide the present work, which is probably among the most frequent in modern SLRs (KITCHENHAM, 2004). The first reason is to summarize the existing evidence on the use of ontologies in e-Gov systems. The second is to identify gaps in research, suggesting new research possibilities and contributions. Moreover, the third reason, no less important than the previous ones, is to provide a theoretical framework to position our research activities properly.

When executing an RSL, the emphasis is on the characteristics that guarantee its completeness and impartiality. These qualities are coverage, reproducibility, and reliability, and should be explicit in the RSL planning (KITCHENHAM, 2004).

3.2. Electronic Government – Foundations and Definitions

The acronym e-Gov used for the concept "Electronic Government" is well disseminated as a concept defined in various ways, in diverse contexts. The United Nations General Glossary provides a generic and extensive comprehensive definition: "Collective term for the digital interactions and delivery models between the citizens and their government (C2G), between governments and government agencies (G2G), between government and citizens (G2C), between government and employees (G2E), and between government and businesses/commerce (G2B)"[6].

In the Brazilian context, we have a historical evolution of the concept, presented in a document called Digital Governance Strategy (DGS) (MP, 2016a). At the turn of the 21st century, the Brazilian government presented e-Gov as an effort to prioritize the use of Information and Communication Technologies (ICT) to democratize access to information. In other words, this presentation starts with a recognition of the low use of ICT by the population. The declared objective of the re-signification was to broaden the debate and popular participation in the construction of public policies. Moreover, it was also to improve the quality and effectiveness of services and information.

Recently, due to factors such as the expansion of broadband Internet access and the popularization of mobile telephony, the Brazilian concept was expanded and took on a new name, Digital Government.

---

[6] See tinyurl.com/y3add234. Accessed on July 29, 2020.

> The Digital Government contemplates the amplification of interactivity and political participation in the State processes, as well as the facilitation of navigation and access to government portals and services in favor of integration, transparency and meeting society's demands. Digital Government aligns itself to the objectives of government communication, such as the strengthening of democracy, accountability to society, communication to citizens, generation of the governmental message in the right place at the right time, and interaction with society. (author's translation) (MP, 2016a)

The Internet has become more popular and has acquired a large variety of uses. Because of this growing diversity, other terms have emerged, all associated with the idea of e-Gov. We can cite the following concepts, among others, which can be seen in academic literature and current newspapers:

- Open Government or e-Governance (BORTOLATO, 2014; HENDLER et al., 2012; LUNA-REYES; BERTOT; MELLOULI, 2014; MATHEUS; RIBEIRO; VAZ, 2015; MCDERMOTT, 2010; WIJNHOVEN; EHRENHARD; KUHN, 2015);
- Electronic Democracy (DOS SANTOS BRITO et al., 2014; LINDERS, 2012; WELCH; HINNANT; MOON, 2005; WIJNHOVEN; EHRENHARD; KUHN, 2015);
- Electronic Procurement (ALVAREZ-RODRÍGUEZ; LABRA-GAYO; DE PABLOS, 2014; BULUT; YEN, 2013; FERNEDA; CRUZ, 2016).

In a search for more generality, we also find the term “Government via Web or Web 2.0 Government.” On this concept, we transcribe the following comment from Tim O'Reilly, cited by Fons Wijnhoven:

> Much like its predecessor, Web 2.0, 'government 2.0' is a chameleon, a white rabbit term, that seems to be used by people to mean whatever they want it to mean (WIJNHOVEN; EHRENHARD; KUHN, 2015, p. 31).

Considering the diversity of terminology, we clarify that the concept of e-Gov is structured in this work in three main axes (MP, 2016a): access to information, service provision, and social participation. We need these definitions to delimit our study and to organize the bibliographic findings. A more detailed description of the various uses and historical evolution of the concept, its definitions, and constructive structures can be found in “An ontology of e-Government” (RAMAPRASAD; SÁNCHEZ-ORTIZ; SYN, 2015a).

Initially, one has the impression that there are significant differences in the semantics of data in these three axes - access to information, provision of services, and social participation. A more detailed examination reveals the clarity that semantic modeling can bring unifying visions through the identification of similarities and points in common.

We believe that Knowledge Representation for e-Gov can contribute to improving the understanding of government formation, composition, and management mechanisms in all their spheres and geographical scope (ALLEMANG; HENDLER, 2011). Moreover, in this way, it will provide the various agents involved with a new and enlightening vision through the consensual and precise definition of the meaning of the terms studied. It is hoped that this broadening of understanding can underpin new lines of action and evolution, providing gains in efficiency and effectiveness in government management and improvement in services and governance (ALMEIDA, 2013) (GUARINO, 1998) (ZUNIGA, 2001).

According to a bibliometric study, the concern with semantics or the organization of e-Gov terms and concepts is not a common concern among researchers in the area (RAMAPRASAD; SÁNCHEZ-ORTIZ; SYN, 2015b). For this reason and by the results presented here, there is a significant research opportunity to seek the theoretical basis for new design standards, both in data models and service organization.

The interplay between Ontologies and e-Gov occurs within Knowledge Representation, in which a domain ontology is used to provide vocabulary about concepts, their relationships, activities, and rules that govern them. Another useful intersecting context for this study is a field of research, inserted in the view of ontology as a discipline, called Social Ontology (SMITH, 2003; TUOMELA, 2016). The definition of Social Ontology refers to the human artifacts of the social tract, such as money, property, governments, and nations (ALMEIDA; SILVA; BROCHHAUSEN, 2017). It is possible to discern, through knowledge about the laws that govern the relations between these objects, which of them are involved in each economic phenomenon. Furthermore, this construction can be expressed in a taxonomic way, regardless of the opinions of the agents included. Considering that these objects can be seen and described ontologically as the product of beliefs (conventions and agreements) and material things (objects and processes), we characterize e-Gov Ontology as a social ontology (SEARLE, 2010; ZUNIGA, 1999).

In this study, which covers different scenarios − governments, society, and companies − we identify generic and specific gains in the use of ontologies for each situation. The generic, broader benefits are those indicated by semantic modeling and the construction of domain ontologies. The specific advantages are those that are the most important for each scenario and set of actors.

Listed below in a simplified manner are some of the general gains of this ontological orientation:

- The possibility of reusing ready-made ontologies and public knowledge bases, making adaptations and extensions (FREITAS, 2003).
- The existence of a reasonable number of public ontologies (so-called “off-the-shelf”) available for use, consultation, and adaptations. In the case of e-Gov, this

possibility seems especially valuable for the possibility of intentional reuse of knowledge from other areas, from engineering to social sciences (PINTO; BAX, 2017).

- Online access to ontology servers that, storing thousands of classes and instances, can function as maintainers of shared knowledge integrity, seeking to ensure vocabulary uniformity (MARCONDES et al., 2009).
- The possibility of integration and interoperation of existing databases and legacy systems through the mapping between formalisms of knowledge representation. These databases can make available a vast mass of data stored today in relational databases through a standard access interface (BATES, 2011).

Considering that governments generally use socio-technical systems, there is an essential gain in two significant Management Information Systems (MIS) areas: financial control and accounting (LAUDON; LAUDON, 2011).

The possibility of integrating data models between these areas reinforces the fundamental role of MIS in providing support to government processes and administrative decisions.

We believe that the ontological approach can contribute to one significant MIS project challenge, which is to expand the scope of the management philosophy. By formalizing the knowledge of socio-technical processes, we can meet the strategic requirements to reduce the substantial investments in systems and the long development time.

At the government level, transparency and governance have achieved improvements, allowing for publicity of budget execution and citizen oversight of spending and investments. In this case, the importance of Open Government Data programs should be highlighted (RIBEIRO; ALMEIDA, 2011). It should also be noted that the mere publication of data recovered from large legacy systems is not sufficient to promote reuse and integration. Even in approaches that use technologies such as XML[7] and RDF[8], there is no representation of the meaning of the concepts recorded there. More modern approaches propose the use of upper ontologies (FAÇANHA; CAVALCANTI, 2014).

Of course, there are no single winnings in either area. Several correlations and intersections in the set of terms used in each scenario cause the benefits of the semantic approach to spread. For example, one cannot fail to notice the similarities between the current government and business management, improving financial and

---

[7] Extensible Markup Language (XML) is a W3C standard that defines a markup language for encoding data.

[8] Resource Description Framework (RDF) is a set of W3C specifications used as a general method for conceptual description or modeling of information.

accounting control of great value to the citizen, taxpayer, and user of government services. Other examples registered are the efforts in the elaboration of development methodologies with a semantic approach specific to e-Government (APOSTOLOU et al., 2005) and the search for a “Global City Indicators Finance Ontology” based on the ISO 37120:2014 standard - Sustainable Development of Communities (WANG; FOX, 2016).

Another benefit, quite broad and vital, is the improvement of transparency of governments favoring Public Governance and Democracy. In this aspect, there is a significant relationship to be highlighted: public transparency will be as sufficient as the quality of the data exposed (MADNICK; ZHU, 2006). The higher the variety and quantity of data available online, the more difficult it becomes for shareholders and public finance regulators to make decisions. Du and Zhou (2012) present an ontology-based model for improving the quality of financial data made public by law.

3.3. SLR Planning

This SLR aims to identify and report the ontologies about e-Gov and related works found in academic literature.

The study intends to answer the research question: What are the current stage of development and utilization of domain ontologies on e-Gov?

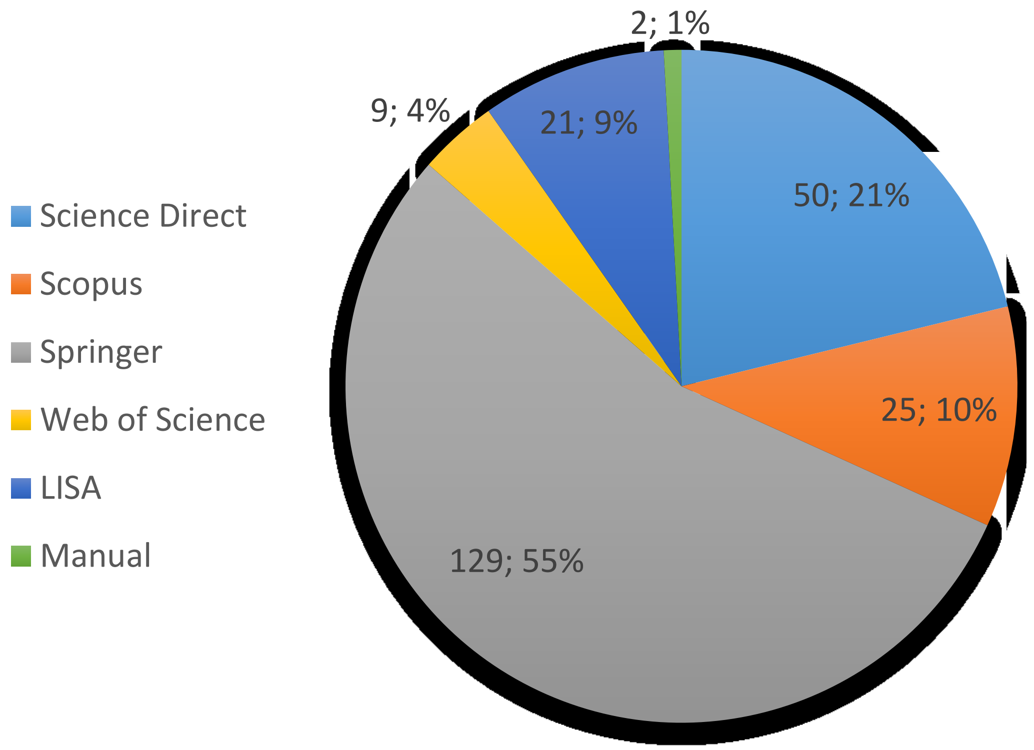
The vision sought in the research is to demonstrate that the efforts employed in the ontological description of the e-Gov domain, a component of the Corporation domain, provide benefits in several use scenarios, namely: access to information, service delivery, and social participation. Moreover, to present the development initiatives of such ontologies, with particular focus on collaborative and public domain projects.

Considering that, in general, works with ontologies are not yet widespread, we find a large number of works that are, so to speak, “in a pre-ontological state.” In other words, we can see the goal of building an ontology, but the researchers are still on the way. Therefore, in this case, the authors choose to present a taxonomy proposal or a controlled vocabulary. Thus, we also include taxonomies and thesaurus in the research. Refer to Figure 1 to clarify the spectrum of semantic artifacts in two dimensions of expressiveness, for use by humans or by machines.

3.4. SLR Conduction

The primary source of references is the theoretical and applied studies of e-Gov ontologies. We are mainly looking for applied studies that may present reusable artifacts. Besides that, we have the general intention to report the most recent works.

Figure 2 - Amount of results per source



Source: The author.

The focus of the searches is the publications accessible by the CAPES[9] bibliographic reference portal, which did not prevent the use of other public repositories (see Figure 2). The first phase, querying the sources according to the search criteria exposed in Figure 2, returned 236 papers.

Table 3 presents the consulted databases with their corresponding query strings. The additional search engines used were Google Search, Google Scholar, and CiteSeerX. We gave priority to publications after the year 2014, which did not hinder the examination of previous works. The review focused on texts in Portuguese and English, without any impediment to the reading of a few cases found in Spanish.

The primary sources used were the following types of publication:

- Review Journals;
- Magazines and Newspapers;
- Books;
- Government Publications;
- Nongovernment publications;
- Conference proceedings;
- Legislation;

[9] CAPES is an educational body of Brazilian Federal Government - Coordination of Superior Level Staff Improvement.

- Research Reports.

The keywords used in the searches are (English – Portuguese):

- e-Government – governo eletrônico
- ontology(ies) – ontologia(s)
- taxonomy(ies) – taxonomia(s)
- vocabulary(ies) – vocabulário(s)

Table 2 - Databases and search strings

Database	Search 1	Search 2
LISA	e-government AND (ontology OR taxonomy OR vocabulary)	"governo eletronico" AND (ontologia OR taxonomia OR vocabulario)
ScienceDirect	e-government AND (ontology OR taxonomy OR vocabulary) year 2014-2019	e-government AND semantics year 2014-2019
Scopus	TITLE-ABS-KEY (e-government AND (ontology OR taxonomy OR vocabulary)) AND PUBYEAR > 2013	TITLE-ABS-KEY ("governo eletronico" AND (ontologia OR taxonomia OR vocabulario)) AND PUBYEAR > 2013
Springer	e-government AND (ontology OR taxonomy OR vocabulary) year 2014-2019	
Web Of Science	(e-government AND (ontology OR taxonomy OR vocabulary)) Allotted time: 2014-2018. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.	

Source: The author.

In general, the most productive search query, with the best recall, was “e-government AND (ontology OR taxonomy OR vocabulary).” The keyword “egov” proved to be very broad, with low precision, but it returned some more generic texts valuable to the research.

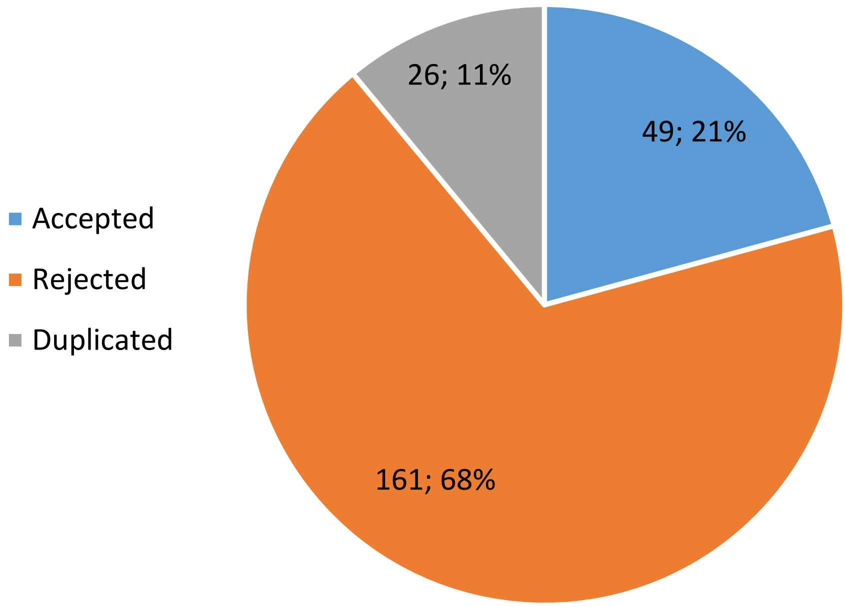
The main selection criteria in the initial phase were the critical reading of the summaries. In this reading, we sought the relationship and correlation between the works, with particular attention to those that present the most appropriate approach to Information Science. In this assessment, we gave priority to works that report a case study, or that present some kind of prototype. Next, we evaluated the quality of publications and selected texts seeking their QUALIS/CAPES qualification and references in other international publications (see overall results in Figure 3).

The reasons for acceptance or rejection of the selected articles are summarized in Figures 4a and 4b, respectively. The letter (I), in parentheses, before the name of the criterion, means it is an Inclusion criterion. Likewise, the letter (E) indicates the Exclusion criteria.

3.5. SLR Report

As a way of organizing the results, we classified the selected papers into three groups according to their primary objective. That is, each of the groups brings together works that have similar objectives. These are the groups:

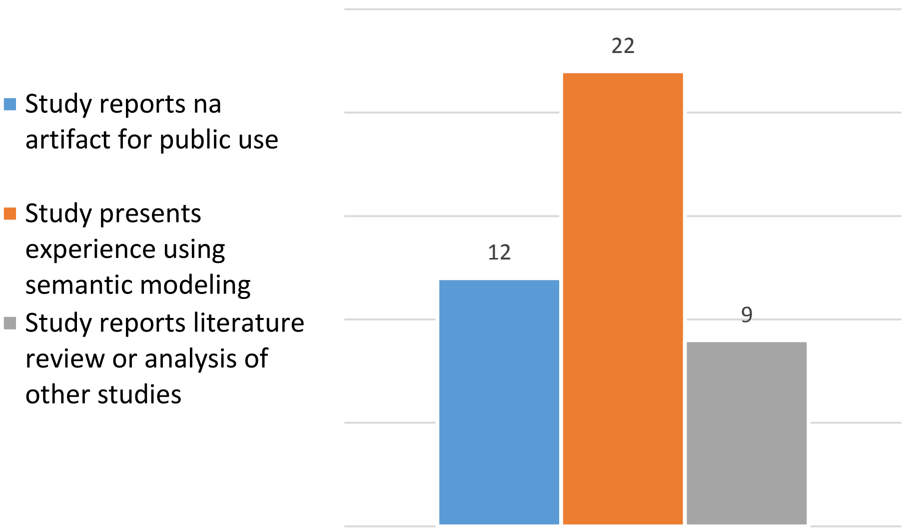
Figure 3 - Initial selection results



Source: The author.

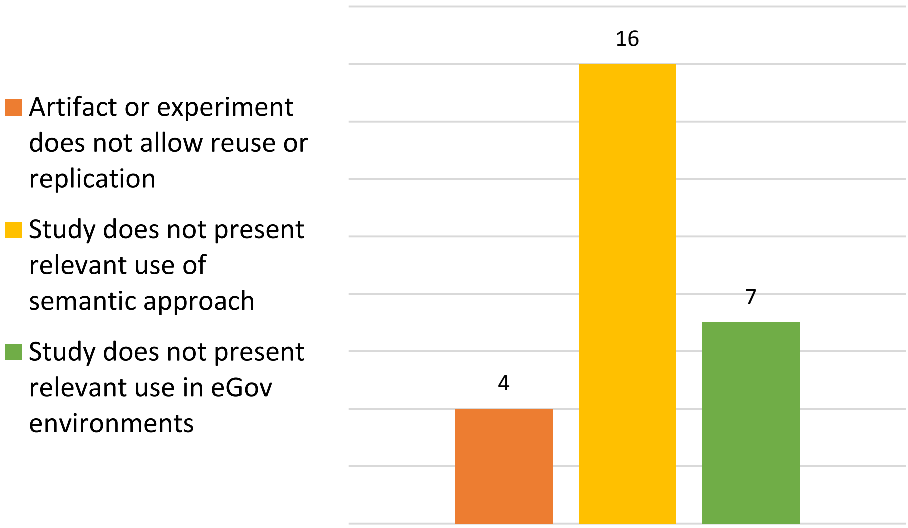
1. e-Gov Ontologies applications, Information Technology - IT application reports - that use ontologies in e-Gov domains.
2. Ontology or Taxonomy standards, proposals, or definitions of standards and norms for e-Gov ontologies.
3. Development methods or experiences, e-Gov ontology development reports with or without methodological proposals.

Figure 4a - Reasons for paper's acceptance



Source: The author.

Figure 4b - Reasons for paper's rejection



Source: The author.

We made this clustering, but it is neither definitive nor unquestionable criteria. It was made to allow the application of the last criteria for defining scope, which was the exclusion of the work of the third group, Development Methods and Experiences. We made this exclusion because of the objective of the SLR, which is “to identify the current stage of elaboration of e-Gov ontologies that are public, open, and collaboratively built.” We recognize the importance of development methodologies and experiences as a source of information about the state of the art. However, our leading idea is the focus of the work on applications and proposals for standardization.

Table 3 - Number of works per cluster

Cluster	Number of works
e-Gov Ontologies applications	11
Ontology or Taxonomy standards	5
Development methods or experiences	8
TOTALS	24

Source: The author.

It is worth noting that there is a sound overlap of objectives between the works studied. For example, it is common for a development report to describe, in addition to the methodology used, the application developed. On the other hand, a standardization proposal includes the evolution methodology as a development and maintenance method. To deal with these dubious cases, we consider the main objective of the work as a classifying factor among the three groups.

Besides the works that are the object of this SLR, we found some articles and theoretical studies from two specific areas: Public Administration and Computer Science. These papers are not the focus of the SLR but are, somehow, part of the analytical foundation itself.

3.5.1. **e-Gov Ontologies applications**

Our research recorded 11 works that report the construction of e-Gov services and systems using methodologies or semantic tools. Classifying the works according to the types of services built, we found “intelligent cities” and “service portal” with two works each. The other types of applications had one work each: purchasing, finance, social networks, real estate registration, social security, public security, and legislation.

On the subject “smart cities,” two works of this study present different perspectives and good results of modeling and implementation of ontologies. Consoli et al. (2017) demonstrate a broad semantic model that integrates diverse and heterogeneous databases, such as geo-referenced data, public transport, urban problems, road maintenance, and waste collection, among others. By using RDF coding and the theories on Linked Open Data (LOD)[10], the work presents, besides the semantic model, a prototype implementation with good experimental results. One can know more

[10] Linked Data is a set of design principles for sharing machine-readable interlinked data on the Web. Linked Open Data (LOD) is a blend of Linked Data and Open Data: it is both linked and uses open sources.

about LOD theory, and its application in e-Gov see, respectively (BIZER; HEATH; BERNERS-LEE, 2009) and (GEIGER; VON LUCKE, 2012).

