

## UMA ONTOLOGIA DE DOMÍNIO PARA A PRESTAÇÃO DE CONTAS DOS GESTORES PÚBLICOS FEDERAIS: ONTOACCOUNT

Reuber da Silva Fonseca

 <http://lattes.cnpq.br/6490326251279486> –  <https://orcid.org/0000-0002-3316-9684>  
[reuberf@gmail.com](mailto:reuberf@gmail.com)

Universidade Federal de Minas Gerais (UFMG)  
Belo Horizonte, MG, Brasil

Gercina Angela Lima

 <http://lattes.cnpq.br/3183050056105009> –  <https://orcid.org/0000-0003-0735-3856>  
[limagercina@gmail.com](mailto:limagercina@gmail.com)

Universidade Federal de Minas Gerais (UFMG)  
Belo Horizonte, MG, Brasil

### RESUMO

A difusão de inovação em gestão pública e o uso de tecnologias em prol do acesso à informação, da transparência, e da participação social são marcas dos esforços governamentais para o desenvolvimento e expansão do paradigma do Governo Aberto no Brasil. O presente artigo pretende divulgar os resultados de uma pesquisa de doutorado que investigou o paradoxo entre a vasta quantidade de dados governamentais disponibilizados na tomada e prestação de contas dos administradores públicos federais e a baixa percepção e apropriação do conhecimento gerado pelo cidadão. Seu propósito central consistiu em realizar uma modelagem do domínio de conhecimento da tomada e da prestação de contas anual dos gestores públicos federais a fim de facilitar e acelerar o desenvolvimento de agentes de software cívicos. Do ponto de vista metodológico, desenvolveu-se uma pesquisa exploratória qualitativa com procedimentos técnicos de pesquisa bibliográfica e documental. Utilizou-se o *framework Protégé 5.5.0* e a metodologia *OntoForInfoScience*, proposta por Mendonça (2015), para desenvolver o protótipo de ontologia. Como resultado, foi criada a ontologia de domínio chamada *OntoAccount* para representar os conceitos e os relacionamentos do domínio, além de possibilitar responder a questionamentos sobre as instâncias do domínio da tomada e prestação de contas anual dos gestores públicos federais.

**Palavras-chave:** Ontologia de domínio. Representação do conhecimento. Prestação de contas.

### A DOMAIN ONTOLOGY FOR THE ACCOUNTABILITY OF FEDERAL PUBLIC MANAGERS: ONTOACCOUNT

### ABSTRACT

The diffusion of innovation in public management and the use of technologies in favor of access to information, transparency, and social participation are hallmarks of government efforts to develop and expand the Open Government paradigm in Brazil. The present article intends to disclose the results of a doctoral research that investigated the paradox between the vast amount of governmental data available in the taking and rendering of accounts of federal public administrators and the low perception and appropriation of the knowledge generated by the citizen. Its main purpose was to carry out a modeling of the domain of knowledge of the taking and annual accountability of federal public managers in order to facilitate and accelerate the development of civic software agents. From the methodological point of view, an exploratory qualitative research was developed with technical procedures of bibliographic and documental research. The *Protégé 5.5.0* framework and the *OntoForInfoScience* methodology, proposed by Mendonça (2015), were used to develop the ontology prototype. As a result, a domain ontology called *OntoAccount* was created to represent the concepts and relationships of the domain, in addition to making it possible to answer questions about the instances of the domain of taking and rendering annual accounts of federal public managers.

**Keywords:** Domain ontology; Knowledge representation; Accountability

DOI <http://dx.doi.org/10.1590/1981-5344/41347>

Recebido em: 29 setembro 2022.

Aceito em: 23 fevereiro 2023.

## INTRODUÇÃO

A evolução tecnológica contínua e ininterrupta dos recursos computacionais contemporâneos ampliou as capacidades de captura, de processamento, de disponibilização e de acesso à informação pública. Do outro lado, a adoção de novos referenciais político-administrativos, pela Administração Pública, baseados no conceito de Governo Aberto, compatibilizou agendas governamentais no intuito de transformar governos em entidades mais responsáveis, responsivas e eficientes, e cidadãos em agentes mais conscientes e participativos.