Another work on the same theme – “smart cities” – indicates exciting possibilities of using semantic models applied to the formulation of Evidence-Based Public Policies (EBPM) (HUNTER, 2015). EBPM is a term used in recent decades and applied in various fields of government referring to situations where rigorously established objective evidence guides the political decisions. The broader objective is to seek a mechanism to develop more effective and efficient government actions based on scientific evidence. In this work, the authors present a high-level ontology that has a consistent and systematized approach to modeling the data and indicators of a smart city. The principle of the modeling work is simple but very hard to achieve. The basis is the consensus about the meaning of the data between two entities: the agencies that produce the field data and the public policymakers. The authors also present a case study with real data and an analysis of possibilities for continuous improvement in policy formulation.

In dealing with the subject “service portal,” there is a fundamental approach, which is the issue of semantic interoperability between services. At the Brazilian moment, this subject is especially important, and we will adopt, as a comparative example with the works studied, some definitions made in the gov.br portal. Public service must have the following characteristics: interaction, customization, sufficiency, standard process, and direct utility to society. The first work studied (ORDIYASA et al., 2017) addresses the difficulties of reconciling all these characteristics with consistent performance of the services offered by the portal. Given the inherent heterogeneity of the services, caused by the age of the systems, old models and formats, evolution of knowledge of the domain, among other reasons, it is a challenging task to build a coherent process to define and monitor the Quality of Service (QoS) indicators of the portal. Using a deep-learning[11] approach, the authors propose an adaptive ontology that can help in the analysis of bottlenecks and maximize the joint performance of the systems accessible in the portal.

Another work on the “service portal” subject focuses on the problem of construction of integration ontologies, not on modeling but on the discovery and registration of links between websites, documents, and the portal (SANTOSO et al., 2016). Through the construction of a web crawler[12], portal administrators traverse government websites and documents, indexing them among themselves, automatically

---

[11] Deep Learning is a branch of Artificial Intelligence studies based on a set of algorithms that model high-level abstractions of data.

[12] A web crawler, or network crawler, is a computer program that searches the Internet, or part of it, in a methodical and automated way, selecting and extracting data, or creating links between pages.

and continuously, based on the concepts of a domain ontology. The authors report significant improvements in portal maintenance costs and query assertiveness.

A semantic application that has shown great interest from governments is the area of government procurement, also known as e-procurement, which is a wide-ranging set of socio-technical solutions that can bring a significant gain in efficiency, agility, and transparency for governments. The selected work presents far-reaching research on current solutions and future trends (ALVAREZ-RODRÍGUEZ; LABRA-GAYO; DE PABLOS, 2014). The authors conclude that governments are the largest buyers of diversified items worldwide. The authors present, among others, the research efforts and solution building by the countries of the European Community (EC). Given its language and social diversity, the EC sponsors one of the most mission-critical projects in this area, the Tenders Electronic Diary (TED). According to the authors' estimates, it accounts for approximately 20% of the member countries' GDP[13]. Another significant issue is the vision of a chain of solutions that the e-procurement process fosters: e-notification, e-access, e-submission, e-evaluation, e-awarding, e-ordering, e-invoicing, e-payment, e-auction, e-certificates, e-signatures, e-catalogs.

In this study, two Brazilian papers were selected, one on finance (CHEVITARESE; BAX, 2014) and another on social security (RIBEIRO; PEREIRA, 2015). In the study on finances, the authors highlight the benefits of using the LOD standard by the Brazilian Congress, exploring the example of the Federal Public Budget. In social security, the authors detect the flaws in the Government Controlled Electronic Vocabulary, which Brazilian acronym is VGCE, when dealing with the classes and concepts involved in the subject of the title. The authors present proposed solutions based on a review of the aforementioned semantic artifact, controlled vocabulary. There are related papers about VCGE (PINTO; ALMEIDA, 2018) (RAMIREZ, 2015) and its official documentation (MP, 2016b).

Another area of applications where semantic modeling has proved to be very promising is legal studies, with several advances in social ontological modeling, integrated with the BFO (SMITH, 2012). The next study explored one recent driver of research in this area, where the authors present a Privacy Ontology according to new laws on the matter (PALMIRANI et al., 2018). The General Data Protection Regulation (GDPR) creates new outstanding obligations for governments when it comes to citizens' data. The ontology proposed in this paper is named PrOnto. It intends to represent the knowledge of the agents involved in the processes of use, storage, and movement of personal data, defining concepts, and their relationships. The high-level objective is to

[13] Gross domestic product (GDP) is a monetary measure of the market value of all the final goods and services produced in a specific period in a country.

create a framework of rules for self-diagnosis and ensure government agencies comply with them.

The use of social networks can pose a threat to citizens' privacy. However, due to its constitution, social networks are a terrain of complicated legislation and restricted government action. The proposal of a semantic model to represent social networks can be a step forward in providing more security to users while providing control tools for governments (KRIJGSMAN; HOFMAN; HOUBEN, 2015).

Enterprise Application Integration (EAI) is a concept that has been developed since the 2,000s (LINTHICUM, 2000; MENTZAS; FRIESEN, 2009). It is a reference to the computational means and system architecture principles used in the process of EAI. By proposing an Ontology for Public Security, Santos et al. (2018) assume that Knowledge Representation has a significant role in EAI initiatives, representing concepts of each legacy system and presenting an "interlanguage" that allows the integration of data and services. In this work, the field studied is Public Security, especially the subdomain of the Violent Crimes against Life.

In 2012, in an international standardization effort, the International Organization for Standardization (ISO) published the ISO 19.152 – Geographic Information Land Administration Domain Model (LADM)[14] – defining concepts and models for Territorial Management. This standardization has significant importance for Municipal Administrations in city management processes. The work under review (ÇAĞDAŞ; STUBKJÆR, 2015) combines this ISO standard with the thesaurus building standard – ISO 25.964 The International Standard for Thesauri and Interoperability with Other Vocabularies[15] – to propose semantic management models for data sets held in public records, academic resources, and laws. The proposal uses the W3C standard Simple Knowledge Organization System (SKOS)[16] and integrates with various international vocabularies currently in use.

3.5.2. **e-Gov Ontologies patterns**

Considering that one of the most important goals of using ontologies is interoperability between services and systems, it becomes highly valuable to create standards that allow increasing the scope of interoperation. In other words, allowing joint operation and data exchange between systems of the same organization is already a high gain. However, the most considerable improvement is to create automatic

[14] See www.iso.org/standard/51206.html. Accessed on July 30, 2020.
[15] See www.iso.org/standard/53657.html. Accessed on July 30, 2020.
[16] See www.w3.org/2004/02/skos/. Accessed on July 30, 2020.

interoperation mechanisms between systems, new and legate, from any origin. These systems can come from different organizations, different countries, and different working cultures. The interoperation concept definition is not straightforward either, as it is applied in many dimensions: technical, semantic, political, human, intercommunity, and international (ALMEIDA, 2002). Because of these characteristics, we can note the greatness of this challenge, which, in our current stage of knowledge, is faced through the creation and use of interoperability standards established by consensus. The broader the scope of this consensus, the closer we are to achieving our goal.

In the case of e-Gov systems, the interoperability capability defines, in principle, the attributes of effectiveness, efficiency, and efficacy of the entire government mechanism operated electronically (RYHANEN; PAIVARINTA; TYRVAINEN, 2014). This work, the first to be reported in this area in our research, reviews the literature on Generic Data Models (GDM) employed in governments. GDMs are modeling artifacts that define central concepts of the systems involved, their attributes, and their relationships. In this work, the authors reaffirm the generally accepted idea that GDMs should be independent of physical implementation and developed in a formal language. The GDMs found are analyzed in three categories, according to their theoretical basis of creation: entity-relationship (ER), metadata, and ontologies. One of the authors' essential conclusions is reproduced below.

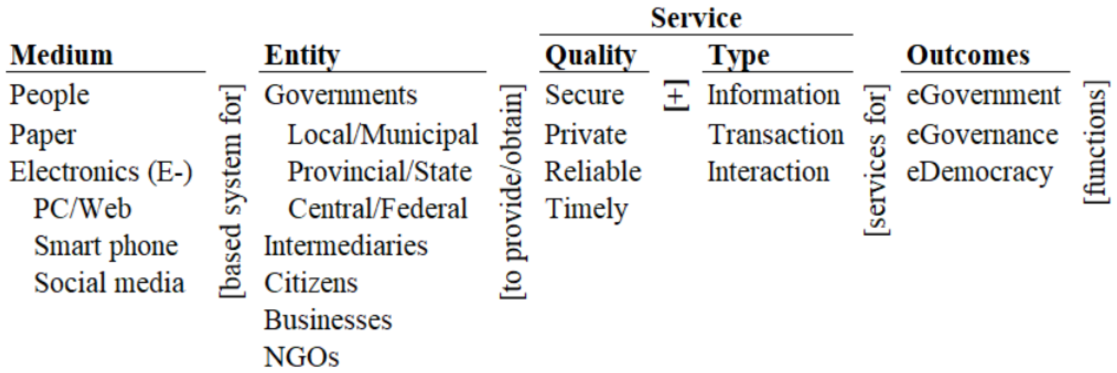
> "The technical orientation brings a significant advantage for ontologies, because they are often presented in machine-readable format and are therefore processable at runtime, reducing the chance to misuse or otherwise incorrectly interpret the data model." (RYHANEN; PAIVARINTA; TYRVAINEN, 2014, p. 115).

The establishment of universal standards is a major technical, political, and social challenge due to the need for a broad consensus amongst developers. The next challenge is the widespread use of the models, which will only achieve their project objectives, justifying the tremendous human effort undertaken in the acquisition of consensus, if they are used to a great extent. The next work review, namely, Van Compernolle et al. (2016), was written two years after the research presented in the previous paragraph. In this review, the authors analyze the state of the art in implementation of GDM, here called the Core Data Model, referring to the same type of artifact. The article lists the existing models and analyzes the most used ones based on four criteria - namely, creation, use, maintenance, and coordination - complementarily examining overlapping concepts between the models. We reproduce below one of the authors' recommendations based on their investigation of the feasibility and use trends.

> On a more high-end level, we suggest to compare information systems and the way European governments (and in this administrations) are structured. In particular, we focus on the aspect of autonomy, coercion and the concept of federalism. Both in public administration research,

> political science and in information management studies, the concept of federations (or federated systems) occur. (VAN COMPERNOLLE et al., 2016, p. 336).

Figure 5 - Ontology of e-Gov Services - illustrative components



Source: (RAMAPRASAD; SÁNCHEZ-ORTIZ; SYN, 2015a)

The practical construction of generic and interoperable e-Gov solutions lacks, as seen, consensus and standardization. Under another perspective, we also need an academic effort to achieve these goals. One of the studies focuses on the proposition of models of a higher level than the GDMs, and that allows a comparative analysis between the implementations of practical models (SÁNCHEZ-ORTIZ; RAMAPRASAD; SYN, 2018). In this model, which had been previously proposed by the same authors (RAMAPRASAD; SÁNCHEZ-ORTIZ; SYN, 2015a), the components are grouped into five basic categories, as shown in Figure 5. This model is proper to systematically enumerate all the possible components, in 1,260 components, of an e-Gov system. By statistically analyzing 453 academic papers, the authors identify all the component clusters in each category, presenting them graphically in a dendrogram. Analyzing these groupings follows the conclusion that state of the art in e-Gov studies presents many gaps and bottlenecks, being focused on public data supply and essential services. There is considerable room for studies and solutions proposals in the most advanced areas, such as e-Governance and e-democracy.

An ontological model developed for the Greek government, Greek e-Government Interoperability Framework (eGif)[17] uses uncomplicated concepts for alignment with other European ontologies (FRAGKOU; GALIOTOU; MATSAKAS, 2014). This model is part of ongoing work to achieve compliance with the Greek Government Services Portal with the guidelines of Interoperability Solutions for Public Administrations, Businesses, and Citizens (ISA2) of the European Community. The method under development, based mainly on Linked Open Data (LOD) technology, aims at enriching and organizing the information presented on the portal.

[17] See www.e-gif.gov.gr/portal/page/portal/egif. Accessed on March 15, 2020.

3.6. SLR Conclusions

In this section, we reported the results of a Systematic Literature Review on the theme “public ontologies on e-Gov.” We intend to introduce application building works and standards proposals in the period 2014 − 2019.

The study used systematic review techniques to identify academic and technological tendencies in the area. There is also an attempt to identify the standards currently used for the reuse of public artifacts. This technique enables dealing with the challenge of grouping studies and drawing representative profiles of trends in academic and practical fields.

The orientation of the work is the realization of descriptive research, which seeks to demonstrate state of the art in a specific aspect of e-Government studies. In the evolution of such opportunities, advances are sought in the area of conceptual modeling supported by ontologies with priority to public ontologies based on high-level ontologies of full acceptance. The LIS contribution is manifested in the modeling quality of ontological approaches, similar to the techniques for building controlled vocabularies.

The findings demonstrate the existence of some outstanding opportunities to use knowledge representation techniques in e-Gov. The use of domain ontologies inserted in a high-level semantic framework composed of well-founded and widely accepted top and intermediate ontologies is particularly opportune.

4. Socio-technical approaches for e-Government

Government is our fundamental subject of study, and the aspect that interests us most is its operationalization by electronic means, the e-Government. This section presents an approach to the government studies that combine two visions, the social vision, here represented by Information Science, and the technical vision that comes from Computer Science and Software Engineering.

On the social aspect, we discussed a crucial issue, which is the nature of government. Can we treat government studies as we do with another widely explored area, which is business administration in general? To better position this question in this thesis, one wonders if it is possible to use the structure of the Social Study to model government corporations.

On the technical side, we present two fundamental components of our experiment: the OBDA techniques and the concept of ‘document’ in D-Acts ontology. By using these components, we built a conceptual model and a demonstration on how to use the theory here presented to mitigate the problem of interoperability among distinct databases.

The remaining of this section is organized as follows: Section 4.1 discusses the ontological nature of Government; Section 4.2 presents the theoretical foundations of Social Studies and their implications for KR, and finally, Section 4.3 relates all details about the technological structure we use in the experiment.

4.1. Government Corporations: Public or Private?

In order to build a useful generalization, we look in this thesis for an ontological model that can cover the main characteristics of government corporations, defined here as an entity composed by the government agency itself and its public bodies. To make a construction like this, we need a well-founded ontological analysis of the nature of government corporations. In particular, we are interested in establishing whether they could be studied like any other corporation. The reason for starting with this question is practical. If we can unify the ontological approach for both government and commercial companies, we can achieve good modeling effort savings through the reuse of standard artifacts and components.

To understand the nature of corporations, in addition to contemporary theories, one can draw on the historical context. Even though early formally organized groups have had a public character since English Crow started establishing them in the 15th century or even before, private variants were later developed. We currently face several forms of institutions, including publicly held corporations, closely-held corporations, limited liability corporations, C corporations, S corporations, professional corporations, non-profit corporations, and so forth. (HODGSON, 2006; IWAI, 2010; KRAINER, 2012).

In this section, we focus on finding the distinctions between two kinds of institutions, public and private corporations. The roots of institutions are briefly explained in Section 4.1.1; this is extended through excerpts of interdisciplinary literature in Section 4.1.2. Finally, in Section 4.1.3, we present our formulation of the distinction, using the speech and document acts theories.

4.1.1. **The Nature of Corporations – a Brief Overview**

This section is based on another published paper that we summarized here (ALMEIDA; RIBEIRO; BARCELOS, 2019). An understanding of institutions as legal entities had already existed when the English Crown began to charter business organizations in the 15th century. The notion of such institutions is a product of Roman civil law, which had already established forms similar to those we see in business today (WILLISTON, 1888).

In this context, institutions have specific core attributes: i) it is a legal unit with its legal rights and responsibilities; ii) it is an entity distinct from the individuals who are members; iii) it acquires legal status by an act of the state. These core attributes were adopted by jurists in the United States, where an institution began to possess additional legal attributes: i) it could contract, sue and be sued; ii) it could acquire and dispose of property; iii) it has its seal by which it could act as a body distinct from its members; iv) its membership may change without affecting its perpetual existence; v) it cannot commit assault; vi) it cannot serve as a trustee (BLUMBERG, 1993).

The classical formulation of institutional attributes has come to be known as the “artificial person” doctrine, one of the several that arose to explain the notion of the corporate personality (BLUMBERG, 1993) (IWAI, 1999, 2007, 2010). The U.S. law maintains that a private or a public corporation must be treated as a person. This extension of rights and obligations from a natural person to a corporation arose from the interpretation of the word “person” in the 14th Amendment.

Debates in countries like France, Germany, and Italy led to the emergence of different theories (MACHEN, 1911): fiction theory, concession theory, group personality theory [or realist sociological theory], bracket theory [or symbolist theory], purpose theory, Hohfeld's theory, and Kelsen's theory. For our current purpose, the essential is the variety of approaches rather than distinctions between these alternatives (ALMEIDA; RIBEIRO; BARCELOS, 2019).

4.1.2. **Interdisciplinary Perspectives**

In the 18th century, the goals of institutions were still perceived as allowing the accomplishment of some public action. Indeed, early U.S. corporations were organized to address public needs, such as bridges, canals, turnpikes, and water companies (BLUMBERG, 1993). It seems that institutions operating in the private sector are in clear distinction to those operating for the public good, as they have different goals, reflected in different management practices, but there are contrary points of view (PINTO; ALMEIDA, 2020b).

From the perspective of Economic and Management Sciences, fields inherently interested in institutions, one sees significant contrasts in stakeholders and sources of funding. By definition, public organizations are funded by the citizens, while private organizations are funded by owners and shareholders. Indeed, issues regarding funding have a considerable impact on governance practices. Here, within Economic and Management perspectives, there are other angles to see corporations.

Table 5 - Differences between Public and Private Organizations

Summary on Differences Between Public and Private Organizations: Main Points of Consensus	
Topic	**Proposition**
I. Environmental Factors	
I.1. Less market exposure	Results in less incentive and lower allocational efficiency. Means lower availability of market indicators and information.
I.2. Legal, formal constraints (courts, legislature hierarchy)	Cause tendency to proliferation of specifications, controls. It allows more external sources of formal influence.
I.3. Political influences	Means diversity of informal external influences on decisions. It leads to a greater need for support of "constituencies."
II. Organization-Environment Transactions	
II.1. Coerciveness	More likely that participation in consumption and financing of services will be unavoidable or mandatory.
II.2. Breadth of impact	Broader impact, greater symbolic significance of actions of public administrators.
II.3. Public scrutiny	Greater public scrutiny of public officials and their actions.
II.4.Unique public expectations	Greater public expectations that public officials act with more fairness, responsiveness, accountability, and honesty
III. Internal Structures and Processes	
III.1. Complexity of objectives, evaluation, and decision criteria	Greater diversity and vagueness of objectives and criteria Greater tendency of goals to be conflicting
III.2. Authority relations and the role of the administrator	Less autonomy on the part of public administrators. Fragmented authority over subordinates, reluctance to delegate.
III.3. Organizational performance	Greater cautiousness, rigidity, less innovativeness. More frequent turnover of top leaders.
III.4. Incentives and incentive structures	Difficulty in devising incentives for effective performance. Lower valuation of pecuniary incentives by employees.
III.5. Personal characteristics of employees	Variations in personality traits, such as higher dominance, higher need for achievement, on the part of government managers.

Source: adapted from (ALLISON, 1992).

Table 5 shows, for example, in the Management field, how Allison addresses the public/private distinction but does not present any fundamental characteristic that differentiates the two (ALLISON, 1992). Boyne reviews theoretical arguments about such differences and identifies hypotheses on the impact of publicness on corporate environments, goals, structures, and managerial values (BOYNE, 2002). Besides, evidence from more than 30 empirical studies regarding differences between public and private corporations are critically evaluated:

> Only three of the publicness hypotheses are supported by a majority of the empirical studies: public organizations are more bureaucratic, and public managers are less materialistic and have weaker organizational commitment than their private sector counterparts […] Whether the existing evidence understates or overstates the distinctiveness of public agencies is therefore unclear (BOYNE 2002, p. 118).

According to Boyne, corporations operating in the private sector have many differences relative to public sector government units and agencies. As described above, the goals are very different, as are administration practices and routines—the kinds of difference very relevant to Management Science. Financial management, on the other hand, identifies as the source of the most significant differences: funding sources and the character of stakeholders. These differences have a considerable effect on accounting and governance practices (ALLISON, 1992; CIEPLEY, 2013; MEIER; O'TOOLE, 2011; RAINEY; CHUN, 2009).

Within the Information Systems literature, one can find a variety of initiatives on international cooperation and research addressing corporate information systems in the context of Semantic Web. In general, such studies do not have clear theoretical concerns with public/private distinctions, even though this has been identified as an ontological approach, e.g., Falbo et al. (2014); European (2015).

The use of conceptual vocabularies such as “public” and “private” often generates much more confusion than understanding because each suffers from theoretical premises with contextual, historical and temporal assumptions and connotations (WEINTRAUB, 1997). Much of the contemporary debate is still characterized by the ambiguity of post-modernism biases, which prevents the explicit development of alternatives (ACKROYD; FLEETWOOD, 2000). There are even lines of thought that reject a clear boundary between private and public institutions:

> Corporations are government-like in their powers, and government grants them both their external “personhood” and their internal governing authority. They are thus not simply private. Yet they are privately organized and financed and therefore not simply public (CIEPLEY, 2013, p. 139).

In addition to fields related to management, Social Ontology offers alternative methods for explicating distinctions between public and private institutions. The analysis of private organizations include two main dimensions: i) the descriptive dimension, which describes how to divide corporations into units and subunits, ii) the prescriptive dimension, which explains duties, obligations, and responsibilities that corporations have to manage (ALMEIDA; SILVA; BROCHHAUSEN, 2017). The former dimension considers the ways people cognitively divide the world by creating departments, sectors, and other forms of divisions within companies; the latter explains how individual duties are related to and are transformed by companies' duties.

Theories introduced in Sections 4.2.3 and 4.2.4 can give support to formulate a new hypothesis. In the context of document acts, one can argue that the creation of a corporation occurs through the recording of spoken rules; then, a document works as an input for a document act. In the next section, we detail these alternatives.

4.1.3. **Private and Public Corporations: Are They Different?**

In the previous section, we examined the nature of corporations through Economic and Management points of view. We also mentioned the existence of ontological alternatives to compare public and private corporations. In this section, we focus on comparisons that can lead to a better understanding of the ontological status of institutions. The remaining of the present section is based on Pinto and Almeida (2020).

In a trivial sense, anyone could view a private corporation as different from a public one. Such differences may seem obvious for some, though not so apparent for others. These ad-hoc views, including some presented here in previous sections, are not suitable for ontological studies that take a more severe and precise approach to the nature of entities as a critical criterion for constructing a classification system (ALMEIDA; TEIXEIRA, 2020). To advance an ontological analysis, here we try to understand the identity of an entity. We do not approach theories about identity criterion here because of their inherent complexity of such issues, but one interested in a contemporary approach can see (DUMMETT, 1981).

The criterion of identity is usually attributed to Frege[18], who asked how one can know whether “a” is identical to “b,” when “a” and “b” are entities (FREGE, 1994). To answer this question from an ontological perspective, one should refer to the properties that objects of the same category share to the extent of being identical. From this perspective, the distinctive property is described in terms of essences — an Aristotelian principle embedded in the computational artifacts we adopt here, namely BFO, Information Artifacts Ontology (IAO), and D-Acts ontologies. In seeking essential properties, we rely on documents and their function in bringing institutions into existence. Before presenting a final analysis, we need to present four preliminary distinctions.

The first distinction held is between natural things and human artifacts; for example, an orange is a “natural thing” and a car is a “human artifact.” In general, to say that something is natural is to say that it is mind-independent – the claim that some

[18] Friedrich Ludwig Gottlob Frege (1848 – 1925), was a German philosopher, logician, and mathematician.

properties of things are constitutively independent of the human – for example, an orange, a person, a tree, or a mountain (KHLENTZOS, 2018). On the other hand, human artifacts are not natural things, since they maintain a trace of human intentions applied to their design, such as cars, hammers, and software. Thus, they also cannot be classified as natural things are classified. In general, a human artifact is not classified by its features, but by its functions: artifacts are intentionally created to serve a function and retain some of the intentionality of the human being who created it (PRESTON, 2013).

The second distinction involves the word "public," which has here the sense of something that can be known to all individuals in society. Artifacts depend not only on one human being or his singular intentional state but on the intentionality that transcends their creator (THOMASSON, 2009). For example, if someone creates a spoon, that spoon is recognized as something used to eat food. One can create a personal "spoon" by folding a piece of cardboard, but even though it can function as a spoon, it would not be recognized as a "real" spoon: an object created for a specific purpose, by adult people, in a particular context. Therefore, although human being's intentional states create a spoon, for the artifact to have success beyond the individual, it must be later associated with public norms.

Thus, an institution is a human artifact describable with two already mentioned dimensions (see Section 4.1.2): the descriptive dimension, which is a Spatio-temporal perspective, and normative dimension, which is a social perspective. The normative dimension is relevant here because documents are artifact and inputs of document-acts, which are later processes able to convey rules and norms, through which people can create and maintain institutions (ALMEIDA; SILVA; BROCHHAUSEN, 2017).

The third distinction, between individual and collective intentionality, develops the need for public norms. Collective intentionality is more than a collection of individual intentional states; this fact explains much of the cohesion of a society. In short, the creation of an artifact is subject to certain norms, because artifacts are recognizable as something that has to be used (i.e., applied) in a certain way (rather than in other ways) by a planned community and in a collective context (DIPPERT; LYAS, 1994). Considering the scope of human artifacts created for ordinary purposes, we follow Thomasson in considering the existence of public artifacts as the result of public norms in a process highly dependent on cultural aspects, despite the intentionalities involved (THOMASSON, 2009). A full discussion of intentionality is beyond the scope of this thesis, but see (SEARLE, 1995), (GILBERT, 1992), and (TUOMELA, 2016).

The fourth and last distinction addresses regulative versus constitutive rules (SEARLE, 2010), which we employ to better explain the notion of a "public norm." The above rule, which merely regulates, reflects acts performed independently of the rule.

Constitutive rules not only regulate behaviors but also create the possibility of the existence of the same behavior they regulate. For example, “driving on the right side of the road” is a regulative rule that limits driving behaviors, in which the act of driving is separated from the rule. On the other hand, there is no possibility of the game outside of chess rules; that is, the rules of chess creates the possibility of chess.

Searle uses this distinction to explain how institutions are created by institutional facts that only exist through collective intentionality, in the scope of a complex network of constitutive rules (SEARLE, 2010). For example, “John is a driver licensed to drive in the U.S.” is an institutional fact that can exist only within a constitutive rule system created and collectively accepted within the US. One relevant last consideration of Searle's framework is the kind of constitutive rule called a “standing declaration”: a declaration that has a doubled direction of realization. Searle explained the notion concept of double direction within speech acts, but Smith extends it to documents (SEARLE, 2010; SMITH, 2012). As a document, in one direction, it can make a thing come to existence; a thing, on the contrary direction, can make changes in documents. For example, following the direction of realization human-mind to the world, a blueprint brings a building to existence; the building (follow the opposite direction, from human to mind) make changes in the original blueprint by replacing it with an “as-built” blueprint.

This framework allows us to elaborate comparisons between public and private corporations, which constitutes the remainder of this section. Our attempt considers documents as the core entities of analysis since once they are taken as inputs of acts, they can assign social obligations to people, both individuals and groups (SMITH, 2014).

First, we can see that institutions, public or private, are human artifacts, as they depend on human intentionality for their creation. As human artifacts, they must be public in the sense that they should be recognized as something associated with social use. To the same extent that a spoon could be made of cardboard (i.e., defined purely by function), one can also create a private corporation and maintain it for only personal use. In contrast, a public corporation cannot exist as a secret, and by definition, it cannot be a private tool.

Secondly, standing declarations can make something happen by representing the possibility of the very same thing happening again. For example, the existence of someone who satisfied the condition of being the oldest son of a dead king, in the context of Medieval Europe, brought a new king into existence (SEARLE, 2010). Similarly, standing declarations also explain the creation of institutions: one declares that anyone who makes a declaration of a specified kind (in the context of a complex system of rules) will have constituted the existence of a corporation. Institutions, however, cannot be addressed as it was in the case of kings in old Europe because they require explicit rules in the context of a complex legal structure, and thus are dependent on written language. Documents have the power to accomplish these requirements and, documents capable of this in social contexts are documents capable of document-acts. They work as inputs for document acts that can create institutions.

Finally, the kind of document-act that creates a public corporation is different from the kind that creates a private corporation. While the former is a law, the latter is a statute of the corporation. Then, if we should point one distinctive and unique property of a document that makes evident the difference between public and private corporations, we choose the deontic powers embodied in the documents that create and maintain an institution. Besides, while a deontic power that creates public corporations emerges from society as a whole, the deontic power that creates private corporations emerges either from a small group or from only one individual.

4.2. The Social Studies Framework

Ontology is a term that originated in the German 16th-century philosopher, Rudolph Goclenius the Elder (1547 – 1628). This scholastic philosopher coined the term to refer to metaphysical studies dating back to Ancient Greece. Since the 1990s, the use of the term within Computer Science and Information Science lacks specificity as "ontology" is used to name both models as well as the theoretical principles that underlie those models (PINTO; ALMEIDA, 2020b).

A relevant notion within ontology studies is the distinction between "Spatio-temporal ontology" and "social ontology." Spatio-temporal ontology corresponds to an attempt to classify both natural categories and the relations by which entities are tied together (MUNN; SMITH, 2013). In contrast, the approach of ontology as a theory for social entities, which is social ontology (SEARLE, 2010).

In the remainder of this section, Sections 4.2.3 and 4.2.4 address ontology as a kind of social theory.

4.2.1. **The spatiotemporal foundation: BFO as theory**

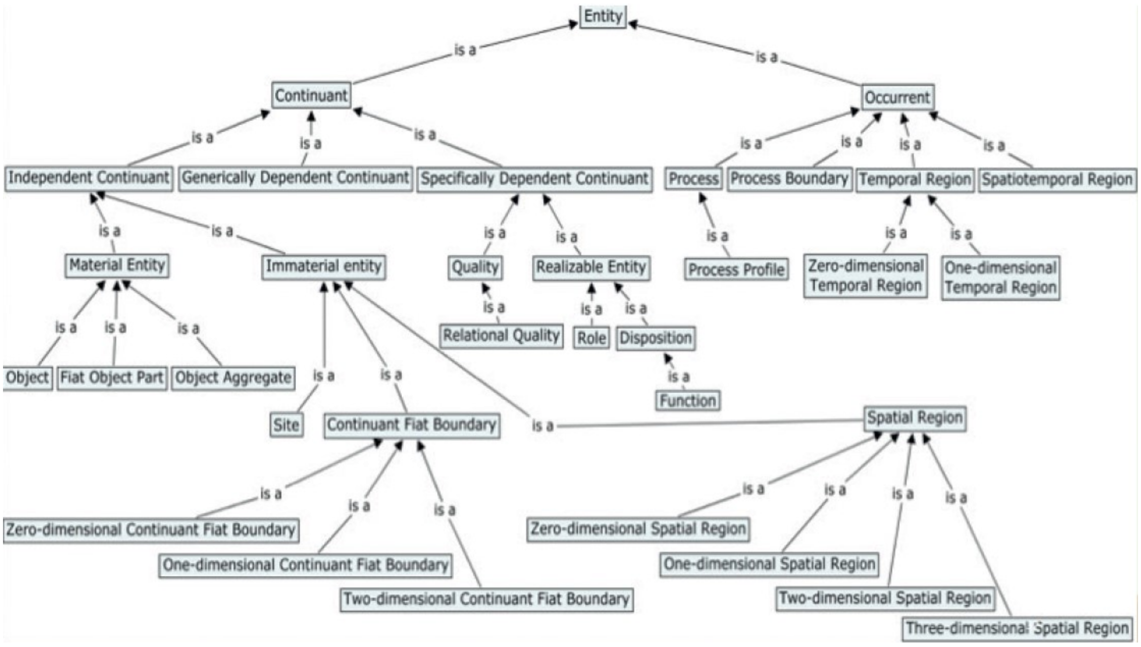
Basic Formal Ontology (BFO) is the fundamental basis of this thesis. BFO is a top-level formal ontology designed to be used to support various types of information systems applications in distinct areas such as information retrieval, scientific analysis, and socio-technical systems. For complete information about BFO, see its home page[19] and Barry Smith academic page[20].

All the following explanations come from the seminal work "SNAP and SPAN: Towards Dynamic Spatial Ontology" (GRENON; SMITH, 2004).

Using a definition closer to philosophy than to software engineering, we say that a formal ontology is domain neutral. It means that it contains just those most general terms, such as object and process, which we can apply in all scientific disciplines whatsoever. A material, or domain, ontology is called domain-specific because it contains terms that apply only in a subset of disciplines (ARP; SMITH; SPEAR, 2015).

These formal characteristics of an upper-level ontology bring to the conceptual ontological modeler to a set of fundamental features for his work, such as indefinite expandability, consistency with increasing content, independence of content and context, accommodation of any kind of concept and the possibility to use different levels of granularity (HIRST, 2009).

Figure 6 – BFO top level ontology, its levels and categories



Source: (ARP; SMITH; SPEAR, 2015)

The BFO is adopted as the starting point of our construction because it is the basis for the D-Acts, which is the intermediate ontology that we will use. The BFO is composed of some levels of well-characterized entities (see Figure 6). We will describe here some levels, only those necessary to understand the essential principles of the

[19] Available on basic-formal-ontology.org/. Accessed on June 19, 2020.
[20] Available on ontology.buffalo.edu/smith/. Accessed on June 19, 2020.

BFO, as well as its central premises. The BFO normative rules are formalized with a set of 16 good practice recommendations[21].

The first level of the BFO has the generic name of the entity. The second level has two distinct groups of entities. On one side, it considers substantial entities called continuant; on another side, it considers procedural entities called occurrents. Continuant ones last over time, maintaining their identity. Examples of continuants are a person, a machine, an orchestra, and a law. Occurrents happen and develop over time. Examples of occurrences are the breathing and functioning of a body organ, a part of one's life, and a business process governed by specific rules.

The third level of the BFO contains three categories in the continuum branch: (i) independent continuants, (ii) specifically dependent continuants, and (iii) generally dependent continuants. Independent continuants are the bearers of qualities, some qualities that they inherit, the red color inherited by a tomato. Specifically dependent continuants are entities that depend on one or more specific independent continuants for their existence, for example, the pain in my head depends on me, and the role of a professor in a university depends on a person. Genetically dependent continuants are entities that also depend on independent continuants, but in contrast to specifically dependent continuants, the body that functions as a carrier may change over time. One example is Shakespeare's Hamlet, which has many copies bound.

Under occurrences, the third level of the BFO contains four classes: (i) processes, (ii) process boundaries, (iii) temporal regions, and (iv) space-time regions. Processes are entities that are unfolded in time, have temporal parts, and always maintain a relationship of participation with independent continuants. Examples are the digestion process, the course of a disease, and the legislative process of proposing and voting on a law.

The fourth level of the BFO continuum contains essential categories, such as (i) material entities, e.g., objects and aggregates; examples of objects are an apple and a mountain; examples of aggregates are an orchestra and a pile of stones; (ii) qualities are properties of entities, e.g., the color or smell of something; and (iii) realizable entities are entities whose instances contain periods of updating in the course of their existence, e.g., the role of antibiotics in curing disease and the disposition of people to grow. The fourth level of occurrences also contains categories to represent the world, such as process profiles and temporal regions.

[21] See www.obofoundry.org/principles/fp-000-summary.html. Accessed on August 11, 2020

4.2.2. **Essentials of Social Ontology**

The history of philosophy registers that the first movements to explain social acts, no matter how they are spoken or written, comes from Reinach[22] on his works analyzing civil law from a phenomenological point of view (SMITH, 2012). In his initial writings, Reinach characterizes the spontaneous acts that he defined as the set of experiences self-directed. In other words, these are the type of experiences a person may have in which he has active participation. Moving on this characterization, Reinach stated there is a type of action that refers to another subject and must be perceived by this subject. They called it "non-self-directable acts" or, more specifically, "social acts" (ALMEIDA; SILVA; BROCHHAUSEN, 2017). In the next sections, we will examine two subtypes of social acts, namely "speech acts" and "document acts."

Reinach developed, in 1913, the first systematic theory of performative uses of language, not only in promising and commanding but also in warning, entreating, accusing, flattering, declaring, baptizing, and so forth. Reinach called these phenomena social acts, a term that remains unchanged until now. Reinach applied his method, especially on the act of promising, as a legal phenomenon such as a contract. As a former student of law, he described his theory as a "contribution to the general ontology of social interaction" (SMITH, 2003, p. 6).

Based on the previous work of Reinach and the Aristotelian conception of language, Austin[23] moves ahead of the studies of statements and propositions. Austin, in 1946, said that, when one says, "I know that S is P," he is saying, "I give others my authority for saying that S is P." Austin called this type of sentence, "performative utterances."

> Performative utterances are those uses of language, often involving some ritual aspect, which are themselves a kind of action and whose very utterance brings about some result. Of an utterance like "I promise to mow your lawn" we ask not whether it is true, but whether it is successful (SMITH, 2003, p. 7).

Following Austin, Searle[24] provides a theoretical framework within which the three dimensions of utterance, meaning, and action involved in speech acts could be seen as being unified together. Searle's general theory has three components of a significant role: rules, meanings, and facts. Moreover, there are two types of rules: i) Regulative rule: regulates antecedently existing forms of behavior; ii) Constitutive rule:

---

[22] Adolph Bernhard Philipp Reinach (1883 – 1917), was a German philosopher, jurist, and law theorist.

[23] John Langshaw Austin (1911 – 1960), was a British philosopher of language.

[24] John Rogers Searle (1932) is an American philosopher, with contributions on philosophy of language, philosophy of mind, and social philosophy.

creates or defines new forms of behavior. Constitutive rules have the basic form: X counts as Y in context C. These type of rules rarely occurs alone. Thus, we may have to say “acting in accordance with all or a sufficiently large subset of these and those rules by individuals of these and those sorts counts as playing basketball” (SMITH, 2003, p. 9).

The central hypothesis of Searle's book, as explained by Smith: “speech acts are acts characteristically performed by uttering expressions in accordance with certain constitutive rules” (SMITH, 2003, p. 9). The difference between merely uttering sounds and performing speech acts is to mean something by an utterance. Then, this leads us to the conclusion that all speech acts involve both a speaker and a hearer, returning to Reinach's original statement about non-self-directable acts.

At this point, Smith explains how Searle built the concepts of promising and obligating. These two concepts are so closely related that when we exemplify one of them, we are at the same time explaining the other. Searle gave these concepts new relevance and outlined the legal consequences of promising, using his theory of constitutive rules. The constitutive rules allow us to perform certain singular activities. Under them, our behavior can be interpreted by ourselves and others in terms of certain exceptional types of institutional concepts. The use of constitutive rules as a human behavior-guiding system is a central point to understand our social conduct as language-using animals. These institutional concepts are the key to understand what we characteristically call human behavior.

Institutional fact is a fact whose existence presupposes the existence of specific systems of constitutive rules called institutions. Institutional facts exist because we, as human beings, treat the world and each other in a particular cognitive way within institutional contexts. Institutions are observer-relative features in the world, in contrast with observer-independent features that are equal to natural facts. Every institutional fact thus lies beneath a system of constitutive rules to form ‘X counts as Y in context C’. Pertinent examples of ‘what an institution is’ means they are systems of constitutive rules, they are money, property, marriage, and government (SMITH, 2003).

This brief introduction to social ontology needs one more concept, the collective intentionality to meet our goals. We can accept a definition that “[…] intentionality distinguishes a property of mental phenomena, namely, the property of being directed to an object, real or imaginary” (ALMEIDA; PESSANHA; BARCELOS, 2018, p. 10). As noted by Smith, Searle accepts what has been called “primacy of the mental,” acknowledging that “language is derived from intentionality and not conversely” (SMITH, 2003, p. 16). Thus, as we have individual and collective intentionality and, in general, they are diverse but convergent, we can say that the collective intentionality is directed to build and formalize, through language, our system of constitutive rules.

> Ever since the Cognitive Revolution, Sapiens has thus been living in a dual reality. On the one hand, the objective reality of rivers, trees, and lions; and on the other hand, the imagined reality of gods, nations, and corporations. As time went by, the imagined reality became ever more powerful, so that today the very survival of rivers, trees, and lions depends on the grace of imagined entities such as gods, nations, and corporations. (HARARI, 2014, p. 29)

4.2.3. **Speech Acts**

The theory of “speech acts” was proposed by Austin as a method for analyzing philosophical questions according not only to the truth of sentences but also to the use of ordinary sentences to perform acts. For Austin, one can “do things with words”, for example, when one uses several forms of utterances to ask things to other people (AUSTIN, 1962). These are forms of speech acts, the basic units of meaning, constituted by three connected dimensions of acts: locutionary acts, illocutionary acts, and perlocutionary acts.

The most critical dimension is the illocutionary act because it contains an element called illocutionary force that represents the kind of act performed. For example, in the proposition “I promise to pay you tomorrow,” there is an utterance (the verb promise) that constitutes the act of promising rather than the description of mental states. When a person utters the sentence, the promise is concretized; in other words, the force that characterizes the act is the promise (ALMEIDA; SILVA; BROCHHAUSEN, 2017).

The result of a speech act is the combination of: i) a proposition, which can be true or false; ii) the semantic content, which mirrors facts of the world; and iii) the illocutionary force, which concretizes the content. This theory provides a classification for speech acts theory and adds seven components to the illocutionary force. The illocutionary force aims to determine the commitments, that is, the relationships established, in addition to expressing how the content is related to institutional environments. To specify the force of a speech act, consider the following characterizations (SEARLE; VANDERVEKEN, 1985):

- The point that represents the purpose of an act: whether it is an assertive, a commissive, a directive, a declarative, or an expressive act. (e.g., a commissive point commits one to do something, as in promising.)
- The degree, which corresponds to the strength of the point: for example, assertives are represented by a sequence of verbs (assert, claim, state, deny, argue, suggest, and so forth) in which “assert” produces an assertive stronger than the “suggest.”

- Content conditions, which are required by the propositional content so that a speech act be achieved. For example, invoicing can only refer to payments and not to salaries.
- Preparatory conditions, which provide the success of the act. For example, in placing an order, the buyer presupposes that the supplier sells specific products.

The importance of speech acts in institutional contexts relies on their ability to bring about new social entities as obligations and claims, to which promises and orders give rise. In Searle's theory, institutions are created by a simple function: X counts as Y in C, where X is an entity, Y is the status of such entity in a context C (SEARLE, 2010). For example, Trump (X) counts as president (Y) in the United States (C). Institutions are a particular case of this formula in which an issue emerges: there is no member X as a physical entity. A divide, called a “standing declaration,” solves this issue: they are of a specified kind that allows one to make something happen simply by promising that it will really happen. In the case of institutions, executing and fulfilling the statute of incorporation X counts as the creation and maintenance of a corporation Y, within a specific legal code C (SMITH, 2003).

Institutional facts emerge when human beings approach each other and the world in specific ways within various unique contexts. Such specific ways depend on cognition and provide the means to build all institutional settings. Institutions are, therefore, observer-relative features of the world in contrast with “natural facts,” which are observer-independent features. A system of rules in the form “X counts as Y in context C” sustains every institutional fact that people produce and recognize within institutions, just as they create institutions as themselves observer-relative contexts.

Thus, the institutional reality is created through speech acts, by which duties and obligations are delivered throughout an institution, unfolding new forms of social interactions. However, the evanescence of speech acts, a consequence of their inherent orality, limits their possible impacts on institutional environments.

4.2.4. **Document Acts**

While speech acts exist only in the moment of their performance, documents can also convey acts, allowing them to persist in time even while absorbing modifications through the document's life cycle. This is the basic premise of the theory of “document acts” (SMITH, 2012): promises and obligations can be established through speech acts in small face-to-face communities, but such compromises cannot be maintained in large impersonal societies. Indeed, obligations transcend local character; human memories are not sufficient to assure the fulfillment of promises in all social contexts.

The Theory of Document Acts describes how people and institutions use documents to engender new social entities. A discrete document, as an input of a document-act, is the bearer of social and institutional powers (ethical and legal) that contributes to bringing about a variety of social effects. A document-act is, in turn, a kind of process. Examples of documents and the respective social entities created through their associated document-acts are presented in Table 6. For example, a contract (i.e., a document) generates an obligation (e.g., to pay some amount of money) through a document-act (e.g., the process of signing the document); a statute of a corporation (i.e., a document) brings a corporation into existence (e.g., with the obligation of delivering specific product or service) through a document-act (e.g., the process in which local government stamps the statute of a corporation); and so forth.

These kinds of documents not only record information but also create social entities and make new types of social arrangements possible. All of them are persistent entities that maintain their identity over time and throughout their history. They can be manipulated, archived, destroyed, signed, registered, inspected, or transferred. In this sense, documents make it possible to use endless kinds of social relations that work as extensions of human memories (SMITH, 2014).

Table 6 - Input Document and Respective Social Entities Created

Document	**Entity created**
contract	obligation
statute of corporation	corporation
deed	privilege
title deed	property right
patent	exclusive right
statement of accounts	audit trail
marriage license	bond of matrimony
stock certificate	capital
diploma	qualification
registration of baptism	legal name
insurance certificate	insurance coverage
license	permission
IOU note	obligation to pay

Source: SMITH (2014)

The relationship between speech acts and document acts involves the commitment between the two kinds of acts. For example, the commitment of a contract, which was only uttered in speech acts, cannot survive in this complex society. The modern institution requires a kind of document that remains the same over time, working as an input of acts that fulfill the obligations in different situations.

We observe that this notion of “document” does not challenge any established theories of Library & Information Science (LIS), in which documents are central and valued subjects. Some documents may indeed be used to record information. However, we address here those additional functions that create claims and obligations for people, as well as new social facts in society (PINTO; ALMEIDA, 2020b).

4.3. The Technological Framework

In this section, we present the two main technical components of our experiment: the OBDA techniques, and the D-Acts ontology. Both of them are public access artifacts and work in a complementary way to execute the database integration.