A difusão de inovação em gestão pública e o uso de tecnologias em prol do acesso à informação, da transparência, e da participação social são marcas dos esforços governamentais para o desenvolvimento e expansão do paradigma do Governo Aberto no Brasil. Nesta seara, é relevante problematizar o cenário de produção, de disponibilização e de consumo de dados abertos governamentais provenientes da tomada e prestação de contas do administrador público federal e discutir alternativas para representar este domínio de conhecimento de modo a aumentar a percepção e a apropriação pelo cidadão.

Dentre os propósitos da prestação de contas anual, se sobrepõe a finalidade de informar aos cidadãos (principais provedores de recursos e destinatários de serviços públicos) o alcance dos objetivos de interesse coletivo estabelecidos. Esta pesquisa, contudo, questiona a efetividade dos dados que compõem a prestação de contas em subsidiar a geração de conhecimento para o controle social sobre o Poder Público. Questiona-se, sobretudo, a ausência de política que apoie a disponibilização dos dados em formato estruturado e semântico a fim de favorecer o reuso das informações por máquinas.

Ao longo dos anos, uma extensa quantidade de dados sobre resultados da gestão pública federal, provenientes dos processos de tomada e de prestação de contas, foi disponibilizada ao público em formatos não estruturados. (ex: *Portable Document Format - PDF*). Contudo, entende-se que a maneira isolada como estes processos são relatados implica barreiras para o reuso e agregação deste conhecimento. Uma ideia, ainda que incompleta, do valor destes dados para os indivíduos e organizações, pode ser alcançada

analisando o seu potencial de reuso e de construção de novos contextos de análise. Logo, a ausência ou incipiência de uma política de disponibilização destes dados em formato estruturado e semântico, tendo como fim o seu reuso, é um problema a ser enfrentado por esta pesquisa.

Neste cenário político e tecnológico, emerge a necessidade de "libertar dados" dos relatórios que compõem a tomada e prestação de contas e de se desenvolver estratégias, abordagens e instrumentos para a construção de modelos conceituais representativos desta realidade e consistentes com as necessidades e interesses dos cidadãos. Dentre as propostas, as ontologias são instrumentos úteis de organização e representação do conhecimento para representar domínios para uso tanto de humanos como de agentes de software.

Este estudo, portanto, é baseado na premissa de que a ausência de modelagem semântica formal e explícita do domínio de conhecimento da tomada e da prestação de contas anual dos gestores públicos federais prejudica o eventual consumo dos dados públicos por agentes que queiram assimilar o valor público gerado na gestão. No tocante à publicação dos dados públicos provenientes da prestação de contas, cumpre-se alertar que se esta for inadequadamente instituída, constituir-se-á em barreira ao avanço da participação e do controle social. Em última análise, isto pode levar à alienação política do cidadão.

Deste modo, esta pesquisa teve como objetivo desenvolver uma modelagem de conhecimento baseada em um protótipo de ontologia instanciada com dados estruturados provenientes do domínio de conhecimento da tomada e da prestação de contas anual dos gestores públicos federais a fim de oferecer suporte ao desenvolvimento de aplicações cívicas. Durante os trabalhos, buscou-se conhecer e revelar significados e relações semânticas presentes no domínio e contribuir para a publicação de dados abertos estruturados e semânticos deste domínio. As relações semânticas foram exploradas por meio de uma solução ontológica com vista a atribuir sentido e significado ao processo e ao conteúdo dos relatórios, oportunizando o reuso de recursos e a troca de informação e conhecimento entre pessoas e aplicações.

## DADOS ABERTOS, DADOS ABERTOS CONECTADOS E WEB SEMÂNTICA

Segundo a *Open Definition* (2014), o significado do termo “aberto”, no contexto do movimento de acesso livre e aberto ao conhecimento, é aquele que qualquer pessoa pode acessar, utilizar, modificar, e compartilhar com exigência máxima de preservação da proveniência e da abertura. A *Open Knowledge Foundation* (OKF), define Dados Abertos como os dados “que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras” (OKF, 2020).