The experiment that we describe in Part II of this work is entirely founded on these components. To perform the integration of heterogeneous data, we use the Ontop, a well-known integration framework, through an integrated development environment provided by Protégé, an ontology editor (MUSEN, 2015). To build our integration ontology, we extend the D-Acts ontology, one of the components of the OBO Foundry framework.

4.3.1. **Ontology-Based Data Access**

Interoperability between information systems has been a problem to be solved in the planning of governments around the world. This concern is based on two needs that are present in several countries: the provision of digital public services and the co-existence of large legacy databases. Governments seek solutions for the interoperability of the systems, which enable the integration of data and services. Our focus here is not on problems related only to big government data, but on how to represent these data as valuable and useful knowledge. In this section, we seek to clarify the role that ontologies can play in building alternatives to solve this problem. We adopt the view that ontologies, as a knowledge organization system, have the potential to participate in a solution called OBDA (ALMEIDA, 2002; GUARINO, 1998; SMITH, 2002).

OBDA is a model based on the use of knowledge representation and reasoning techniques to manage the resources, data, metadata, services, and processes of contemporary information systems. OBDA provides tools and techniques to enable uniform access to data stored in multiple and heterogeneous sources. The main objectives of OBDA are i) To provide a high-level conceptual understanding of the data; ii) To provide a useful vocabulary for querying; iii) To enhance incomplete data with the knowledge of ontology. A full explanation of OBDA is out of the scope of this thesis, but there is a considerable literature on this subject available (BIENVENU et al., 2013; KOGALOVSKY, 2012; SKJÆVELAND et al., 2015).

Considering the problem of integration of legacy relational databases, we have, in principle, two alternative solutions (SAVO et al., 2010). One solution is to make a complete reengineering of the systems, the design of a new unified database that meets all the applications and the replacement of the original applications. This approach, in general, is not feasible due to organizational and cost problems. The other solution is, from the standpoint of standard software engineering, the creation of a new information system executing the task of communication with old systems and having hard-coded integration features. This new system will be configured as a software layer typically with two components. The first component is a mediation scheme, representing the new structure using the user's view and language. The other component is a mapping that relates the source data to the elements of the mediation scheme. To perform the work with this software, called firmware, we have two methods, materialized and virtual triples. In the first method, we will build a new database to be filled with data from legacy systems. Using the virtual method, we will not move the data or replicate it, and the queries made to the system are answered by directly accessing the sources. In this work, we adopted the virtual approach, considering it preferable in dynamic scenarios such as the government, where data frequently changes, and users need up-to-date information.

In the practical part of this work, we choose to use a free OBDA system that has a lot of examples, templates, and tutorials on the Internet. We are using Ontop, from the Free University of Bolzen-Bolzano (ANTONIOLI et al., 2013, 2014; CALVANESE et al., 2017; RODRIGUEZ-MURO; KONTCHAKOV; ZAKHARYASCHEV, 2013).

> The Ontop system exposes relational databases as virtual RDF graphs by linking the terms (classes and properties) in the ontology to the data sources through mappings. This virtual RDF graph can then be queried using SPARQL by translating the SPARQL queries into SQL queries over the relational databases. This translation process is transparent to the user. The architecture of Ontop, […], can be divided in four layers: (i) the inputs, i.e., the domain-specific artifacts such as the ontology, mappings, database, and queries; (ii) the core of the system in charge of query translation, optimization, and execution; (iii) the APIs exposing standard Java interfaces to users of the system; and (iv) the applications that allow end-users to execute

> SPARQL queries over databases. We explore each of these components in turn (CALVANESE et al., 2017, p. 2).

At OBDA architectural design, we need to build a conceptual layer as an ontology. This artifact, called integration ontology, will define shared concepts and model the domain of knowledge. It brings the user the additional value of hiding the structure of the legacy data source. The user makes queries over this high-level conceptual view and can be carefree about formats, physical organization, and attributes of legacy data. SPARQL queries are translated by the OBDA system into SQL language and submitted to the originating DBMS. The OBDA components tie together the ontology and data sources through a declarative specification, called RDB2RDF mapping (relational database to resource description format). The designer writes the mapping using the standard W3C R2RML language, and this artifact has as output a virtual RDF graph, a virtual graph materialized and stored at an RDF triplestore. However, materialization operation may be costly and does not take advantage of the persistency mechanism already available in our mature and reliable DBMS (CALVANESE et al., 2017).

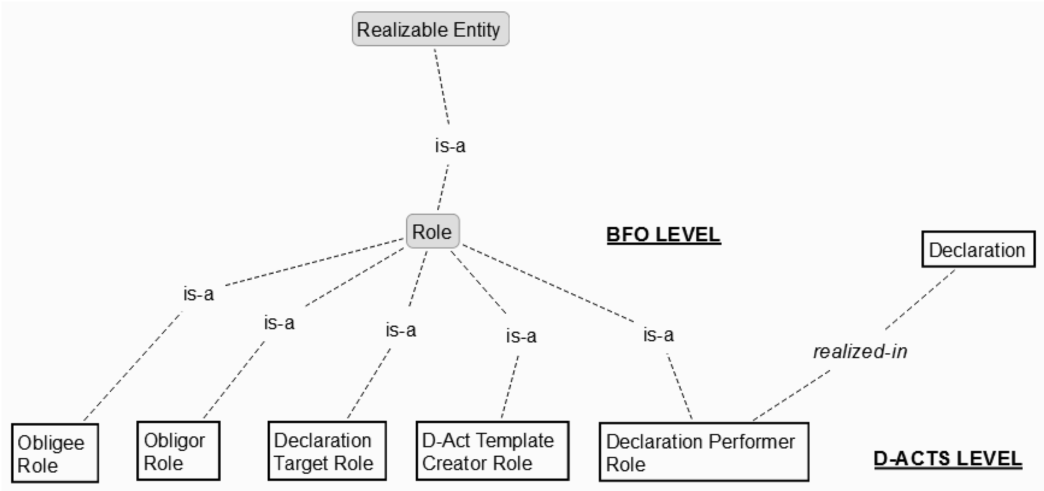
4.3.2. **Documents in Documents Acts ontology**

The Document Acts theory can be implemented and operationalized for practical purposes through "document acts ontology" (hereafter D-Acts ontology) (BROCHHAUSEN; ALMEIDA; SLAUGHTER, 2013). This ontology uses pre-existing resources developed based on the BFO (SMITH et al., 2007), for example, the IAO. While IAO provides alternatives for representing information artifacts as documents that record information (CEUSTERS, 2012), D-Acts ontology represents another kind of document that is used as an input of a document act to deliver obligations. For the sake of simplicity, we do not present BFO and its tenets here, but a full account about it can be found in Barry Smith's works (ARP; SMITH; SPEAR, 2015; GRENON; SMITH, 2004). Likewise, we do not present a complete account of D-Acts Ontology, but the key definitions are listed below, while the classification scheme is represented in Figures 7 and 8.

- Declaration Performer Role: a role inhering in a human being or an institution that is realized by the bearer acting as the agent in a declaration.
- Document-Act Template Creator Role: a kind of role, inhering in human beings or institutions that prepare a document to be the specified input to a document act.
- Social-Act: a process carried out by a self-conscious being; it is spontaneous, in the sense that it is directed toward other conscious beings and needs to be perceived.

- Declaration: a social act that brings about or revokes social-acts; declarations are sometimes actions, for instance, the signing of a document.
- Document-Act: a declaration performed using a document to temporally extend the effects of the declaration.

Figure 7 - Roles in D-Acts Ontology



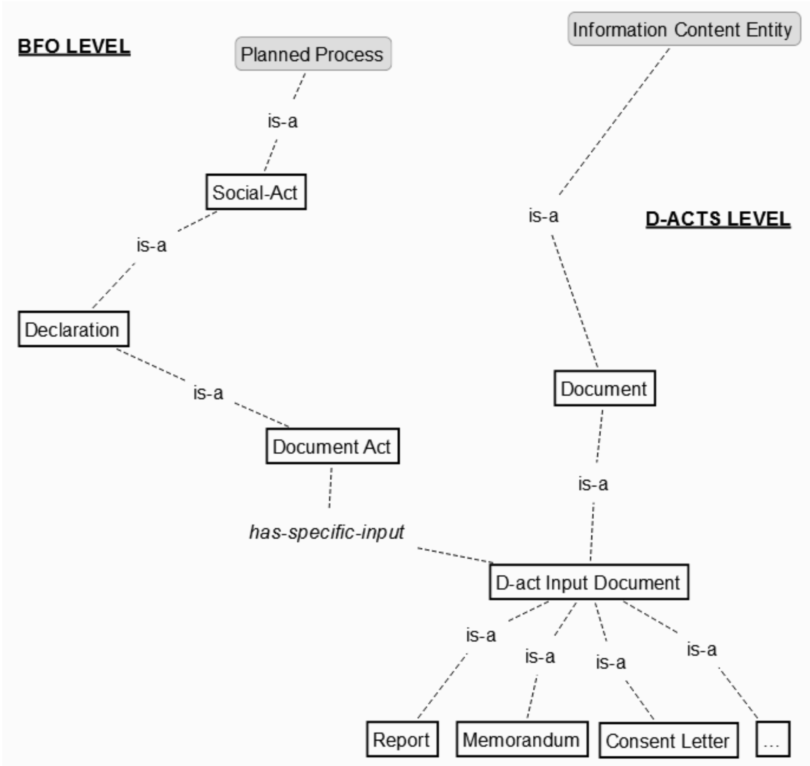
Source: (ALMEIDA; RIBEIRO; BARCELOS, 2019)

Figure 7 presents an essential branch of D-Acts Ontology that concentrates on the roles required to trigger document acts. Figure 8 presents another branch of D-Acts Ontology that gathers documents and the acts they contain. These two branches are linked by the class “Declaration” as seen in both Figures 7 and 8. Classes depicted by a shaded rectangle are not D-Acts Ontology classes, but BFO or IAO classes that illustrate the connection between middle and top-level ontologies.

D-Acts Ontology assumes that claims and obligations are subtypes of BFO classes. An example can illustrate the operation of D-Acts Ontology, as follows. Consider that John claims a piece of land that was unclaimed before. This act creates John's claim to a specific piece of land. This claim is concretized as John's role as claimant of the land. John's claimant role can be performed in multiple ways, for example, in the process of leasing the land to a third party, in the process of selling the land, and so on. Once John sells the land to Mary, his claimant role goes out of existence, and now the claim is concretized in a new role, Mary's claimant role.

Within this complex network of roles, acts, and realizations, one can explicate the basic operations of institutions. Such operations are structured in processes performed by people acting in institutions endowed with ethical and legal powers, namely, deontic powers. Using documents, people can trigger new processes while registering them for future assessments of responsibilities and performance.

Figure 8 - Document and Document Acts in D-Acts Ontology



Source: (ALMEIDA; RIBEIRO; BARCELOS, 2019)

To follow the BFO framework and the kinds of concepts involved, some explanations of D-Acts and IAO models are necessary:

- An "information content entity" is an "object" that is about "some thing." Its dependence is broad, which means that it is encoded in some media, but it can migrate from a bearer to another: for example, as the content of a book can migrate from a paper book to a digital file.
- One level below this, a "directive information entity" is an "information content entity" whose relations and properties indicate to their bearer how to realize them as a process. For example, a consent letter indicates that a nurse has the right to perform processes of extracting blood.
- One more level below, a "document" is an "information content entity," characterized as a collection of information content entities to be understood together as a whole.

On the one hand, IAO's "document" is a document that only records information; on the other, IAO's "directive information entity" also encompasses devices that realize processes. Thus, a "directive information entity" is a kind of document that realizes acts, which is precisely the sense of documents within the D-Acts ontology. Under IAO's "directive information entity," we found a class "plan specification," which we took to be a parent for budgets in the taxonomy not yet constructed. In the context of

D-Acts, a plan specification works as an input for a document act, which, in this context, represents all instances of Brazilian law involved in budget enforcement and approval.

Finally, we would emphasize that the BFO framework provides formal definitions required for computational inferences. We do not expect frameworks to do justice to the richness of the word “document” in the context of LIS (PINTO; ALMEIDA, 2020b).

5. Building the Integration Ontology

In this section, we will join the conceptual modeling theoretical framework with the practical tools that come from D-Acts. We intend to build a new model to demonstrate how we can use D-Acts to understanding a government process like the budgeting process.

We demonstrate in Section 4.1 that the differences between private and public corporations have their origins mainly in two aspects, the subject to whom their acts are directed and their deontology. Private corporations direct their acts, performed through their documents, to their stakeholders, like shareholders, employees, and customers. On the other hand, all the people under a government perceive the document acts from this government.

Furthermore, the deontic powers that establish a government or a business company are very different. One comes from the law, and the other comes from a corporation statute. One reflects the collective intentionality from a group of citizens, and the other the founder’s individual will. We hypothesize that regardless of these differences, we can model a similar process in both corporations in the same way. Our example is precisely the case of the budget process, be it public or private.

The rest of this section is organized as follows: Section 5.1 discusses the complete model for the Budget Process we have built using the top-level ontologies; Section 5.2 presents the new classes of the D-Acts ontology we have created for this new model; Section 5.3 explains the methodology used in the practical experience.

5.1. The complete model

The following explanation takes as reference Figure 9 and IAO Ontology on its latest release 2019-06-07[25]. For each class or relation mentioned, we cite the formal

[25] Available on github.com/information-artifact-ontology/IAO/. Accessed on June 23, 2020.

definition and the IRI[26]. The IRI range reserved for D-Acts entities on IAO is from IAO_0021000 to IAO_0021999. We use the IRI's abbreviated form, omitting the first part, which is, for all OBO ontologies, http://purl.obolibrary.org/obo/. We call this section "the complete model" because it describes how we model our knowledge of the budget process to fit into the D-Acts ontology. In this section, we present our new classes in ontological relationships with the existing top level ontologies: D-Acts, IAO, OBI, RO, and BFO.

We start our model by examining the class 'planned process' – "A processual entity that realizes a plan which is the concretization of a plan specification," OBI_0000011. This class has a D-Acts subclass 'social act' – "A planned process that is carried out by a conscious being or an organization, and is self-generated and directed towards another conscious being or an aggregate of conscious beings, an organization or an aggregate of organizations, and all that must be perceived," IAO_2100003.

As a subclass of 'social act,' we have 'deontic declaration' – "A social act that creates or revokes a deontic role," IAO_2100009. Here we find our starting point, the class 'deontic document act' – "A deontic declaration creating or revoking a deontic role by lawfully manipulating (signing, stamping, publishing) a document," IAO_2100001. This hierarchy demonstrates the dynamic nature of a social act. It prompts us to propose a new class named 'publicBudgeting,' defined as "The process to enact the budget document according to budget law guidelines."

At another branch of the proposed model, we find the class 'organization', as a subclass of 'material entity' – "An organization is a continuant entity which can play roles, has members, and has a set of organization rules. Members of organizations are either organizations themselves or individual people. Members can bear specific organization member roles that are determined in the organization's rules. The organization rules also determine how decisions are made on behalf of the organization by the organization members," OBI_0000245.

Thereby, we state that "organization" has a new subclass named "budgetOffice," and we define this new class as "The organization that plays the role to realize the budgeting actions." OBI ontology has a precise definition of the word "realize" in this context: "Relation between a process and a function, where the unfolding of the process requires the execution of the function. Class level: P realizes F iff[27]: given any p

[26] Internationalized Resource Identifier (IRI) is an Internet protocol standard, which builds on the Uniform Resource Identifier (URI) protocol by greatly expanding the set of permitted characters.

[27] Iff stands for "if and only if" as a logical operator between statements, where either both statements are true or both are false.

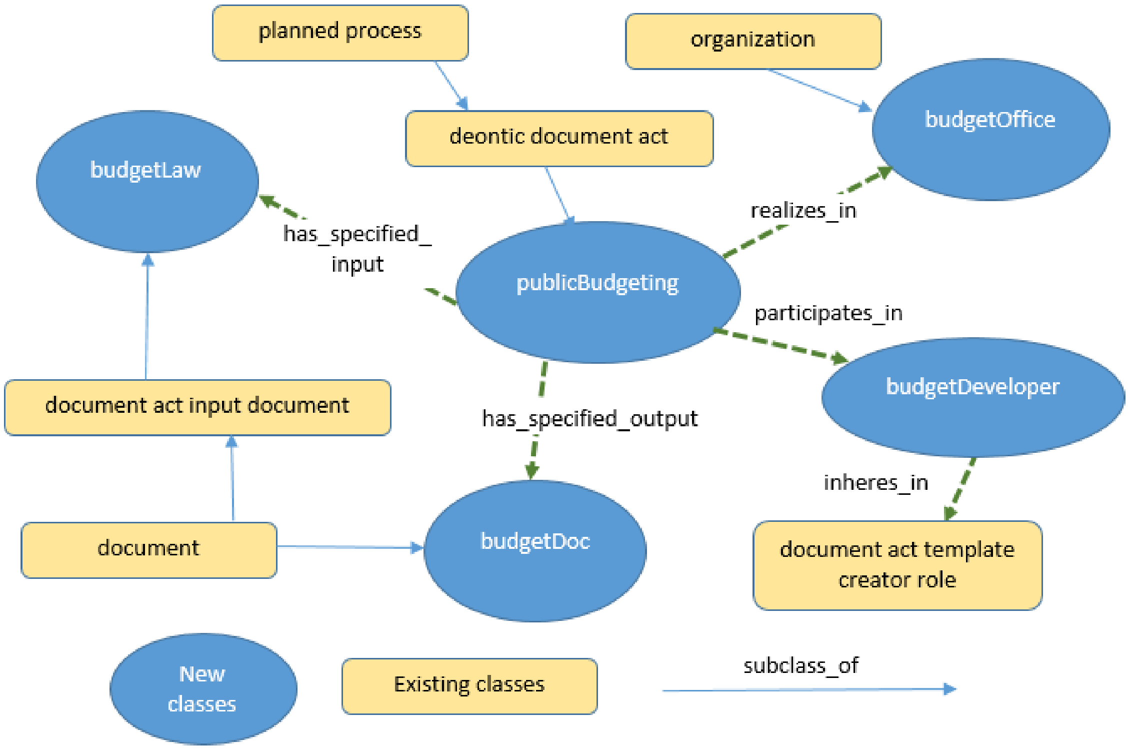
that instantiates P, there exists some f, t such that f instantiates F at t and p *realizes* f. Here, *realizes* is the primitive instance-level relation," OBI_0000308.

Going on through the model, we can observe that we need to define some people, some social (e.g., human) work to develop the budgeting process. These people are the "budgetDeveloper," a new subclass of "document act template creator role." This class is a subclass of "role," a "specifically dependent continuant" defined at the upper-level. Both the classes, "role" and "specifically dependent continuant," are BFO classes, BFO_0000023 and BFO_0000020 respectively, and their explanation and definition are beyond the scope of this thesis, but one can see (ARP; SMITH; SPEAR, 2015) or the website Basic Formal Ontology 2.0 Specification And User's Guide[28].

D-Acts define the class "document act template creator role" as "A role that inheres in a human being or organization or aggregate of any of the aforementioned that prepares a document that is the specified input to a document act and is the input document of a document act," IAO_2100021. Moreover, the relation "inheres in" is "A relation between a specifically dependent continuant (the dependent) and an independent continuant (the bearer), in which the dependent specifically depends on the bearer for its existence," RO_0000052. In other words, budgetDeveloper means the people or aggregate of people that prepare the budget document that is the input to the "public budgeting." These people have the relation "participates in" with the "publicBudgeting" class. Relation Ontology (RO) defines "participates in" as "A relation between a continuant and a process, in which the continuant is somehow involved in the process," RO_0000056. In other words, budgetDeveloper is somehow involved in the process of publicBudgeting.

[28] Available on ncorwiki.buffalo.edu/index.php/Basic_Formal_Ontology_2.0. Accessed on June 23, 2020.

Figure 9 - Public Budgeting model using D-Acts



Source: the author.

Analyzing our central new class, "publicBudgeting," we see that the "deontic document act is a deontic declaration and (has_specified_input some document act input document) and (has_specified_output some document)." These rules come from IAO Ontology at IRI IAO_0021001. Therefore, we need an input and an output to our process.

As the input, we have a D-Acts class called "document act input document" defined as "A document that is intended to be the specified input in a document act. It has a plan specification as a part that specifies the intended socio-legal entities that are created through the document (objective specification) and how the document act is to be performed (by signing, by stamping, and so on) (action specification)", IAO_0021007. Therefore, we propose the creation of a new class called "budgetLaw" to be that specified input. This class may have one instance called Annual Budget Law Of The Federal Government Of Brazil[29]. The Federal Budget is a plan that indicates how much and where to spend federal public funds for one year, based on the total amount collected by the government from taxes. The Executive Branch is the author of the proposal, and the Legislative Branch needs to turn it into law.

Another document we need to be the output of the process is the document now called budgetDoc. Both the input and the output are subclasses of "document," which is defined as "A collection of information content entities intended to be understood

[29] See the real website at tinyurl.com/yxa52wp6. Accessed on June 23, 2020.

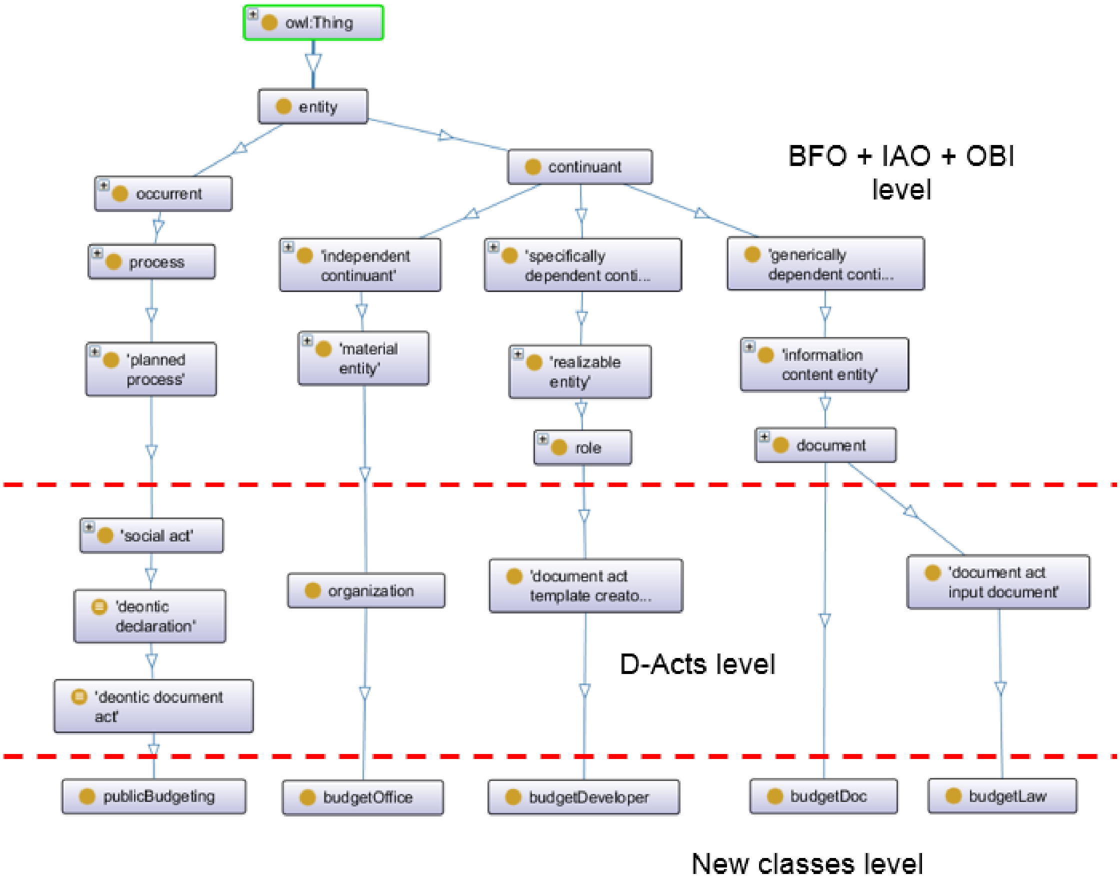
together as a whole," IAO_0020020. Then, we can rewrite our central statement as follows: " 'publicBudgeting' is a 'deontic document act' that 'has_specified_input' a 'document act input document' called 'budgetLaw' and 'has_specified_output' a 'document' called 'budgetDoc.'"