Isotani e Bittencourt (2015) esclarecem que a abertura de dados se concretiza por meio da retirada de controles legais, técnicos e econômicos sobre os dados. Este movimento, contudo, implica em custos e benefícios para produtores e consumidores de dados abertos. Uma proposta de classificação da maturidade das publicações de dados abertos foi apresentada por Berners-Lee (2006). O Esquema 5-Estrelas dos Dados Abertos tem como propósito encorajar pessoas e instituições (em especial as governamentais) a desenvolver ações de publicação e consumo de dados estruturado na Web de forma aberta. O sistema de classificação alvitado atribui maior número de estrelas para os dados mais abertos, isto é, aqueles com maior facilidade de conexão a outros dados. Em síntese, o esquema expressa as seguintes conclusões segundo o número de estrelas:

- 1 Estrela: O recurso está disponível na *Internet* em qualquer formato e com licença aberta. (exigência mínima para que seja considerado dado aberto);
- 2 Estrelas: Seguindo todas as regras acima, o recurso está disponível de forma estruturada;
- 3 Estrelas: Seguindo todas as regras acima, o recurso está disponível em formato não proprietário;
- 4 Estrelas: Seguindo todas as regras acima, o recurso está disponível nos padrões RDF e SPARQL;
- 5 Estrelas: Seguindo todas as regras acima, o recurso conecta seus dados a outros dados a fim de fornecer um contexto.

Kim e Hausenblas (2012), refletindo sobre o esquema de avaliação, apresentado acima, identificou um conjunto de custos e benefícios para

produtores e consumidores de dados abertos em cada nível de classificação. Os resultados foram estruturados no Quadro 1.

**Quadro 1** - custos e benefícios de dados abertos na web

Nível	Categoria	Consumidores	Produtores
1 estrela	Benefícios	Podem visualizar, imprimir, armazenar, usar, modificar e compartilhar os dados.	É fácil publicar os dados e comunicar as regras de licenciamento.
	Custos	Requer desenvolvimento de código para extrair os dados presos ao documento.	
2 estrelas	Benefícios	Podem tudo anteriormente e adicionalmente podem processar os dados diretamente com um software proprietário e exportá-los para outros formatos suportados.	Continua fácil publicar os dados.
	Custos	Para manipular os dados é necessária uma licença de software proprietário.	
3 estrelas	Benefícios	Podem tudo anteriormente e adicionalmente podem usar, modificar e compartilhar os dados de qualquer forma (sem as restrições do software proprietário).	Continua fácil publicar os dados.
	Custos	Os dados continuam no documento e não são diretamente acessíveis na Web.	Pode ser necessário conversores ou plugins para exportar os dados.
4 estrelas	Benefícios	Podem tudo anteriormente e adicionalmente podem acessá-los diretamente para reusar parte ou todos os dados. Podem conectá-los a outros dados.	Tem controle granular detalhado sobre os itens de dados e pode tê-los referenciados por outros publicadores.
	Custos	Entender a estrutura de dados RDF	Requer mais tempo para organizar e representar os dados (definir URIs e padrões para reuso)
5 estrelas	Benefícios	Podem tudo anteriormente e adicionalmente podem descobrir e conectar mais dados relacionados aprendendo sobre a sua organização.	Os dados são encontráveis na Web, aumentando o seu valor. Com o efeito de rede, passam a obter todos os benefícios de consumidores de dados.
	Custos	Devem controlar "erros 404" e links quebrados de dados nas páginas Web e monitorar o conteúdo de outras páginas Web reutilizadas.	Requer investimento de recursos para manter os dados na Web. Requer reparo de links quebrados ou erros 404.

**Fonte:** Elaborado pelo autor baseado no trabalho de Kim e Hausenblas (2012).

Percebe-se, com a leitura do Quadro 1, que a partir da classificação de 4 estrelas do modelo, os itens de dados podem ser referenciados por outros publicadores, possibilitando o efeito de rede. Também é verdade que este compromisso com a abertura de dados requer mais tempo e investimento para organizar e representar os dados a serem consumidos.

O desenvolvimento da Web Semântica<sup>1</sup>, cunhou o surgimento do conceito de Dados Conectados (do Inglês, *Linked Data*). Segundo Isotani e Bittencourt (2015), “este pode ser definido como um conjunto de boas práticas para publicar e conectar conjuntos de dados estruturados na Web”. Estas boas práticas são recomendações, reveladas em padrões de desenvolvimento, que promovem a interoperabilidade de dados na Web. Os padrões de publicação dos dados conectados na Web permitem a identificação de recursos (objetos reais ou conceitos) habilitando a descoberta e o processamento automatizado de dados por agentes de *software*.

Apesar de existirem ações de publicação de dados conectados de forma fechada na Web, esta pesquisa busca refletir sobre as ações de publicação e consumo de Dados Abertos Conectados (do inglês, *Linked Open Data - LOD*). De acordo com Berners-Lee (2006, tradução nossa), LOD pode ser definido como “um dado conectado que é lançado sob uma licença aberta, que não impede sua reutilização gratuitamente”. Isto é, para ser considerado LOD, o recurso deve existir na Web sem restrições legais que impeçam sua reutilização (princípio de Dados Abertos) e deve observar padrões de desenvolvimento que promovam a interoperabilidade. As iniciativas de LOD, portanto, posicionam-se na interseção dos dois conceitos anteriormente discutidos e refletem decisões de abertura de dados no mais alto nível do Esquema 5-Estrelas dos Dados Abertos proposto por Berners-Lee (2006).

Segundo Berners-Lee, Hendler e Lassila (2001, p. 3, tradução nossa) “para que a Web Semântica funcione, os computadores devem ter acesso a coleções estruturadas de informações e conjuntos de regras de inferência que podem usar para conduzir o raciocínio automatizado.” Para que os sistemas interoperem, eles devem ter a capacidade de descobrir os significados comuns (a semântica) dos dados que encontrar. Logo, esta visão da Web requer compromisso ontológico para se concretizar.

Ontologias podem melhorar o funcionamento da Web de várias maneiras. Eles podem ser usados de uma maneira simples para melhorar a precisão das pesquisas na Web - o programa de pesquisa pode procurar apenas as páginas que se referem a um conceito preciso, em vez de todas aquelas que usam palavras-chave ambíguas. Aplicativos mais avançados usarão ontologias para

---

<sup>1</sup> A Web Semântica pode ser definida como uma extensão da Web atual na qual os dados recebem um significado mais bem definido (BERNERS-LEE; HENDLER; LASSILA, 2001).

relacionar as informações em uma página às estruturas de conhecimento e regras de inferência associadas. (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 7, tradução nossa).

As ontologias, portanto, são fundamentais para transformar os dados em recursos compreensíveis por agentes de *software*. Deste modo elas permitem romper a barreira semântica existente e criam oportunidades para a interação automatizada por meio do reuso da infraestrutura de dados.

## ONTOLOGIAS E APLICAÇÕES CÍVICAS

As ontologias assumem importância crescente no tratamento temático da informação na *Web*, em especial na *Web Semântica*. (BRÄSCHER; CARLAN, 2010). Os estudos crescentes sobre ontologia carregam contribuições de várias disciplinas e ciências. Cada uma delas, partindo de sua própria visão de mundo e amparadas em fundamentos teóricos e metodológicos mais ou menos distintos, ajuda a moldar o pensar e o fazer ontológico.

Uschold e Gruninger (1996) afirmam que uma ontologia se compromete com uma visão compartilhada de mundo sobre um determinado domínio. Já Guarino (1998), propõe uma definição para ontologia quando diz que:

Uma ontologia é uma teoria lógica que explica o significado pretendido de um vocabulário formal, ou seja, seu compromisso ontológico com uma conceituação particular do mundo. Os modelos pretendidos de uma linguagem lógica usando tal vocabulário são restritos pelo seu compromisso ontológico. Uma ontologia reflete indiretamente esse compromisso (e a conceituação subjacente) aproximando esses modelos pretendidos. (GUARINO, 1998, p. 7, tradução nossa).

Como se observa, o autor propõe a ontologia como uma teoria que reflete o significado dos objetos no mundo. Almeida (2014) identifica dois principais sentidos para o termo neste campo de pesquisa, são eles: a) a ontologia como teoria representativa que descreve fatos e regras que governam parte da realidade; e b) a ontologia como declarações expressas em uma linguagem formal de representação. No primeiro sentido, a função da ontologia é formalizar o conhecimento de um grupo de especialistas com fins computacionais. No segundo sentido, a função da ontologia na modelagem “é tornar explícitos axiomas que restringem modelos, de forma a igualar, tanto quanto possível, os modelos que contém o significado pretendido” (ALMEIDA, 2014, p. 249), permitindo inferências automatizadas.

As autoras Bräscher e Carlan (2010) postulam que as ontologias “definem conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual” (BRÄSCHER; CARLAN, 2010, p. 160). Já Campos (2004), refletindo sobre os princípios subjacentes ao processo de modelização de domínios do conhecimento, compreende a ontologia formal, como um formalismo que sistematiza conhecimento pretendendo a formalização de definições axiomáticas.

Como tratado, a definição de ontologia varia em decorrência da tradição disciplinar em que é operacionalizado. Nesta pesquisa, utilizou-se a definição postulada por Gruber (1993), neste sentido, a ontologia é a especificação formal e explícita de uma conceitualização compartilhada em determinado domínio do conhecimento (GRUBER, 1993).