5.2. Contributions to D-Acts Ontology

Figure 10 shows the new class hierarchy, which was built using Protègè ontology editor. We divide the hierarchy into three levels, top to bottom, namely (BFO+IAO+OBI) level, D-Acts level, and new class level. The upper level contains the primary classes, inherited from our foundational ontology, the BFO. Some of those classes come from BFO itself, and others come from IAO and OBI ontologies. All of them contribute as building blocks to D-Acts ontology, the middle level, from where we built our new classes at the bottom level.

Section 5.1 and Figure 10 presents these new classes and the relationships between them. We propose that a model like this one has the potential to represent other government processes. In other words, it can be a guide or a demonstration on how to model a business process using D-Acts ontology.

Figure 10 - New D-Acts classes



Source: the author.

5.3. The methodology of the experiment

Based on the considerations of scientific research methodology, this research can be classified as follows (MARCONI; LAKATOS, 2003):

- Regarding nature, as applied research, because it aims to indicate design guidelines and generate results of practical application for organizations. Specifically, the application built is suitable in the context of the Brazilian Federal Budget and, by extension, in another similar data context;
- As to the approach to the problem, as qualitative research, since the knowledge learned could not be quantitatively measured;
- As for the objectives, as exploratory research, since it studied a subject still incipient in both academic and technical literature, facilitating its understanding and application. This study seeks to improve the discussion of Information Science as a scientific discipline able to help other fields about solutions related to representation, organization, and retrieval of information, especially to develop socio-technical solutions.;
- It is explanatory research since it sought to propose new ways of modeling the reality into the socio-technical systems;
- The technical procedures also include a case study, using real data and producing pieces of evidence of the correctness and the value of the proposed approach;

To achieve scientific research objectives, a researcher must use a set of intellectual and technical procedures from scientific methods to prove his hypothesis. A scientific research methodology is the description of all the systematic and rational, planned activities the researcher follows to answer his research question. One that read the description text may reproduce the way the researcher followed, may see how he detected errors, and how he corrected the way (COLLIS; HUSSEY, 2013).

In this study, the principles described above may not be the most suitable to follow because of one singular characteristic. It is a study to present a theoretical approach, a corresponding model, and an example of its application. We are not focusing on building a new Information Science theory, but rather on applying a theory – the Social Ontology – to solve real-world problems. Thus, we have here a mixed-methods approach, using a combination of both qualitative and quantitative methods. The technical procedures are, at the same time, bibliographic and practical. We agree with the vision that mixed-method research can enrich and add value to the results achieved (MACKENZIE; KNIPE, 2006).

5.3.1. **Cross-industry standard process**

We use in our research an adaptation of the Cross-industry Standard Process for Data Mining (CRISP-DM), which is an open standard process model that describes conventional approaches used by data mining experts. It is one of the most widely-used analytics models (KURGAN; MUSILEK, 2006). We explain the reasons we chose this model in the following paragraphs.

In reviewing research texts for this work, we were surprised to discover the difficulty of defining methodology and methods. The next paragraph shreds some light about the distinction.

> The most common definitions suggest that methodology is the overall approach to research linked to the paradigm or theoretical framework, while the method refers to systematic modes, procedures or tools used for collection and analysis of data (MACKENZIE; KNIPE, 2006, p. 6).

Consider this definition, and we can see that we are looking for a method, as a systematic process, to apply in our research. We found a generic process, structured as an adaptable model that we can use. The original purpose of this process is to use it for data mining jobs. However, there are several examples in the academic and industrial literature of CRISP-DM implementations adapted in other domains. There is an extension of CRISP-DM to stream analytics (KALGOTRA; SHARDA, 2016) and a specialization for evidence-mining (VENTER; DE WAAL; WILLERS, 2007). We found implementations with various grades of modification and adaptations in social network analytics (ASAMOAH; SHARDA, 2015), ERP systems (SASTRY; BABU, 2013), and in some engineering and management areas (HUBER et al., 2019; SCHAFER et al., 2018; WOWCZKO, 2015; YUN; WEIHUA; YANG, 2014).

The original work to guide us on how to understand the scope of CRISP-DM is an academic article (WIRTH, 2000). Additionally, a practical guide is freely available on the Internet[30].

> We can conclude that CRISP-DM works. The generic process model is useful for planning, documentation, and communication. It is fairly easy to write specialized process models based on generic checklists. Finding the right level of detail is still difficult. But the process is living, and therefore all the documents must be living documents, too (WIRTH, 2000, p.10).

Moreover, what is CRISP-DM?, and how can we use it in this work? The very first idea of CRISP-DM was conceived in late 1996 and consolidated in a public document in 2000 by CRISP-DM Consortium, funded by the European Commission. Today it is a robust, open, and well-proven methodology. No one claims any ownership

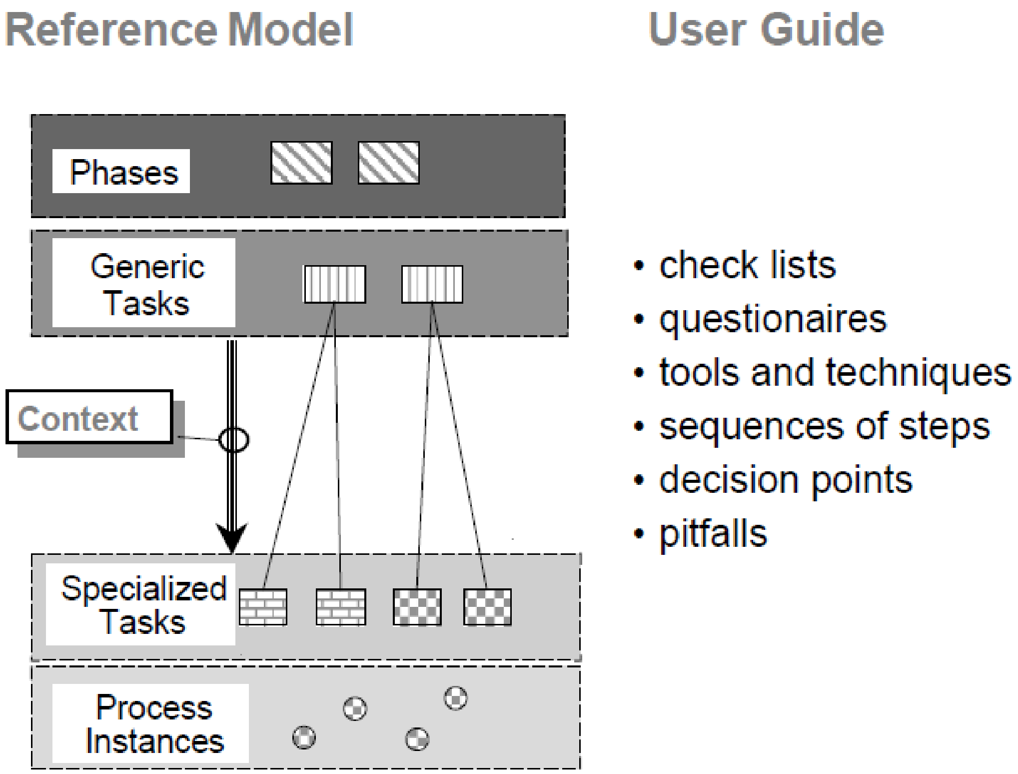
---

[30] We use the copy available at tinyurl.com/75ytjh4. Accessed on May 6, 2020.

over it. In its essence, this work is not a data-mining work, but instead it is a work on ontologies. However, both works have some common properties, of which the most relevant is the use of a database of unknown nature. We follow nearly the way of a data-miner: start with a quantity of data from which we must understand its business meaning, origin, production process, and storage formats. Both the data-miner and the ontologist are searching for something hidden in the data.

The CRISP-DM model is an idealized sequence of events, where the user has the freedom to perform many of the tasks in a different order. It will often be necessary to backtrack some previous tasks and repeat specific steps. Because of its iterative nature, the model does not try to capture all the possible routes through the way to solve one specific problem.

Figure 11 - Four levels breakdown of CRISP-DM methodology



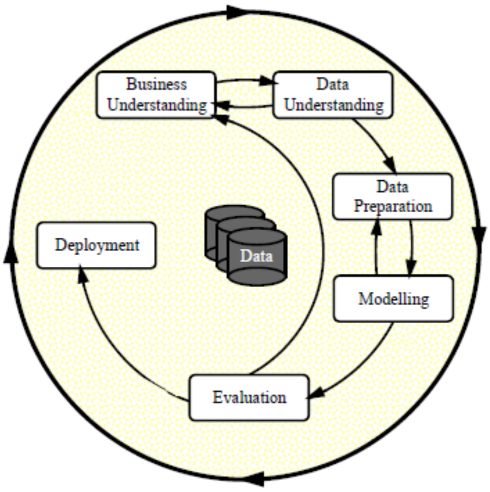
Source: (WIRTH, 2000)

The CRISP-DM methodology is described in terms of a hierarchical process with four levels of abstraction. From the higher level to the lower or, in other words, from general to specific, we have phases, generic tasks, specialized tasks, and process instances (see Figure 11).

In addition to the hierarchical organization, the model provides an overview of the life cycle of the project. At the first level, we broke down the reference model in six iterative phases (Figure 12). In the following, we briefly outline each phase.

- Business understanding: the focus at the first phase is on understanding the project objectives and requirements from a domain perspective. Here we will explain the hypothesis under test.
- Data understanding: the first two phases are closely related. In this phase, the analyst must understand the data and the reality it represents.
- Data preparation: this phase covers all activities to prepare, from the raw data, the final dataset that will be the input of the modeling tool. Tasks in this phase include record, and attribute selection, data cleaning, and transformation of data for modeling tools (in this case, the ontology editor is the modeling tool). Sometimes, in the professional environment, analysts refer to this phase as ETL – Extract, Transform, and Load.
- Modeling: is at this stage that our methodological model has the essential adaptation to the original. Modeling, in CRISP-DM, is related to choose statistical tools for data mining. Here is the phase when we build two complementary models: a class model from an existing ontological framework and an OBDA model suitable for our mapping and data retrieval.
- Evaluation: at this stage of the project, we have built and executed our model, in the developer/test environment. It executes without errors and appears to be adequate to prove our hypothesis. The results must be coherent with what we wanted to reach.
- Deployment: the creation of a model is not the end of the project; it must be converted into a useful knowledge organization and presented to the intended users. In our case, we use the complete project and its results to prove the feasibility of the ontological approach to integrate relational databases. This phase is accomplished through the publication of both the experience demonstrated here, and a tutorial on the use of the technique. We see this publication as our delivery to the research community and professionals. Therefore, this phase is outside the scope of this thesis.

Figure 12 - Phases of the CRISP-DM process model



Source: (WIRTH, 2000)

5.3.2. **Context of the experiment**

To better explore the characteristics of documents in corporations, we chose to consider a very well-known one: a budget. One can make budgets for a planned event or a new venture. Other people can make budgets for an engineered product or even for a simple vacation trip. Also, there are several types of budgets: personal, private, and governmental, for example. Federal budgets are public and common documents that can be seen as a tool for planning since it contains data on all predicted and realized values, both revenues, and expenditures within a given period.

The Brazilian government consists of hundreds of public corporations, which are required by law to publish their projected expenditures in a document called the “Brazilian Federal Budget” (BFB). According to the Brazilian law, the BFB should be publicly available through an Open Linked Data format within a Semantic Web structure for transparency purposes. Thus, this public document made available is not the paper document but, rather, a semantic model called the “Brazilian Federal Budget Ontology” (BFBO), which was created to facilitate the access and data extraction by society as a whole (ARAÚJO; SANTOS; SILVA, 2015).

The website “Ontological Model for Expenditure Classification of the Brazilian Federal Budget”[31] explains the BFBO ontology. It provides additional documentation (e.g., associated laws, user guides, reference manuals, and many others) as well as all BFBO datasets since 2000. We collect all our raw data from this

[31] See vocab.e.gov.br/2013/09/loa. Accessed on May 6, 2020.

website. According to the ontological artifacts presented above, we have tried to combine generic classes of top-level BFO models to BFBO classes, as explained in the next sections.

Part II – THE EMPIRICAL RESEARCH

In the previous part, we have discussed our Theoretical Structure, highlighting some fundamental matters on this broad theme, Knowledge Representation. In Section 2, we talk about Conceptual Modeling as the foundational theory to link KR to Information Systems.

The SLR presented in Section 3 approaches several aspects of our focus, namely, Ontologies about Electronic Government. We covered distinct knowledge areas, from Computer Science to public management, and concluded that there is a vital space for contributions in this field of study. Moreover, our primary goal is to add our results to the joint effort of researchers around the world.

After that, in section 4, we present a fundamental cornerstone of our study, the Social Ontology, as the theoretical basis of the socio-technical vision we want to develop for the new e-Gov systems.

Dealing with this challenge, in Section 4, yet, we approach the pervasive problem of Semantic Interoperability through Ontologies using the OBDA technique. To demonstrate the real use of the modeling techniques and the potential of OBDA to integrate legacy databases, we develop a solution using government data from the Brazilian Open Data Portal. We will describe and report this case study in the next sections.

Section 5 is dedicated to the scientific research methodology for this study, including the methodological steps taken to achieve the research goal and to answer the research question.

In Part II, we present how we carried out the research, in both conceptual and practical terms. On the conceptual side, we deal with models, while on the practical side, we describe the application of the ideas and techniques that were studied.

This part is organized into two sections. The first, Section 7, describes the preparatory phases of the experiment and the other, Section 8, reports on the execution of each phase and shows evidence of the results achieved. There are five subsections in each section, and they have linked accordingly, in order of execution.

6. Preparatory phases

This section presents and discusses the preparatory procedures for conducting the whole experiment. We explain each task and justify the choices we have made. At the same time, we intend to provide evidence and other conditions so that other experimenters can reproduce the entire experiment.

6.1. Domain Understanding

Here we have one of the specializations we did to the original model: we use the word “domain” instead of “business.” “Domain” is more concise to use in an ontological study like this and have no loss of meaning in this context. We adopted the definition from Prieto-Díaz, from the field of software engineering: “In the context of software engineering it [domain analysis] is most often understood as an application area, a field for which software systems are developed.” (PRIETO-DIAZ, 1990, p. 50).

Our hypothesis here is that we can use public ontologies to design an OBDA model that integrates heterogeneous data using an RDB to RDF Mapping Language (R2RML) mapping structure. R2RML is a W3C standard language for expressing customized mappings from relational databases to RDF datasets[32]. At this phase, we have two desired outputs from generic tasks (see Figure 12): set the objectives and produce a project plan.

The objectives of the experiment are:

1. Modify IAO Ontology adding some new concepts and relationships at d-Acts, which is part of IAO;

2. Demonstrate the use of the ontology d-Acts-budget (modified d-Acts with some new classes as budgets concepts) to represent data at Brazilian Federal Budget (BFB);

3. Demonstrate the use of d-Acts-budget to retrieve real data from BFB stored as Relational Database (RDB) using SPARQL;

4. Demonstrate the possibility of using d-Acts-budget to integrate BFB to a hypothetical RDB.

At this phase, the second generic task, produce a project plan, has three specialized tasks, namely: initial assessment, determine success criteria and build the complete project plan.

In the initial assessment, we made the following inventory of available resources for the experiment in terms of hardware and operational software:

- Notebook Dell, with i5 processor, 8GB RAM, running Windows 8.1;
- Protégé version 5 – Ontology Editor[33];
- Ontop version 3 – A Virtual Knowledge Graph System[34];
- H2 version 1.4 – In-memory SQL database[35].

---

[32] Available on www.w3.org/TR/r2rml/. Accessed on July 31, 2020.
[33] Available on protege.stanford.edu. Accessed on July 31, 2020.
[34] Available on ontop-vkg.org. Accessed on July 31, 2020.
[35] Available on www.h2database.com. Accessed on July 31, 2020.

All standards adopted here come from W3C in its latest version, of which all related documents are freely available at Semantic Web Standards homepage[36]. These standards are RDF, SPARQL, Ontology Web Language (OWL), and R2RML.

Another standard language we use in this experiment is Structured Query Language (SQL), also in its latest version, which displays specification documents freely available at the Database SQL Language Reference homepage[37].

One of our most valuable project preconditions is to promote the reuse of public ontologies. Thus, we use the entire ontological framework available at BFO Foundry, as explained in Section 4. Explicitly, they are the main reused ontologies, all in their latest version, namely, BFO, IAO, D-Acts.

To complete all the statements to the initial assessment, we explicitly made the following assumptions:

- We adopted open standards in all the procedures and codes.
- The data is available in raw format at Brazil Open Data Portal (*Portal Brasileiro de Dados Abertos*)[38].

To assert the “criteria of success” for this experiment, we state the following: with appropriate mappings relations, the experiment should prove that a SPARQL query over two different relational databases returns the same data from two SQL queries over different databases.

The third specialized task is to build a project plan for the experiment. We briefly list the steps to conduct the case study. According to the model (see Figure 12), we classify these steps as specialized tasks. The tasks are listed below, without much descriptive accuracy or chronological order.

1. Creation of new classes at d-Acts, generating new ontology d-Acts-budget.
2. Creation of a simplified relational model to BFB.
3. BFB sampling from Brazil Open Data Portal, in at least two annual historical series.
4. Creation of a routine for reformatting BFB data from CSV format to the RDB database.
5. Loading of BFB samples in H2.
6. Mapping from BFB to OWL using R2RML.
7. Loading of R2RML code in Ontop.
8. Verifying the consistency and correctness using Ontop reasoner.

[36] Available on www.w3.org/standards/semanticweb/. Accessed on July 31, 2020.
[37] Available on tinyurl.com/y2rmc4a3. Accessed on August 4, 2020.
[38] Available on dados.gov.br/. Accessed on August 4, 2020.

9. Connecting H2 and Protege using JDBC[39] driver.

10. Executing some simple SPARQL queries on Protege to retrieve data instances from the RDB database.

11. Comparing the queries result with original data.

12. Design and load of a hypothetical RDB with some new attribute to BFB, with the same identifiers (primary key).

13. Including this new attribute as a new relationship in d-Acts-budget.

14. Executing some simple SPARQL joining results from both RDBs.

15. Compare the results from the SPARQL query with the original data and with the results returned by the correspondent SQL queries.

6.2. Data Understanding

The second phase of CRISP-DM requires the acquisition of experiment data, as listed in project resources. At this phase, we have four generic tasks, each one with a correspondent report as output: data collection, data description, data exploration, and data quality.

All Brazilian open government data is available at a central website, with a search box to select the specific dataset the user wants. We accessed this portal website and searched for the string "*loa federal*." "LOA," in Portuguese, which stands for "*Lei Orçamentária Anual*" ("Annual Budgetary Law")[40]. The word "federal" has the same spelling and meaning, both in Portuguese and in English. On the LOA website[41], we have a brief explanation of the data: its concept, provenance, associated legal elements, and a list of resources available to explore the data. We chose the resource named "*Consulta Pública ao Orçamento Federal*" ("Public Inquiry into the Federal Budget"). The next page provides the user with several options to explore the Federal Budget[42]. We chose the option "*Consulta Livre*" ("Free Queries").

This query builder is our primary data sampling tool, as the total budget data is enormous, with over 2 million records per year. To select records in the database, we must provide some parameters to the query builder. We used the query twice, once to extract budgeted values and once to extract spent values, in order to build two relational files with the same primary key and different attributes.

---

[39] Java Database Connectivity (JDBC) is an application-programming interface (API) that uses the Client-server model, which defines how a program may access a database.
[40] Translations made by the author.
[41] Available at /www.dados.gov.br/dataset/orcamento-federal. Accessed on June 16, 2020.
[42] See tinyurl.com/ycbjyyum. Accessed on June 19, 2020.

The first step is to select the basic filters needed. The first filter is the reference year, and we selected 2019. We made this choice because it is the most recent year in which we have data on two fields we need, budgeted_value and spent_value. For the next year, 2020, we do not have complete data on the spent value. The second filter used was "*Orgão Orçamentário*" ("Budgetary Office"). We arbitrarily chose "*Ministério da Educação*" ("Ministry of Education"). We left the other filters with their default values.

The second step to extract our sampling was to select which fields we wanted to include in our results. We made this choice in two parts: fields to be included as detail and fields by which we would summarize the results. We left the first part of the query builder panel with three values, "*Detalhar por Orgão Orçamentário*, "*Detalhar por Ação*," and "*Detalhar por Unidade Orçamentária*" ("Detail by Budgetary Office," "Detail by Action," and "Detail by Budgetary Item," respectively). After uploading the raw data extracted to the database, we renamed the fields "Budgetary Office" and "Budgetary Item" as "beneficiary" and "office," respectively.

In the second part of the query builder panel, we have two selector keys, "*Apresentar Dotação Inicial*" and "*Apresentar Valores Pagos*" ("Summarize by Budget Value" and "Summarize by Spent Value", respectively) The extraction is run twice, the first with the key labeled "Summarize by Budget Value" switched on, and the second with the key labeled "Summarize by Spent Value," switched on. All other selector keys at this part of the query builder's panel remain switched off.

The third step in sampling is to visualize the results in a report format. We need to convert the report data into a file we can work on. To do this, we click the mouse right button over any part of the report. In the context menu, we chose "*Enviar para Excel*" ("Send to Excel") and saved the generated file. The first run generated the file *budgetedValues.xlsx*, and the second run generated the file *spentValues.xlsx.* With this action, we finished our first generic task in this phase, the data collection phase.

The next generic task is a description of the data, in which the main objective is to evaluate whether the acquired data meets our requirements. As both files are formatted as a report, we had to remove all print formats to use the data in its raw format. As a report, it has footer and header lines, which were also removed. The first column has the same value, 2019, for all lines, and then it may be removed. The second column is the Budgetary Office, and we leave it unchanged. The value column must be an integer, and then we must round it to the next integer, to be sure we do not have decimal points. All these "cleaning tasks" can be performed using a simple spreadsheet editor.