Enquanto ferramenta apta a representar domínios de conhecimento computacionalmente, as ontologias adquirem importância crescente na Web semântica. Para que os dados possam ser consumidos independentemente da estrutura original, as ontologias precisam atuar. Considerando o papel das ontologias na Web semântica, estas também poderão sustentar o desenvolvimento de aplicativos cívicos baseados em conhecimento.

A *Internet* e a *Web* têm inspirado pesquisadores de várias matizes de pensamento a discorrer sobre o papel que estas tecnologias assumem no processo democrático. Há uma expectativa delas poderem contribuir para o aperfeiçoamento das instituições e das práticas democráticas. Nesta pesquisa importa discutir o significado que se atribui ao conceito de aplicações cívicas e revelar sua ligação com outros conceitos supracitados. Segundo Jäske e Ertiö (2019, p. 21) aplicações cívicas podem ser definidas como “tecnologias ascendentes que usam dados abertos para resolver desafios de governança e políticas”.

A adoção de padrões de publicação e consumo de dados abertos na Web permitiu ao homem, à máquina ou a uma combinação entre os dois explorar dados que antes estavam ocultos em documentos. Segundo O'Reilly (2011), “a magia dos dados abertos é que a mesma abertura que permite transparência também permite inovação” (O'REILLY, 2011, p. 26, tradução nossa). Na visão do autor, este movimento se dá a partir da reutilização

criativa dos dados abertos governamentais (DAG) no desenvolvimento de aplicativos.

Assim sendo, os panoramas possíveis para a utilização de dados abertos governamentais são ilimitados, amparando ações empreendedoras, governamentais e cívicas. Estes podem viabilizar a criação de novos modelos de negócios baseados em plataformas e aplicativos; o aperfeiçoamento e desenvolvimento de serviços públicos; e o incremento de ações de participação, deliberação e controle sobre o Estado e outras organizações.

Os DAG habilitam formas de participação direta do cidadão (*crowdsourcing*) na formulação e execução de serviços públicos potencialmente mais efetivos, além de fornecer bases para economia de custos no governo. Por meio de APIs de DAG, desenvolvedores de software podem usar dados governamentais para escrever códigos que criam interfaces novas e profícuas com o governo e/ou com o mercado. Algumas destas aplicações podem auxiliar ou mesmo substituir funções de governo. Um aplicativo, por exemplo, que possibilita aos indivíduos de uma cidade relatar buracos na via pública, iluminações e sinalizações deficientes, entre outros problemas, pode auxiliar um órgão público (eventualmente sobrecarregado) a oferecer melhores serviços públicos (O'REILLY, 2011).

É crescente a adoção pelo poder público dos padrões W3C para a *Web* semântica, permitindo o reuso e integração de DAG (LEE; ALMIRALL; WAREHAM, 2015). Este movimento promoverá a padronização de publicação de DAG entre os poderes públicos constituídos e alavancará os efeitos de rede que permitirão, entre outros feitos, o desenvolvimento de plataformas comuns entre cidades. Estas mudanças, caso se concretizem, podem conduzir a gestão pública a uma visão de “*government as a platform*”<sup>2</sup> (O'REILLY, 2011).

## **ONTOACCOUNT: METODOLOGIA E ESPECIFICAÇÃO**

Para a construção de uma ontologia existem diversas metodologias de referência. A metodologia de construção adotada nesta pesquisa é a *OntoForInfoScience*, proposta por Mendonça (2015). Ela compreende uma pré-etapa e oito etapas de desenvolvimento de ontologias. São elas: 1) especificação; 2) aquisição e extração de conhecimento; 3)

---

<sup>2</sup> Uma abordagem para a transformação digital do setor público apresentado por O'Reilly, (2011).

conceitualização; 4) fundamentação ontológica; 5) formalização; 6) avaliação; 7) documentação e 8) disponibilização.

A escolha desta metodologia se deu em função de quatro qualidades atribuídas a ela: 1) possui um melhor detalhamento das etapas de construção de ontologias; 2) foi concebida para apoiar especialistas em organização e representação do conhecimento; e 3) tem como origem etapas de metodologias já conhecidas (*Methontology*, *NeOn* e o *101 Method*); e 4) foi desenvolvida no âmbito de um Programa de Pós-Graduação, o que confere uma garantia razoável de que a sua proposição obedeceu ao rigor científico.