The last operation to be done in this task is to export the files to a CSV format[43], generating the new *budgetedValues.csv* and *spentValues.csv* files.

Now, to accomplish the next generic task, we explore the data using a simple data visualization tool and check some aspects of both of the datasets we have. We examine the composition of the primary key (PK), as a concatenation of two fields, “office” and “action,” with special attention to the correspondence between the PK in each of the files and the magnitude order of numeric fields.

To evaluate the quality of the data, we dealt with aspects such as completeness, accuracy, and physical integrity. Once all the data we use comes from an official source, it is supposed to be correct, no matter how complete it is because our experiment is about data formats and retrieval, and our attribute choices are arbitrary. We take specific care of the physical format. We have special Portuguese characters on every string. Therefore, we carefully code all data using UTF-8[44].

6.3. Data preparation

In the third methodology phase, we prepare our data according to the planned experiment. We convert our CSV files to relational tables into an H2 database[45] (see Figure 13).

Figure 14 schematically explains how we transform the raw data, two CSV files, and then into two databases using H2 DBMS[46] and SQL commands. Each of these databases has one relational table. Both tables have the same basic format, and the only difference is the “value” field, whereas one has the budgeted values, and the other has the spent value.

As demonstrated in Section 5, we have an ontological model as a correspondence with the relational model. We call the ontological model as “OBDA class model” (Figure 14) because in here we include the new classes and relationships we created to map one model to another.

---

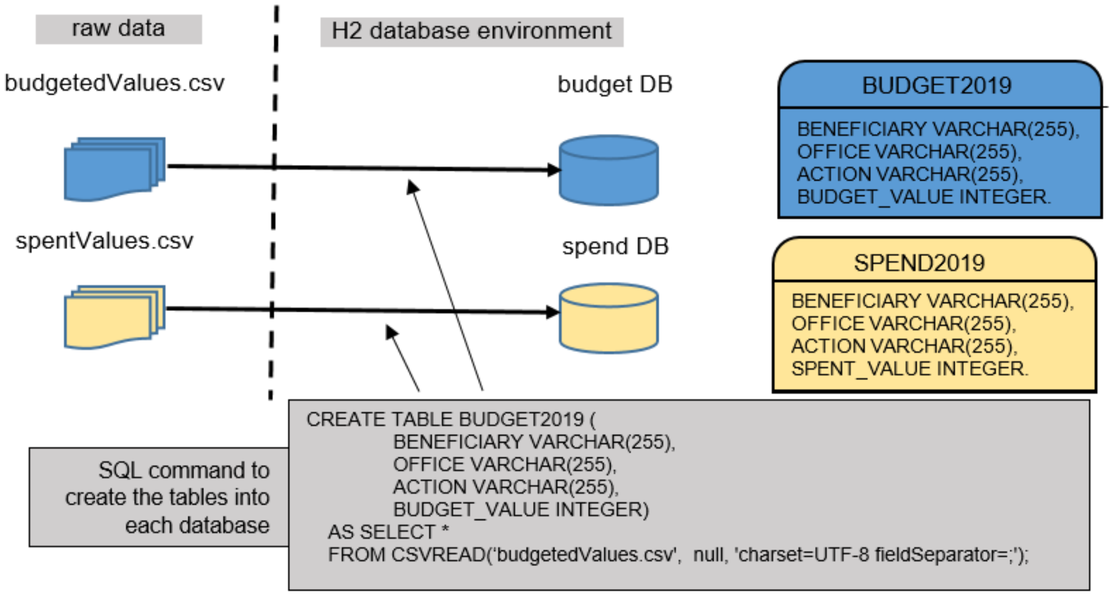
[43] A Comma Separated Values (CSV) file is a plain text file that contains a list of data. These files are often used for exchanging data between different applications.

[44] 8-bit Unicode Transformation Format (UTF-8) is a variable width character encoding capable of encoding all 1,112,064 valid character code points in Unicode using one to four one-byte (8-bit) code units.

[45] Available at www.h2database.com. Accessed on May 6, 2020.

[46] The database management system (DBMS) is the software that interacts with end users, applications, and the database itself to capture and analyze the data.

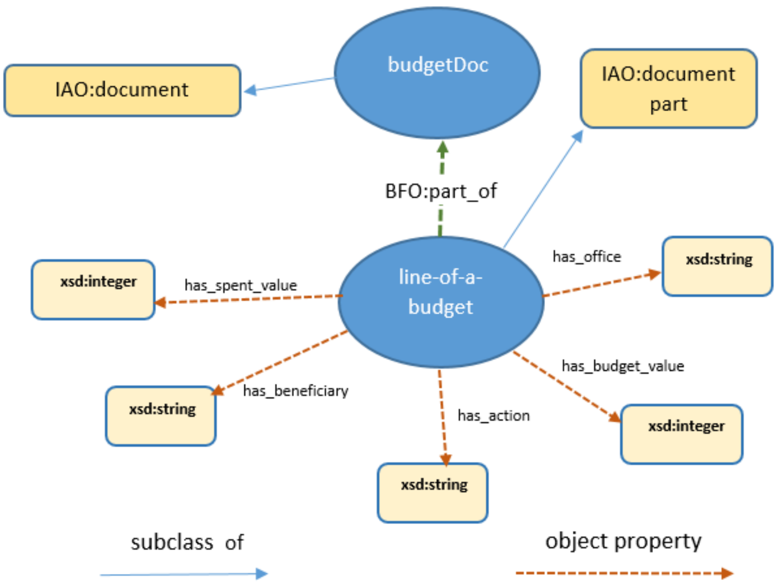
Figure 13 - Data preparation schema



Source: the author.

Additional information required for this Methodological Phase is related to the data that we are going to use for the experiment. There are no additional records, and we do not create any new attributes. We describe all the cleaning tasks in the previous section. We did not make any modifications to the original raw data, which remains unchanged, as extracted using the free tool.

Figure 14 – OBDA class model



Source: the author.

6.4. Modeling

In this Methodological Phase, we join the two models, the relational model with the OBDA model, using R2RML. There are some implementations and techniques to use R2RML, with small differences (BROWNE et al., 2019; SLEPICKA et al., 2015; TAHERIYAN et al., 2016; ZHOU et al., 2013). We base our experiment on the studies and tools made available by Diego Calvanese, as already cited (CALVANESE et al., 2017). The tools are publicly available at the Ontop gitpage[47].

> Ontop is a Virtual Knowledge Graph system. It exposes the content of arbitrary relational databases as knowledge graphs. These graphs are virtual, which means that data remains in the data sources instead of being moved to another database. Ontop translates SPARQL queries expressed over the knowledge graphs into SQL queries executed by the relational data sources. It relies on R2RML mappings and can take advantage of lightweight ontologies. (https://ontop-vkg.org/guide/. Access at May-12, 2020)

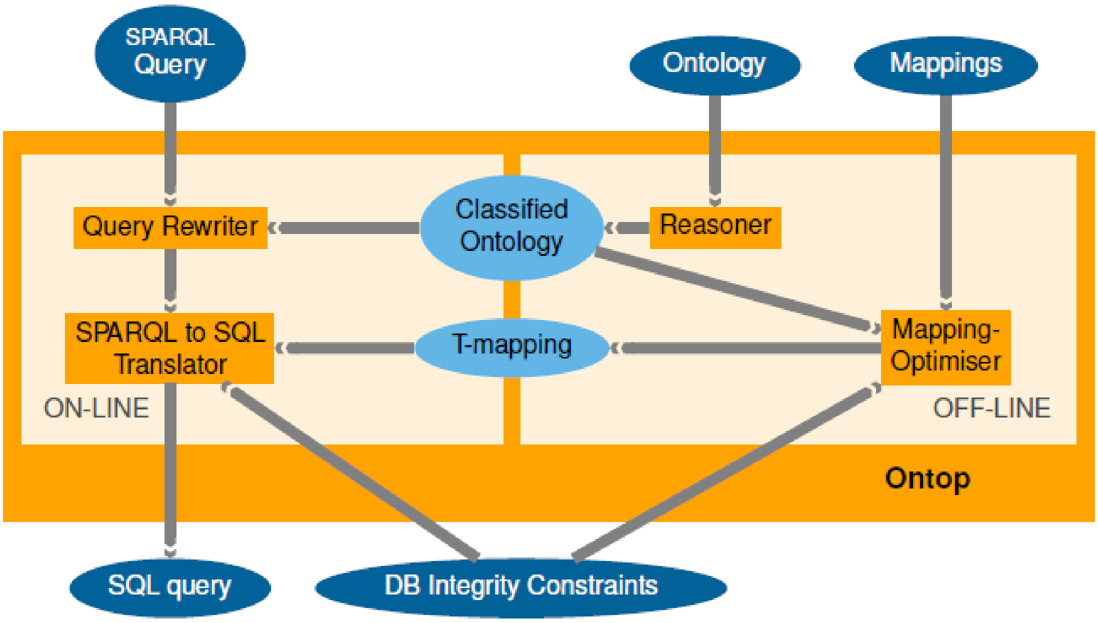
R2RML mappings provide the ability to view existing relational data in the RDF data model. We can express the mappings according to our models in a structure and target vocabulary of our choice. R2RML mappings are themselves RDF graphs and written down in Turtle[48] syntax. In figure 21, we show an excerpt of one mapping, and the complete Turtle file is available in Annex 1.

To process a SPARQL Query over relational databases, we need some structural software components; all of them are publicly available at the Apache Jena project[49]. Apache Jena is a free and open-source framework for building Semantic Web applications.

To develop an OBDA application, we need an Integrated Development Environment (IDE). An IDE is a software application that provides comprehensive facilities to computer programmers for software modeling and development. In this experiment, we use Protégé, which is one of the most used IDE by semantic web developers

[47] Available on github.com/ontop/ontop. Accessed on May 6, 2020.
[48] See github.com/ontop/ontop/wiki/TurtleSyntax. Accessed on June 24, 2020.
[49] Available on jena.apache.org/. Accessed on May 6, 2020.

Figure 15 - ONTOP workflow – example of an OBDA framework



Source: (CALVANESE et al., 2017)

To exemplify the workflow of our experiment, we use Figure 15, which explains the Ontop way of work. The essential characteristic of this architecture that drives our experiment from the beginning is the use of internationally open standards. The three components we create are the databases (SQL), the ontology (OWL), and the mappings (R2RML). We do not write any new code, and we do not use any proprietary tool. We design the experiment to prove the possibility of OBDA using open standards, real data (from Brazilian Open Data Portal), and free tools (H2, Protégé, Ontop).

6.5. Evaluation

At this stage, we evaluate our partial findings so far. Our previous steps dealt with factors such as the accuracy and generality of the model. We now assess the degree to which the model meets our business objectives and seek to determine if there is any business reason for this model to be deficient.

The CRISP-DM methodology establishes that it is time to evaluate the results, approve the models, review the entire process, and determine the next steps. We undertake these generic tasks comparing the results with the objectives listed in Section 7.1 Domain Understanding. All the pieces of evidence to this evaluation are presented in Section 8.

- **Objective 1**: Modify IAO Ontology, adding some new concepts and relationships at d-Acts, which is part of IAO. We create two new classes in the IAO Ontology, budgetDoc as a subclass of IAO:document, and line-of-a-budget as a subclass of

IAO:document_part. We link these two new classes with the BFO relationship part_of. In order to reflect the original BFB model, we create five object properties to represent the attributes of each budget line.

- **Objective 2**: Demonstrate the use of the ontology d-Acts-budget (modified d-Acts with some new classes as budget concepts) to represent data at BFB – Brazilian Federal Budget. When we query the OBDA model using SPARQL and query relational DB using SQL, we find the same results (see the confirmations in Section 8.5). Thus, we say that the concepts are apparently the same. It is essential to notice that the ontology has no instances. All the data originates from the relational DB, which was uploaded with raw data extracted from BFB using a sampling tool.
- **Objective 3**: Demonstrate the use of d-Acts-budget to retrieve real data from BFB stored as RDB - Relational Database – using SPARQL – Protocol and RDF Query Language. We have done this retrieval using Protégé with Ontop plugin. This plugin includes the reasoner, the mapping editor, and a SPARQL end-point.
- **Objective 4**: Demonstrate the possibility of using d-Acts-budget to integrate BFB to a hypothetical RDB. This objective is a combination of the two previous objectives. Objective 2 relates to the capacity of the representative model, and Objective 3 concerns the operational possibility of the two models, relational and conceptual (OBDA). Thus, the joint achievement of both objectives confirms the feasibility of using the modified ontology as an integrating element of two relational data banks.

7. Execution of the experiment

This Section presents and discusses the results achieved with the experiment described in the previous section. We enlightened how we conduct each step and provide evidences about the outcomes.

For each phase or step performed, we present the results achieved and try to clarify and display each technical component of the phase.

7.1. Results for Domain Understanding

Examining our results, one may ask whether we have a complete understanding of our underlying study object, the BFB. We started our work from an existing domain ontology, designed by government budget experts. Therefore, we assume that we started from actual knowledge by assumption, and we have not changed any of the fundamental concepts in the original ontology.

As shown in the previous section, we planned and realized a complete OBDA experiment. The initial evaluation we made is correct, and the available resources, namely hardware, software, and data, are consistent with the objectives. They have proven to be sufficient to perform all the planned tasks.

7.2. Results for Data Understanding

The physical structure of the BFB is a simple table without PK or external links. This feature simplifies data selection and extraction. The query builder makes it much easier to choose attributes and summarize. Of course, designers have designed this tool to explore the database for data mining applications.

We made two extractions, using the parameters explained in Section 7.2, one run for budgeted amounts, and another for spent amounts. We retrieved 1,789 records from the Budget database and 1,221 records from the Spend database. We cleaned the data removing empty records, and rounding the decimal values.

To build the relational database, we rename the original fields to new English names to make sense in this work (see Figure 14). It is important to note that there seems to be a need to create a PK joining the two new fields (office + action). Although this PK is not essential for the OBDA experiment, we do it only to clarify and simplify the understanding of R2RML mappings.

7.3. Results for Data Preparation

We uploaded the data to H2 DBMS, as shown in Figure 14, and verified the contents of each database using a simple SQL command.

Figure 16 - First 10 lines from Budget database

```
select BENEFICIARY , OFFICE , ACTION , BUDGET_VALUE from BUDGET.BUDGET2019;
```

BENEFICIARY	OFFICE	ACTION	BUDGET_VALUE
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	0Z00 - Reserva de Contingência - Financeira	655882076
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	0Z01 - Reserva de Contingência Fiscal - Primária	2829285905
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	09HB - Contribuição da União, de suas Autarquias e Fundações para o Custeio do Regime de Previdência dos Servidores Públicos Federais	746815904
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R2 - Implantação do Novo Campus do Instituto Nacional de Matemática Pura e Aplicada por Organização Social (Lei 9637/98)	10000000
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior	285270484
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R4 - Apoio à Expansão da Rede Federal de Educação Profissional, Científica e Tecnológica	255250894
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20GK - Fomento às Ações de Graduação, Pós-Graduação, Ensino, Pesquisa e Extensão	18170000
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20RH - Gerenciamento das Políticas de Educação	146000000
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20RJ - Apoio à Capacitação e Formação Inicial e Continuada para a Educação Básica	27000000
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20RU - Gestão Educacional e Articulação com os Sistemas de Ensino	1000000

(10 rows, 24 ms)

Source: the author.

Figure 17 - First 10 lines from Spent database

select SPEND_BENEFICIARY , SPEND_OFFICE , SPEND_ACTION , SPENT_VALUE from SPEND.SPEND2019;

SPEND_BENEFICIARY	SPEND_OFFICE	SPEND_ACTION	SPENT_VALUE
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior	13645027
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R4 - Apoio à Expansão da Rede Federal de Educação Profissional, Científica e Tecnológica	20370982
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20GK - Fomento às Ações de Graduação, Pós-Graduação, Ensino, Pesquisa e Extensão	9037857
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20RH - Gerenciamento das Políticas de Educação	27140365
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20RJ - Apoio à Capacitação e Formação Inicial e Continuada para a Educação Básica	19548
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	20TP - Ativos Civis da União	114906375
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	212B - Benefícios Obrigatórios aos Servidores Civis, Empregados, Militares e seus Dependentes	7104313
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	212H - Manutenção de Contrato de Gestão com Organizações Sociais (Lei nº 9.637, de 15 de maio de 1998)	326870621
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	216H - Ajuda de Custo para Moradia ou Auxílio-Moradia a Agentes Públicos	1425447
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	219U - Apoio ao Funcionamento da Rede Federal de Educação Profissional Científica e Tecnológica	519380

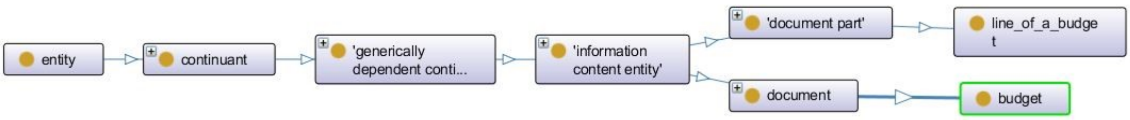
(10 rows, 27 ms)

Source: the author.

Figures 16 and 17 show the first ten lines of each database. The SQL command that generates these reports is described in the top left corner of the figure frame. The data is not sorted, and there is no reference to the PK as the report is only a sequential list. Each line of these tables comes from one record at CSV original files.

Figure 18 shows the hierarchy of classes built at Protégé. The figure shows only "subclass_of" type relations, and the two new classes added, "budget" and "line_of_a_budget." As we can see, each table in the database is a budget, and each line of this table is line_of_a_budget.

Figure 18 - Class hierarchy for mappings



Source: the author.

7.4. Results for Modeling

Since the mid-2000s, OBDA has become a widely accepted approach to cope with the challenge of integrating heterogeneous data sources. As we saw in Section 3, most of the socio-technical systems operate a single all-knowing, self-contained central database. In general, these databases have the characteristics or dimensions that define big data: velocity, variety, and volume. According to the 3Vs model, the challenges of big data management result from the continuous expansion of all three dimensions (KADADI et al., 2014).

The critical point of OBDA modeling is mapping. As Calvanese said: "The process of creating mappings is probably the most complex step in setting up an OBDA system" (CALVANESE et al., 2017, p. 5). In this context, an integration ontology is a specification of a conceptualization shared between two designers. The first person is the systems analyst who built the original relational data model, and the other is the ontologist who mapped the initial concepts to the ontology (ALMEIDA, 2013). This difficulty is related to knowledge differences that may exist among them. In the next paragraphs, we will explain the reasoning that led us to design the class hierarchy exposed in Figure 19.

The designers built the Brazilian Federal Budget Ontology – BFBO (ARAÚJO; SANTOS; SILVA, 2015) on top of a prior classification model that relates each expense to a unique numeric code. This model is similar to a glossary, inspired by a standard for classifying purposes of government activities called Classification of the Functions of Government (COFOG)[50], a document developed by the Organization for Economic Co-operation and Development and published by the United Nations Statistical Division. Brazilian law enforces the gathering of expenditure values in elements called "Expenditure-Items" according to specific criteria. The main goal of the budget is to register how governing bodies spend money while detailing which public unit is responsible for the expenditure.

BFBO reveals a poor structure when one considers the richness of alternatives that could be used to organize this data. The law classifies BFB according to only two axes: the institutional classification and the functional classification. Primarily, BFBO is, indeed, merely a list of items with values attached, each item corresponding to a line in a paper document. It does not provide any organization in the form of generic classes, but starts with the root concept "thing" and uses representation languages typical from Semantic Web initiatives. Also, classes in BFBO are not adequately defined; for example, "year" must be a property and not a class.

To perform our experiment, we made BFBO classes fall under D-Acts and IAO models while maintaining BFO as the top-level. Based on the foundations described in Section 4.3.2, we created a class called "budget_doc" a subclass of "document," as the starting point to better organize BFBO (see Figure 14). With these and other amendments to the original BFBO, the proposed structure reflects a more suitable structure of classification in contrast to the original static and straightforward list. By "suitable," we mean that from our amendments, the classes of a computational artifact received theoretical considerations to be correctly classified – and not defined in an ad-hoc way - according to Information Science theories.

7.5. Results for Evaluation

To evaluate our results, we start selecting two records from the original databases, one from the Budget and another from the Spend. Figures 19 and 20 show each record, respectively, and the correspondent SQL query that retrieves it. The fields

[50] Available at bit.ly/2TLvugs. Accessed on April 15, 2020.

“office” and “action” are equivalent to a composite primary key. We choose these values at random, and the values chosen do not change the experiment at all. It is essential to note that the primary key does not exist in the original databases. Moreover, we do not modify the legacy data in any way.

Figure 19 - Target record at Budget database

```
SELECT * FROM BUDGET.BUDGET2019
 WHERE
  BUDGET_OFFICE = '26101 - Ministério da Educação - Administração Direta' AND
  BUDGET_ACTION = '15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior' ;
```

```
SELECT * FROM BUDGET.BUDGET2019
 WHERE
  BUDGET_OFFICE = '26101 - Ministério da Educação - Administração Direta' AND
  BUDGET_ACTION = '15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior' ;
```

BUDGET_BENEFICIARY	BUDGET_OFFICE	BUDGET_ACTION	BUDGET_VALUE
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior	285270484

(1 row, 16 ms)

Source: the author.

Figure 20 - Target record at Spend database

```
SELECT * FROM SPEND.SPEND2019
 WHERE
  SPEND_OFFICE  = '26101 - Ministério da Educação - Administração Direta' AND
 SPEND_ACTION = '15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior' ;
```

```
SELECT * FROM SPEND.SPEND2019
 WHERE
  SPEND_OFFICE  = '26101 - Ministério da Educação - Administração Direta' AND
 SPEND_ACTION = '15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior' ;
```

SPEND_BENEFICIARY	SPEND_OFFICE	SPEND_ACTION	SPEND_VALUE
26000 - Ministério da Educação	26101 - Ministério da Educação - Administração Direta	15R3 - Apoio à Expansão das Instituições Federais de Ensino Superior	13645027

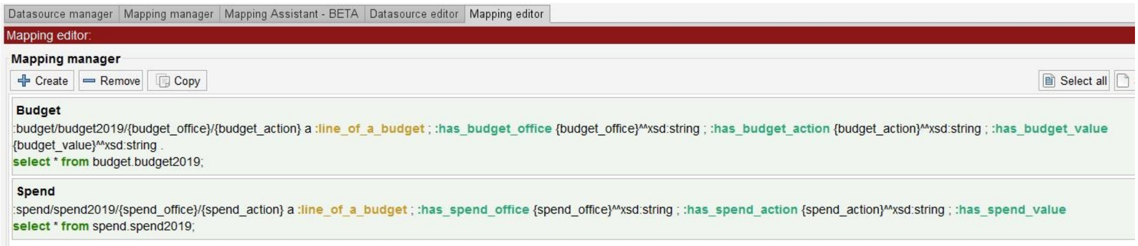
(1 row, 15 ms)

Source: the author.

Figure 21 shows the mapping we did when modeling with the ONTOP plugin at Protègè. It shows two mappings, one for the Budget database and another for the Spend database. Written in a simplified language provided by ONTOP, the mappings will be later translated into standard R2RML language format. The R2RML code produced by this mapping is in Annex 1. Both mappings have the same general format because the mappings must reflect the database structure, thus being very similar. The first line of the mapping is only a title, in bold letters. The second line, called target, is a triple in the standard format subject/predicate/object described using a Turtle-like syntax. The subject is the database name, as referred to in the JDBC plugin, plus the primary keys, even though they do not exist in the original database. Ontop uses this information to optimize query performance, removing self-joins. The predicate is the letter “a,” which is a short form to the predicate “rdf:type.” The object is our new class, line_of_a_budget, with its “objectProperties.” We modeled three ObjectProperty, one for each column we want to retrieve from the database. Each ObjectProperty is associated with the column name and type, as the same defined in the database data model. The third line is the part called “source” in Ontop terminology. It contains the SQL code to retrieve the records we want for this mapping.

Figure 22 shows the SPARQL editor with the query that will be sent to Protégé internal SPARQL end-point and analyzed by the Ontop reasoner. In the lower part of the figure, we see the values returned for the fields, budget_value, and spend_value. They are the same values returned by the SQL query shown in Figures 19 e 20.

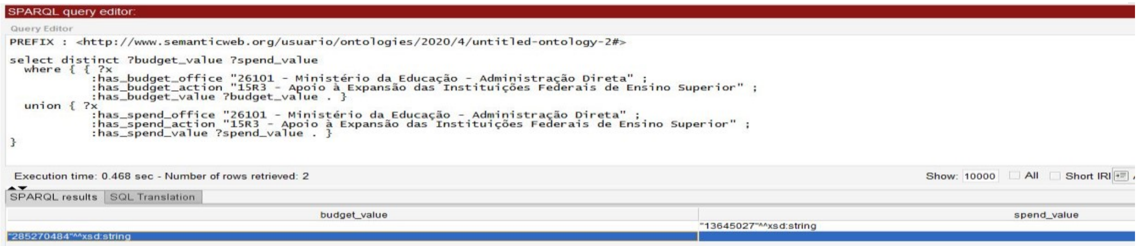
Figure 21 - Mappings in simplified Turtle language



Source: the author.

Therefore, we can say that the SPARQL query returned the same values as the SQL query, providing proof that the entire OBDA scheme works as planned.

Figure 22 - SPARQL query to retrieve the select records



Source: the author.

8. Final Remarks

The research we presented so far aimed to identify the feasibility of integrating government data using both public ontologies and OBDA techniques. In grounding the research on a theoretical foundation called Social Ontology, we conclude that the components needed to develop the aimed solutions are already public. By combining a set of components and standards, the research shows how to develop the conceptual model and build a technical element to perform the integration, as mentioned above. The practical experiment is now complete: it began with data extraction from public databases, followed with modeling and building an intermediate ontology integrated with BFO, and gathered all together using an OBDA tool. We believe that this initiative breaks new ground in dealing with challenges associated with developing new systems and services to the citizen using legacy government databases.

8.1. Research Timeline and Conclusions

The research started with a personal observation, which has been enlarged over decades of work as a systems engineer regarding difficulties in understanding information systems with no proper documentation, but only programs and database models available for reference. Activities of writing and maintaining software application requirements cause problems for almost all teams of software developers. Various research approaches have attempted to answer questions involved in such a situation, different practical and theoretical procedures have been presented, but the whole situation remains without a sound solution. Our point is that the underlying issue in such problems is the same: people have difficulties in representing knowledge about a domain. This point is the same whether the representation is processed for humans (requirements) or by computers (data model). One of the most severe consequences of this issue is the lack of interoperability between systems, databases, and digital services. The interoperability issue is a concern that is on the agenda of all companies and governments.

Our focus here is on social-technical systems for the government, considering that it is an area that presents one of the biggest challenges in dealing with data. In recent years, one can observe companies' progressive migration to more modern system architectures, including large ERP-type solutions and other not so contemporary software engineering resources. These new architectures partly solve, for private companies, some challenges of heterogeneous data integration. However, this movement is not replicated in many cases in government systems, which remain using old and massive isolated data silos. It is crucial to realize the value that these data-sets have for citizens insofar as they are composed mainly by sensitive and private personal information. Government databases store personal data on the health, finances, and personal life of each citizen. Moreover, we must consider that these databases have cost the public coffers a considerable investment of time and money, which cannot be wasted. An additional component of this context is the growing need to provide more and better e-Gov services to the population, allowing democratic access to public services and improving public governance.

How can Information Science address these enormous challenges? The alternatives should consider the development of modern solutions and the preservation of public investments in existing databases. The approach we follow as an alternative in this thesis is based on Knowledge Representation, particularly on Conceptual Modeling supported by ontologies. The research introduces the possibilities and benefits of an ontological approach, not only in building solutions but also in understanding the information to be represented. To build solutions, we present a case study showing how to integrate two relational databases using ontologies. Besides, to provide a better understanding of the framework, we discuss philosophical tenets about government institutions' ontological status, contrasting to the private companies.

One central premise that guides our research from the beginning is open knowledge and free software. All the knowledge represented by the ontologies we use is on the public domain and comes from the BFO open portal, a well-founded and widely accepted ontological framework. The software techniques and standards come from W3C, and they are applied to a well-documented method called OBDA. The motivation to adopt this set of resources is based on the value we place on voluntary cooperation and the search for consensus among academics and professionals. It is the style of agreements that this thesis proposes, one that is founded on public standards and reusable artifacts.

The thesis examines the literature on Public Ontologies on e-Gov produced over the past four years through a systematic review. Although several studies – 236 articles selected and 49 papers examined – no strong focus on ontologies was found. The results demonstrate some opportunities for the use of KR techniques in the e-Gov field. The philosophical principles explored in this thesis demonstrate the value of having a factual theoretical basis to discuss a complex subject such as Social Acts.

The approach of the Documents Acts Theory was beneficial for this study since it allowed us to make comparisons between government and private institutions. The generalization that proved to be adequate and justified was used to demonstrate the D-Acts ontology application in the conceptual modeling of a government process, using real data. This case study allows us to consider that this way of modeling can indicate solutions and alternatives to our research when used in other government systems, as mentioned earlier's initial concerns.

This research differs from previous studies, which are reported in the SLR, in integrating KR's ontological approach to developing a fully functional software component, which can be used in a real project to integrate relational databases. The combined use of a software engineering technique with a set of well-build ontologies indicates new perspectives for modern systems development. The use of domain ontologies in AI systems is a promising approach both in Natural Language Processing (NLP) and Automatic Inference applications.

Incorporating KR techniques into software projects can help solve challenging problems such as requirements analysis and conceptual modeling. The integration of heterogeneous data can also be performed without duplication of data, preserving databases in the current state. The technique of creating virtual triples is the alternative demonstrated in this work. We intend that this technique can compose a comprehensive set of tools, which become known and widespread. Moreover, the dissemination of these tools can encourage the use of KR and its related techniques in software engineering projects. Our glimpse of the future sees the Semantic Web as an open-world idea and democratically shared knowledge. Through the formalization provided by KR, we foresee the full, robust, controlled, and ethical use of Web resources. Our research question was answered, although we admit that it is only a part of the answer. We can now develop systems and services in the cloud that indeed incorporate formalized knowledge, but some issues remain open: who will provide and formalize this knowledge?

We need to clarify to everyone involved in the human endeavor that the Internet has become the importance of collaborative work and shared knowledge. We need experts from all fields willing to learn how to represent their knowledge on the web using each domain's ontologies. We need governments to seek the uniformity of terms and concepts, strengthening the establishment of standards and KOS's public availability. Companies should be encouraged to make their domain ontologies available for society's benefit, and software developers should be taught, from college, to use the appropriate philosophical principles to represent the world.

This research illustrates the case of data integration and raises the question of how to use these techniques and methods. We work with our case study in the laboratory, using some advanced IDEs, in a controlled environment. It can be said that this solution is not directly applicable to a real software project, but our experience and expertise suggest that it is sure that it is. What we need is education, a vision of the future, and a willingness to share knowledge to be formalized in a public way.

Our intention in continuing studies is to advance the research provided here to complement the current e-Gov efforts in Brazil. We have the opportunity to adopt a top-level ontology that improves our standardization of terms and concepts following world trends. Creating a government portal to store ontologies, vocabularies, taxonomies, and all KR standards in a reusable way would be a significant advance for the country.

8.2. Statement of Vision

It is common to refer to the fabulous world of existing systems on the planet only as "the cloud." This diffuse and imprecise figure is part of our most significant modern social challenges, from the maintenance of unfair economic systems to the collapse of social and political consensus through climate and ecological problems. This construction is typically socio-technical, as it contains much technology and cultural, historical, and social foundations. However, the most crucial formative characteristic of the "cloud" is the lack of a single purpose. One cannot identify explicit collective intentionality in the construction and integration of global systems. The opacity of the "cloud" contributes to its sector-based appropriation, which can be made by individuals, social groups, corporations, and governments.

Here we take James Bridle's phrase in his book "The New Dark Ages," which was inspiring for this final remark:

> Across the sciences and society, in politics and education, in warfare and commerce, new technologies do not merely augment our abilities, but actively shape and direct them, for better and for worse. It is increasingly necessary to be able to think new technologies in different ways, and to be critical of them, in order to meaningfully participate in that shaping and directing. If we do not understand how complex technologies function, how systems of technologies interconnect, and how systems of systems interact, then we are powerless within them, and their potential is more easily captured by selfish elites and inhuman corporations. Precisely because these technologies interact with one another in unexpected and often-strange ways, and because we are completely entangled with them, this understanding cannot be limited to the practicalities of how things work: it must be extended to how things came to be, and how they continue to function in the world in ways that are often invisible and interwoven. What is required is not understanding, but literacy (BRIDLE, 2018).

Regarding the "cloud," humanity faces today two challenges, parallel and related: the literacy of people and the formalization of knowledge. We need information users who are critical concerning the games of demand and let others know that they understand the meaning of the common idea "if you do not pay for the product, it is because you are the product." Besides, we also need to make technical staff aware of democratic principles in the idealization, design, and development of these new global systems. This activity is too critical to be dominated by corporations and governments.

The challenges facing users, analysts, teachers, parents, governments are to collaborate so that the immensity of our network of knowledge is placed at the service of humanity and global welfare. We must recognize that all the data and algorithms that exist today are artifacts, products of human work, of men and women, at all times. Besides, their understanding and use is a shared right. From now on, I put myself at the service of this vision.

REFERENCES

ACKROYD, S.; FLEETWOOD, S. **Realist Perspectives on Management and Organisations (Critical Realism: Interventions)**. London: Routledge, 2000.

ALLEMANG, D.; HENDLER, J. **Semantic Web for the Working Ontologist: effective modelling in RDFS and OWL**. 2nd. ed. Waltham, MA: Elsevier, 2011.

ALLISON, G. T. Public and private management: Are they fundamentally alike in all unimportant respects? In: SHAFRITZ, J. M.; HYDE, A. C. (Eds.). . **Classics of Public Administration**. Belmont: Wordsworth, 1992. p. 457–474.

ALMEIDA, M. B. **Inter-operabilidade entre fontes heterogêneas: um meta-modelo baseado em ontologias**. [s.l.] UFMG, 2002.

ALMEIDA, M. B. Revisiting Ontologies: A Necessary Clarification. **Journal of the American Society for Information Science and Technology**, v. 64, n. 8, p. 1682–1693, 2013.

ALMEIDA, M. B.; PESSANHA, C. P.; BARCELOS, R. Information Architecture for Organizations: An Ontological Approach. **IntechOpen**, p. 26, 2018.

ALMEIDA, M. B.; RIBEIRO, E.; BARCELOS, R. Toward a Document-Centered Ontological Theory for Information Architecture in Corporations. **The Journal of the Association for Information Science and Technology**, 2019.

ALMEIDA, M. B.; RODRIGUES BARBOSA, R. Ontologies in Knowledge Management Support: A Case Study. **Journal of the American Society for Information Science and Technology**, v. 60, n. 10, p. 2032–2047, 2009.

ALMEIDA, M. B.; SILVA, F. B.; BROCHHAUSEN, M. Ontological approach to the normative dimension of organizations: an application of Documents Acts Ontology. **Ciência da Informação**, v. 46, n. 1, 2017.

ALMEIDA, M. B.; SOUZA, R. R.; FONSECA, F. Semantics In The Semantic Web: A Critical Evaluation. **Knowledge Organization Journal**, v. 38, n. 3, p. 187–203, 2011.

ALMEIDA, M. B.; TEIXEIRA, L. M. D. Revisitando os fundamentos da classificação: uma análise crítica sobre teorias do passado e do presente. **Perspectivas**

em Ciência da Informação, v. 25, p. 28–56, 2020.

ALVAREZ-RODRÍGUEZ, J. M.; LABRA-GAYO, J. E.; DE PABLOS, P. O. New trends on e-Procurement applying semantic technologies: Current status and future challenges. **Computers in Industry**, v. 65, n. 5, p. 800–820, 2014.

ANTONIOLI, N. et al. Ontology-Based Data Access: The Experience at the Italian Department of Treasury. **CAiSE Industrial Track**, v. 1017, p. 9–16, 2013.

ANTONIOLI, N. et al. **Ontology-based Data Management for the Italian Public Debt**. Formal Ontology in Information Systems: Proceedings of the Eighth International Conference (FOIS 2014). **Anais**...2014

APOSTOLOU, D. et al. Towards a Semantically-Driven Software Engineering Environment for eGovernment. In: BÖHLEN, M. et al. (Eds.). . **E-Government: Towards Electronic Democracy**. Lecture Notes in Computer Science. [s.l.] Springer, 2005. v. 3416p. 157–168.

ARAÚJO, L. S. O.; SANTOS, M. T.; SILVA, D. A. **The Brazilian Federal Budget Ontology – A Semantic Web Case of Public Open Data**. (ACM Digital Library, Ed.)The 7th International Conference on Management of Digital EcoSystems. **Anais**...Caraguatatuba, Brazil: 2015

ARP, R.; SMITH, B.; SPEAR, A. D. **Building Ontologies with Basic Formal Ontology**. London: Mit Press, 2015.

ASAMOAH, D.; SHARDA, R. Adapting CRISP-DM Process for Social Network Analytics: Application to Healthcare. **AMCIS 2015 Proceedings**, 26 jun. 2015.

ATTARD, J. et al. A systematic review of open government data initiatives. **Government Information Quarterly**, v. 32, n. 4, p. 399–418, out. 2015.

AUSTIN, J. L. **How to do things with words**. London: Oxford University Press, 1962.

BATES, M. J. **Understanding information retrieval systems: Management, types, and standards**. [s.l.] CRC Press, 2011.

BENNETT, M. The financial industry business ontology: Best practice for big data. **Journal of Banking Regulation**, v. 14, n. 3–4, p. 255–268, 2013.

BENNETT, M. **Reuse of Semantics in Business Applications**. 8th International Conference on Formal Ontology in Information Systems (FOIS 2014). **Anais**...Rio de Janeiro: CEUR Workshop Proceedings, 2014

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, p. 4, 2001.

BIENVENU, M. et al. Ontology-based data access. **Proceedings of the 32nd symposium on Principles of database systems - PODS ’13**, p. 213, 2013.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data-the story so far. **International journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1–22, 2009.

BLUMBERG, P. I. The Multinational Challenge to Corporation Law: The Search for a New Corporate Personality. **International Journal of Legal Information**, 1993.

BORGIDA, A. Conceptual Modeling of Information Systems. In: BRODIE, M. L. (Ed.). . **On Knowledge Management Based Systems**. NY, USA: Springer Verlag, 1986. p. 461–469.

BORTOLATO, F. **Ligando Dados Governamentais Abertos: uma ontologia do Processo Legislativo de São Paulo**. [s.l.] IPT SP, 2014.

BOYNE, G. A. Public and private management: What’s the difference? **Journal of Management Studies**, v. 39, n. 1, p. 97–122, 2002.

BRIDLE, J. **New Dark Age: Technology and the End of the Future**. [s.l.] Verso Books, 2018.

BROCHHAUSEN, M.; ALMEIDA, M. B.; SLAUGHTER, L. Towards a formal representation of document acts and the resulting legal entities. **Johanssonian Investigations**, v. 120, n. 139, p. 120–140, 2013.

BROWNE, O. et al. Distributed data and ontologies: An integrated semantic web architecture enabling more efficient data management. **Journal of the Association for Information Science and Technology**, 2019.

BULUT, C.; YEN, B. P. C. E-procurement in public sector: A global overview.

Electronic Government, v. 10, n. 2, p. 189–210, 2013.

ÇAĞDAŞ, V.; STUBKJÆR, E. A SKOS vocabulary for Linked Land Administration: Cadastre and Land Administration Thesaurus. **Land Use Policy**, v. 49, p. 668–679, 1 dez. 2015.

CALVANESE, D. et al. Ontop: Answering SPARQL queries over relational databases. **Semantic Web**, v. 8, n. 3, p. 471–487, 2017.

CEUSTERS, W. **An information artifact ontology perspective on data collections and associated representational artifacts**. Studies in Health Technology and Informatics. **Anais**...IOS Press, 2012

CHEN, P. P. S. The Entity-Relationship Model—toward a Unified View of Data. **ACM Transactions on Database Systems (TODS)**, v. 1, n. 1, p. 9–36, 1 mar. 1976.

CHEVITARESE, M. V. A.; BAX, M. P. Necessity and Feasibility of Adopting LOD Standard by National Congress: a study in the context of the public budget. **Informação e Sociedade - Estudos**, v. 24, n. 1, p. 73–94, 2014.

CIEPLEY, D. Beyond public and private: Toward a political theory of the corporation. **American Political Science Review**, v. 107, n. 1, p. 139–158, 2013.

COCCHIARELLA, N. Nominalism and conceptualism as predicative second order theories of predication. **Notre Dame Journal of Formal Logic**, v. 21, 1980.

COLLIS, J.; HUSSEY, R. **Business research: A practical guide for undergraduate and postgraduate students**. [s.l.] Macmillan International Higher Education, 2013.

CONSOLI, S. et al. Producing Linked Data for Smart Cities: The Case of Catania. **Big Data Research**, v. 7, p. 1–15, 1 mar. 2017.

DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What is a knowledge representation? **AI magazine**, v. 14, n. 1, p. 17, 1993.

DING, L. et al. TWC LOGD: A portal for linked open government data ecosystems. **Journal of Web Semantics**, 2011.

DIPPERT, R. R.; LYAS, C. Artifacts, Art Works, and Agency. **The Journal of Aesthetics and Art Criticism**, v. 52, n. 3, p. 367, 1994.

DOS SANTOS BRITO, K. et al. Using Parliamentary Brazilian Open Data to Improve Transparency and Public Participation in Brazil. **Proceedings of the 15th Annual International Conference on Digital Government Research (dg.o 2014)**, 2014.

DU, J.; ZHOU, L. Improving financial data quality using ontologies. **Decision Support Systems**, v. 54, n. 1, p. 76–86, 2012.

DUMMETT, M. **Frege: Philosophy of language**. [s.l.] Harvard University Press, 1981.

EL IDRISSI ESSERHROUCHNI, O.; FRIKH, B.; OUHBI, B. **Building ontologies: A state of the art, and an application to Finance domain**. 2014 International Conference on Next Generation Networks and Services (NGNS). **Anais**...Casablanca: IEEE, 2014

EUROPEAN, U. **e-Government Core Vocabularies handbook**. [s.l: s.n.].

FABBRI, S. et al. **Managing Literature reviews information through visualization**. ICEIS 2012 - Proceedings of the 14th International Conference on Enterprise Information Systems. **Anais**...2012

FABBRI, S. et al. **Improvements in the StArt tool to better support the systematic review process**. Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering - EASE '16. **Anais**...2016

FAÇANHA, R. L.; CAVALCANTI, M. C. **On the Road to Bring Government Legacy Systems Data Schemas to Public Access**. 8th International Conference on Formal Ontology in Information Systems (FOIS 2014). **Anais**...Rio de Janeiro: CEUR Workshop Proceedings, 2014

FALBO, R. D. A. et al. **Towards an enterprise ontology pattern language**. Proceedings of the ACM Symposium on Applied Computing. **Anais**...Association for Computing Machinery, 2014

FERNEDA, E.; CRUZ, F. W. **Ontological Engineering for Public and**

Private Procurement. Brasília: [s.n.].

FRAGKOU, P.; GALIOTOU, E.; MATSAKAS, M. Enriching the e-GIF Ontology for an Improved Application of Linking Data Technologies to Greek Open Government Data. **Procedia - Social and Behavioral Sciences**, 2014.

FREGE, G. Sobre o Sentido e a Referência (Über Sinn und Bedeutung). **Wittgenstein Studien**, v. 1, n. 1, 1994.

FREITAS, F. Ontologias e a Web Semântica. **Anais do XXIII Congresso da Sociedade Brasileira de Computação**, p. 1–52, 2003.

GANDON, F. A survey of the first 20 years of research on semantic web and linked data. **Ingenierie des Systemes d'Information**, v. 23, n. 3–4, p. 11–56, 2018.

GEELS, F. W. From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. **Research Policy**, v. 33, n. 6–7, 2004.

GEIGER, C. P.; VON LUCKE, J. Open Government and (Linked) (Open) (Government) (Data). **JeDEM - eJournal of eDemocracy and Open Government**, v. 4, n. 2, p. 265–278, 2012.

GILBERT, M. **On social facts**. [s.l.] Princeton University Press, 1992.

GRENON, P.; SMITH, B. SNAP and SPAN: Towards dynamic spatial ontology. **Spatial Cognition and Computation**, 2004.

GUARINO, N. Formal Ontology and Information Systems. **Formal Ontology in Information Systems: Proceedings of the 1st International Conference**, v. 46, n. June, p. 3–15, 1998.

GUBA, E. **The paradigm dialogue**. Newbury Park: Sage Publications, 1990.

GUIZZARDI, G. **Ontological foundations for conceptual modeling with applications**. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). **Anais**...2012

HARARI, Y. N. **Sapiens: A brief history of humankind**. [s.l.] Random House, 2014.

HENDLER, J. et al. US government linked open data: Semantic.data.gov. **IEEE Intelligent Systems**, v. 27, n. 3, p. 25–31, 2012.

HIRST, G. Ontology and the Lexicon. In: STAAB, S.; STUDER, S. (Eds.). . **Handbook on ontologies**. 2nd. ed. New York: Springer, 2009.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, 2003.

HODGSON, G. M. What Are Institutions? **Journal of Economic Issues**, v. 40, n. 1, p. 1–25, 4 mar. 2006.

HORRIDGE, M. et al. A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. **University of Manchester**, p. 118, 2004.

HUBER, S. et al. **DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model**. Procedia CIRP. **Anais**...Elsevier B.V., 1 jan. 2019

HUNTER, J. **An ontological framework to support evidence-based policy making using global city indicators**. 2015 IEEE 1st International Smart Cities Conference, ISC2 2015. **Anais**...Institute of Electrical and Electronics Engineers Inc., 24 dez. 2015

IWAI, K. Persons, Things and Corporations: The Corporate Personality Controversy and Comparative Corporate Governance. **American Journal of Comparative Law**, 1999.

IWAI, K. What is corporation? The corporate personality controversy and the fiduciary principle in corporate governance. In: **Legal Orderings and Economic Institutions**. [s.l: s.n.].

IWAI, K. What is corporation? In: **Legal Orderings and Economic Institutions**. [s.l: s.n.].

JOHANSSON, I. **Ontological investigations: an inquiry into the categories of nature, man, and society.** New York, NY, USA: [s.n.].

KADADI, A. et al. **Challenges of data integration and interoperability in**

big data. 2014 IEEE International Conference on Big Data (Big Data). **Anais**...IEEE, out. 2014

KALGOTRA, P.; SHARDA, R. **Progression analysis of signals: Extending CRISP-DM to stream analytics**. Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016. **Anais**...Institute of Electrical and Electronics Engineers Inc., 2016

KHLENTZOS, D. **Naturalistic Realism and the Antirealist Challenge**. [s.l.] The MIT Press, 2018.

KINGSTON, J.; SCHAFER, B.; VANDENBERGHE, W. Towards a financial fraud ontology: A legal modelling approach. **Artificial Intelligence and Law**, v. 12, n. 4, p. 419–446, 2004.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. TR/SE-0401, p. 28, 2004.

KOGALOVSKY, M. R. Ontology-based data access systems. **Programming and Computer Software**, v. 38, n. 4, p. 167–182, 2012.

KRAINER, R. E. Regulating Wall Street: The Dodd–Frank Act and the New Architecture of Global Finance, a review. **Journal of Financial Stability**, v. 8, n. 2, p. 121–133, 2012.

KRIJGSMAN, M.; HOFMAN, W.; HOUBEN, G.-J. Social web ontology for public services. In: **Case Studies in e-Government 2.0**. [s.l.] Springer, 2015. p. 199–215.

KURGAN, L. A.; MUSILEK, P. A survey of Knowledge Discovery and Data Mining process models. **Knowledge Engineering Review**, 2006.

LAMHARHAR, H.; CHIADMI, D.; BENHLIMA, L. How semantic technologies transform e-government domain. **Transforming Government: People, Process and Policy**, v. 8, n. 1, p. 49–75, 17 mar. 2014.

LANDGREBE, J.; SMITH, B. **Making AI Meaningful Again**. [s.l: s.n.].

LANDGREBE, J.; SMITH, B. **There is no Artificial General Intelligence**. [s.l: s.n.].

LAUDON, K.; LAUDON, J. P. **Sistemas de informação gerenciais**. 9a. ed. São Paulo: Pearson, 2011.

LINDERS, D. From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. **Government Information Quarterly**, v. 29, n. 4, p. 446–454, 2012.

LINTHICUM, D. S. **Enterprise application integration**. [s.l.] Addison-Wesley Professional, 2000.

LUNA-REYES, L. F.; BERTOT, J. C.; MELLOULI, S. Open Government, Open Data and Digital Government. **Government Information Quarterly**, v. 31, n. 1, p. 4–5, 2014.

MACHEN, A. W. Corporate Personality. **Harvard Law Review**, v. 24, n. 4, p. 253, 1911.

MACKENZIE, N.; KNIPE, S. Research dilemmas: Paradigms, methods and methodology. **Issues in educational research**, v. 16, n. 2, p. 193–205, 2006.

MADNICK, S.; ZHU, H. Improving data quality through effective use of data semantics. **Data & Knowledge Engineering**, v. 59, n. 2, p. 460–475, 2006.

MARCONDES, C. H. et al. Bases ontológicas e conceituais para um modelo do conhecimento científico em artigos biomédicos. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 3, n. 1, 2009.

MARCONI, M.; LAKATOS, E. Fundamentos de metodologia científica. **Editora Atlas S. A.**, p. 310, 2003.

MATHEUS, R.; RIBEIRO, M. M.; VAZ, J. C. Brazil towards government 2.0: Strategies for adopting open government data in national and subnational governments. **Case Studies in E-Government 2.0: Changing Citizen Relationships**, v. 55, n. 11, p. 121–138, 2015.

MCDERMOTT, P. Building open government. **Government Information Quarterly**, v. 27, n. 4, p. 401–413, 2010.

MEIER, K. J.; O'TOOLE, L. J. Comparing Public and Private Management: Theoretical Expectations. 2011.

MENTZAS, G.; FRIESEN, A. **Semantic enterprise application integration for business processes: Service-oriented frameworks**. [s.l.] IGI Global, 2009.

MILTON, S.; KAZMIERCZAK, E. Ontology as Meta-Theory: A Perspective. **Scandinavian Journal of Information Systems**, v. 18, n. 1, p. 85–94, 2006.

MORRISON, A.; STEIN, B. The enterprise data lake: Better integration and deeper analytics. **Technology Forecast: Rethinking integration**, n. 1, 2014.

MP, M. DO P. D. E G. **Estratégia de Governança Digital - EGD**. Brasília: [s.n.].

MP, M. DO P. D. E G. **VCGE - Vocabulário Controlado do Governo Eletrônico**. Brasília: [s.n.].

MUNN, K.; SMITH, B. Introduction: What is Ontology for? In: **Applied Ontology: An Introduction**. [s.l.] De Gruyter, 2013. p. 7–19.

MUSAFIR, V. E. N. Brazilian E-Government Policy and Implementation. In: ALCAIDE MUÑOZ, L.; RODRÍGUEZ BOLÍVAR, M. P. (Eds.). . **International E-Government Development : Policy, Implementation and Best Practice**. Cham: Springer International Publishing, 2018. p. 155–186.

MUSEN, M. A. The protégé project. **AI Matters**, v. 1, n. 4, p. 4–12, 16 jun. 2015.

MYLOPOULOS, J. Conceptual modelling and Telos. **Conceptual Modeling, Databases, and CASE: An integrated view of information systems development**, p. 49–68, 1992.

OKOLI, C.; SCHABRAM, K. A Guide to Conducting a Systematic Literature Review of Information Systems Research. **Sprouts: Working Papers on Information Systems**, v. 10, n. 26, p. 10–26, 2010.

OLIVÉ, A. **Conceptual Modeling of Information Systems**. [s.l.] Springer Science & Business Media, 2007.

ORDIYASA, I. W. et al. **Enhancing Quality of Service for eGovernment interoperability based on adaptive ontology**. Proceedings - 2016 2nd International Conference on Science and Technology-Computer, ICST 2016. **Anais**...Institute of

Electrical and Electronics Engineers Inc., 13 mar. 2017

PALMIRANI, M. et al. **Pronto: Privacy ontology for legal compliance**. (A. K\Ho, E. Francesconi, Eds.)Proceedings of the European Conference on e-Government, ECEG. **Anais**...Cham: Springer International Publishing, 2018

PINTO, J. A.; ALMEIDA, M. B. **The role of Controlled Vocabularies in projects for Electronic Government - An overview of the Brazilian context**. CEUR Workshop Proceedings. **Anais**...2018

PINTO, J. A.; ALMEIDA, M. B. Ontologias Públicas Sobre Governo Eletrônico: uma Revisão Sistemática da Literatura. **In press**, 2020a.

PINTO, J. A.; ALMEIDA, M. B. An Applied Ontology-Oriented Study Case to Distinguish Public and Private Institutions Through Their Documents. **In press**, 2020b.

PINTO, J. A.; BAX, M. P. **Domain ontologies and semantic technologies to promote government data access**. CEUR Workshop Proceedings. **Anais**...2017

PRESTON, B. Biological and Cultural Proper Functions in Comparative Perspective. In: **Functions in Biological and Artificial Worlds**. [s.l.] The MIT Press, 2013. p. 37–50.

PRIETO-DIAZ, R. Domain Analysis: An Introduction. **Software Engineering Notes**, v. 15, n. 2, p. 47–54, 1990.

RAINEY, H. G.; CHUN, Y. H. Public and Private Management Compared. In: **The Oxford Handbook of Public Management**. [s.l.] Oxford University Press, 2009.

RAMAPRASAD, A.; SÁNCHEZ-ORTIZ, A.; SYN, T. **An ontology of eGovernment**. Lecture Notes in Computer Science. **Anais**...2015a

RAMAPRASAD, A.; SÁNCHEZ-ORTIZ, A.; SYN, T. **Gaps in Local eGovernment Research: An Ontological Analysis**. Twenty-first Americas Conference on Information Systems, Puerto Rico, 2015. **Anais**...2015b

RAMIREZ, J. P. R. **Vocabulário Controlado do Governo Eletrônico (VCGE): Uma análise com base em em critérios aplicáveis a taxonomias e tesauros**. [s.l.] Universidade Federal de Santa Catarina, 2015.

RIBEIRO, C. J. S.; ALMEIDA, R. F. DE. **Dados Abertos Governamentais (Open Government Data): Instrumento para Exercício de Cidadania pela Sociedade**. XI ENANCIB - Encontro Nacional de Pesquisa em Ciência da Informação. **Anais**...2011

RIBEIRO, C. J. S.; PEREIRA, D. V. A publicação de dados governamentais abertos: Proposta de revisão da classe sobre previdência social do vocabulário controlado do governo eletrônico. **Transinformacao**, 2015.

RODRIGUEZ-MURO, M.; KONTCHAKOV, R.; ZAKHARYASCHEV, M. **Ontology-Based Data Access: Ontop of Databases**. International Semantic Web Conference. **Anais**...Springer, 2013

RUMBAUGH, J. et al. **Object-oriented modeling and design**. [s.l.] Prentice-hall Englewood Cliffs, NJ, 1991. v. 199

RYHANEN, K.; PAIVARINTA, T.; TYRVAINEN, P. **Generic data models for Semantic e- Government interoperability: Literature Review**. Innovation and the Public Sector. **Anais**...2014

SAMEK, W.; WIEGAND, T.; MULLER, K.-R. Explainable Artificial Intelligence: Understanding, Visualizing And Interpreting Deep Learning Models. **Empirical Software Engineering**, v. 19, n. 3, p. 465–500, 2014.

SANTOS, H. et al. From Data to City Indicators: A Knowledge Graph for Supporting Automatic Generation of Dashboards. 2017.

SANTOS, L. A. et al. **Using an Ontology Network for Data Integration: A Case in the Public Security Domain**. (J. Carbonara et al., Eds.)Ontobrás 2018 - XI Seminar on Ontology Research in Brazil. **Anais**...São Paulo: 2018

SANTOSO, H. A. et al. An ontological crawling approach for improving information aggregation over eGovernment websites. **Journal of Computer Science**, v. 12, n. 9, p. 455–463, 2016.

SASTRY, S. H.; BABU, P. Implementation of CRISP Methodology for ERP Systems. **International Journal of Computer Science Engineering**, v. 2, n. 5, 2013.

SAVAGET, P. et al. The theoretical foundations of sociotechnical systems change for sustainability: A systematic literature review. **Journal of Cleaner Production**, v. 206, p. 878–892, 1 jan. 2019.

SAVO, D. F. et al. MASTRO at Work: Experiences on Ontology-based Data Access. **Proceedings 23rd Intl. Workshop on Description Logics**, 2010.

SCHAFER, F. et al. **Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes**. IEEE International Conference on Technology Management, Operations and Decisions. **Anais**...IEEE, 2018

SEARLE, J. R. **The construction of social reality**. [s.l.] Simon and Schuster, 1995.

SEARLE, J. R. **Making the Social World: The Structure of Human Civilization**. [s.l.] Oxford University Press UK, 2010.

SEARLE, J. R.; VANDERVEKEN, D. Speech Acts and Illocutionary Logic. In: **Logic, Thought and Action**. [s.l: s.n.].

SKJÆVELAND, M. G. et al. Engineering ontology-based access to real-world data sources. **Journal of Web Semantics**, v. 33, p. 112–140, 2015.

SLEPICKA, J. et al. **KR2RML: An Alternative Interpretation of R2RML for Heterogeneous Sources**. [s.l: s.n.].

SMIRAGLIA, R. P. Empirical Methods for Knowledge Evolution across Knowledge Organization Systems. **Knowledge Organization Journal**, v. 43, n. 5, 2016.

SMITH, B. Formal ontology, common sense and cognitive science. **International Journal of Human-computer Studies**, v. 43, n. 5, 1995.

SMITH, B. Ontology and Information Systems. **Stanford Encyclopedia of Philosophy**, 2002.

SMITH, B. John Searle: From speech acts to social reality. In: CAMBRIDGE PRESS (Ed.). . **John Searle**. [s.l.] Cambridge University Press, 2003. p. 1–33.

SMITH, B. et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nat Biotechnol**, v. 25, n. 11, 2007.

SMITH, B. How to do things with documents. **Rivista di estetica**, n. 50, p. 179–198, 2012.

SMITH, B. Document acts. In: KONZELMANN-ZIV, A.; SCHMID, H. B. (Eds.). . **Institutions, Emotions, and Group Agents: Contributions to Social Ontology**. 1st. ed. Dordrecht: Springer Dordrecht Heidelberg, 2014. p. 19–31.

SMITH, B.; WELTY, C. **Ontology: Towards a new synthesis**. Formal Ontology in Information Systems. **Anais**...2001

SOMMERVILLE, I. **Software Engineering**. [s.l.] Pearson Education, 2016.

SOUROUNI, A. M. et al. Towards the government transformation: An ontology-based government knowledge repository. **Computer Standards and Interfaces**, v. 32, n. 1–2, p. 44–53, 2010.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. **Knowledge Organization Journal**, v. 39, n. 3, p. 179–192, 2012.

STOREY, V. C.; TRUJILLO, J. C.; LIDDLE, S. W. **Research on conceptual modeling: Themes, topics, and introduction to the special issueData and Knowledge Engineering**Elsevier B.V., , 1 jul. 2015.

TAHERIYAN, M. et al. Learning the semantics of structured data sources. **Web Semantics: Science, Services and Agents on the World Wide Web**, v. 37, n. 0, 2016.

THOMASSON, A. L. The easy approach to ontology. **Axiomathes**, v. 19, n. 1, p. 1–15, mar. 2009.

TUOMELA, R. **Social ontology: Collective intentionality and group agents**. [s.l.] Oxford University Press, 2016.

VAN COMPERNOLLE, M. et al. **State-of-the-Art Assessment on the Implementations of International Core Data Models for Public Administrations**. 9th International Conference on Theory and Practice of Electronic Governance. **Anais**...Montevideo, Uruguay: ACM Digital Library, 2016

VENTER, J.; DE WAAL, A.; WILLERS, C. **Specializing CRISP-DM for**

evidence mining. IFIP International Federation for Information Processing. **Anais**...Springer, 2007

WAND, Y.; STOREY, V. C.; WEBER, R. An Ontological Analysis of the Relationship Construct in Conceptual Modeling. **ACM Transactions on Database Systems**, 1999.

WANG, Z.; FOX, M. S. **Expenditure, revenue and taxes — Towards a standard for representing city finance open data**. 2016 IEEE International Smart Cities Conference (ISC2). **Anais**...IEEE, set. 2016

WEINTRAUB, J. The Theory and Politics ot the Public/Private Distinction. In: **Public and Private in Thought and Practice**. 1st. ed. Chicago: University of Chicago, 1997. p. 42.

WELCH, E. W.; HINNANT, C. C.; MOON, M. J. **Linking citizen satisfaction with e-government and trust in governmentJournal of Public Administration Research and Theory**, 2005.

WIJNHOVEN, F.; EHRENHARD, M.; KUHN, J. Open government objectives and participation motivations. **Government Information Quarterly**, v. 32, n. 1, p. 30–42, 2015.

WILLISTON, S. History of the Law of Business Corporations before 1800. **Harvard Law Review**, v. 2, n. 4, p. 19, 1888.

WIRTH, R. **CRISP-DM : Towards a Standard Process Model for Data Mining**. International Conference on the Practical Application of Knowledge Discovery and Data Mining. **Anais**...2000

WOWCZKO, I. A Case Study of Evaluating Job Readiness with Data Mining Tools and CRISP-DM Methodology. **International Journal for Infonomics**, v. 8, p. 1066–1070, 2015.

YUN, Z.; WEIHUA, L.; YANG, C. **Applying balanced scorecard strategic performance management to CRISP-DM**. Proceedings - 2014 International Conference on Information Science, Electronics and Electrical Engineering, ISEEE 2014. **Anais**...Institute of Electrical and Electronics Engineers Inc., 5 nov. 2014

ZAMBONI, A. et al. **StArt: Uma Ferramenta Computacional de Apoio à Revisão Sistemática**. Salão de Ferramentas. Congresso Brasileiro de Software. **Anais**...Salvador: 2010

ZHOU, S. et al. **R2RML Processor for Materializing RDF View of Relational Data: Algorithms and Experiments**. (V. Monfort, K.-H. Krempels, Eds.)Web Information System and Application Conference. **Anais**...Barcelona, Spain: IEEE, 2013

ZUNIGA, G. L. **An Ontology Of Economic Objects**, 1 abr. 1999.

ZUNIGA, G. L. **Ontology: Its transformation from philosophy to information systems**. 2nd Conference on Formal Ontology in Information Systems. **Anais**...2001

Annex 1 – OWL File


```xml
<?xml version="1.0"?>
<Ontology xmlns="http://www.w3.org/2002/07/owl#"
     xml:base="http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-2"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:xml="http://www.w3.org/XML/1998/namespace"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     ontologyIRI="http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-2">
    <Prefix name="" IRI="http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-
ontology-2#"/>
    <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#"/>
    <Prefix name="rdf" IRI="http://www.w3.org/1999/02/22-rdf-syntax-ns#"/>
    <Prefix name="xml" IRI="http://www.w3.org/XML/1998/namespace"/>
    <Prefix name="xsd" IRI="http://www.w3.org/2001/XMLSchema#"/>
    <Prefix name="obda" IRI="https://w3id.org/obda/vocabulary#"/>
    <Prefix name="rdfs" IRI="http://www.w3.org/2000/01/rdf-schema#"/>
    <Import>http://purl.obolibrary.org/obo/iao/2019-06-07/iao.owl</Import>
    <Declaration>
        <Class IRI="#budget"/>
    </Declaration>
    <Declaration>
        <Class IRI="#line_of_a_budget"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_budget_action"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_budget_office"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_budget_value"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_spend_action"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_spend_office"/>
    </Declaration>
    <Declaration>
        <DataProperty IRI="#has_spend_value"/>
    </Declaration>
    <SubClassOf>
        <Class IRI="#budget"/>
        <Class IRI="http://purl.obolibrary.org/obo/IAO_0000310"/>
    </SubClassOf>
    <SubClassOf>
        <Class IRI="#line_of_a_budget"/>
        <Class IRI="http://purl.obolibrary.org/obo/IAO_0000314"/>
    </SubClassOf>
</Ontology>

<!-- Generated by the OWL API (version 4.5.9.2019-02-01T07:24:44Z)
https://github.com/owlcs/owlapi -->
```

Annex 2 – R2RML file, Turtle format

```
@prefix rr:    <http://www.w3.org/ns/r2rml#> .
@prefix :      <http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-2#> .
@prefix obda:  <https://w3id.org/obda/vocabulary#> .
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl:   <http://www.w3.org/2002/07/owl#> .
@prefix xml:   <http://www.w3.org/XML/1998/namespace> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .

<urn:Budget_1>  a              rr:TriplesMap ;
     rr:logicalTable       [ a           rr:R2RMLView ;
                     rr:sqlQuery  "select * from budget.budget2019;"
                   ] ;
     rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                     rr:objectMap  [ a           rr:TermMap , rr:ObjectMap ;
                               rr:column    "budget_action" ;
                               rr:datatype  xsd:string ;
                               rr:termType  rr:Literal
                             ] ;
                     rr:predicate  :has_budget_action
                   ] ;
     rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                     rr:objectMap  [ a           rr:TermMap , rr:ObjectMap ;
                               rr:column    "budget_office" ;
                               rr:datatype  xsd:string ;
                               rr:termType  rr:Literal
                             ] ;
                     rr:predicate  :has_budget_office
                   ] ;
     rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                     rr:objectMap  [ a           rr:ObjectMap , rr:TermMap ;
                               rr:column    "budget_value" ;
                               rr:datatype  xsd:string ;
                               rr:termType  rr:Literal
                             ] ;
                     rr:predicate  :has_budget_value
                   ] ;
     rr:subjectMap         [ a           rr:TermMap , rr:SubjectMap ;
                     rr:class     :line_of_a_budget ;
                     rr:template
"http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-
2#budget/budget2019/{budget_office}/{budget_action}" ;
                     rr:termType  rr:IRI
                   ] .

<urn:Spend_1>  a              rr:TriplesMap ;
     rr:logicalTable       [ a           rr:R2RMLView ;
                     rr:sqlQuery  "select * from spend.spend2019;"
                   ] ;
     rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                     rr:objectMap  [ a           rr:ObjectMap , rr:TermMap ;
                               rr:column    "spend_office" ;
                               rr:datatype  xsd:string ;
                               rr:termType  rr:Literal
                             ] ;
                     rr:predicate  :has_spend_office
                   ] ;
```

```
        rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                          rr:objectMap  [ a           rr:ObjectMap , rr:TermMap ;
                                        rr:column    "spend_action" ;
                                        rr:datatype  xsd:string ;
                                        rr:termType  rr:Literal
                                      ] ;
                          rr:predicate  :has_spend_action
                        ] ;
        rr:predicateObjectMap  [ a            rr:PredicateObjectMap ;
                          rr:objectMap  [ a           rr:ObjectMap , rr:TermMap ;
                                        rr:column    "spend_value" ;
                                        rr:datatype  xsd:string ;
                                        rr:termType  rr:Literal
                                      ] ;
                          rr:predicate  :has_spend_value
                        ] ;
        rr:subjectMap        [ a           rr:SubjectMap , rr:TermMap ;
                          rr:class     :line_of_a_budget ;
                          rr:template
"http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-
2#spend/spend2019/{spend_office}/{spend_action}" ;
                          rr:termType  rr:IRI
                                        ] .
```

Annex 3 – OBDA File

```
[PrefixDeclaration]
:		http://www.semanticweb.org/usuario/ontologies/2020/4/untitled-ontology-2#
owl:		http://www.w3.org/2002/07/owl#
rdf:		http://www.w3.org/1999/02/22-rdf-syntax-ns#
xml:		http://www.w3.org/XML/1998/namespace
xsd:		http://www.w3.org/2001/XMLSchema#
obda:		https://w3id.org/obda/vocabulary#
rdfs:		http://www.w3.org/2000/01/rdf-schema#

[MappingDeclaration] @collection [[
mappingId Budget
target		:budget/budget2019/{budget_office}/{budget_action} a :line_of_a_budget ; :has_budget_office {budget_office}^^xsd:string ; :has_budget_action {budget_action}^^xsd:string ; :has_budget_value {budget_value}^^xsd:string .
source		select * from budget.budget2019;

mappingId Spend
target		:spend/spend2019/{spend_office}/{spend_action} a :line_of_a_budget ; :has_spend_office {spend_office}^^xsd:string ; :has_spend_action {spend_action}^^xsd:string ; :has_spend_value {spend_value}^^xsd:string .
source		select * from spend.spend2019;
]]
```