A *OntoAccount* foi construída mediante a utilização do *framework Protégé 5.5.0*, um editor de ontologia de código aberto e gratuito, que suporta as especificações OWL 2 e RDF do W3C. A seguir, descrevem-se os procedimentos sucessivos adotados em cada uma das etapas da metodologia e, quando disponível, uma prévia dos resultados.

#### a) Pré-etapa: avaliação da necessidade

Nesta fase foi realizada uma descrição da real necessidade de construção da ontologia ao contrário de outros instrumentos de organização do conhecimento (tesauros, taxonomia etc.).

#### b) Etapa 1: especificação

Nesta etapa foram destacadas algumas dimensões básicas que demarcam o propósito da construção da ontologia de domínio. Seguindo orientações da metodologia *OntoForInfoScience*, foram especificados por meio do *template* de especificação os seguintes pontos: a) o domínio e escopo da ontologia; b) os prováveis cenários de aplicação da ontologia; c) o público-alvo da ontologia; e o d) o grau de formalidade da ontologia.

Além destes elementos que compõe a metodologia, foram acrescentados outros metadados sobre a arquitetura da ontologia e da sua interface *Web*. Entre eles: a) o *namespace* e a *URI* da ontologia; b) a página da *Internet* do projeto; e c) o idioma padrão.

#### c) Etapa 2: aquisição e extração de conhecimento

Os documentos de referência utilizados no desenvolvimento da *OntoAccount* compreendem normas jurídicas e documentos públicos, em especial relatórios. Resoluções, instruções normativas, publicações e acórdãos emitidos pelos colegiados do Tribunal de Contas da União (TCU) em

decorrência da tomada de contas ordinária, assim como relatórios de prestação de contas anual elaborados por universidades federais foram coletadas e analisadas.

Os materiais de referência foram analisados informal e formalmente com o objetivo de identificar as estruturas textuais do domínio e as definições primárias. Uma amostra<sup>3</sup> de relatórios de gestão foi submetida ao processo de extração terminológica automática dos termos mais frequentes por meio da funcionalidade "Wordlist" da ferramenta *Sketch Engine*<sup>4</sup>.

As listas de frequência dos termos foram analisadas em conjunto com a lista de termos provenientes da extração manual dos demais documentos de referência. Os termos resultantes foram organizados em três conjuntos (glossário de conceitos, glossário de verbos e glossário de relações) que serviram de base para a formação de classes, propriedades e relações na etapa de conceitualização.

#### d) Etapa 3: conceitualização

Nesta etapa foram realizadas as atividades de identificação e de análise dos conceitos do domínio a fim de subsidiar a construção de modelos conceituais. Tendo em vista a vivência profissional do autor e a existência de fontes de autoridade legal e formal,<sup>5</sup> que dão suporte a tarefa de definição textual de conceitos, optou-se por não envolver outros especialistas do domínio nesta etapa. Os produtos deste estágio compreendem: a) a tabela de conceitos e propriedades; b) o dicionário de verbos; e c) os modelos conceituais.

A tabela de conceitos e propriedades foi elaborada a partir da análise semântica do glossário de conceitos. A cada fase do processo de conceitualização, foram acrescentados novos elementos à estrutura original do glossário, gerando os seguintes subprodutos: a) o dicionário de conceitos; e b) a tabela de conceito e valores.

---

<sup>3</sup> Foram selecionados de forma intencional os relatórios de gestão de 2018 das seguintes instituições: Universidade de Brasília; Universidade Federal do Rio de Janeiro e Universidade Federal de Minas Gerais.

<sup>4</sup> *Sketch Engine* é uma ferramenta online que analisa textos e identifica o que é típico, raro ou de uso emergente na linguagem. Disponível em: <https://www.sketchengine.eu/>. Acesso em: 07 nov 2021.

<sup>5</sup> Além da legislação específica, foi consultada a ferramenta Vocabulário de Controle Externo (VCE), um tesouro que padroniza a terminologia utilizada na atividade de Controle Externo e auxilia o tratamento e a recuperação das informações no TCU. Disponível em: <https://contas.tcu.gov.br/ords/f?p=701663:1:3919793213294>. Acesso em: 08 Nov 2021.

O dicionário de conceitos foi criado a partir da associação e da inclusão da definição textual dos conceitos selecionados a partir do glossário. Nesta fase, também foram identificados os sinônimos. Para fins de definição textual dos conceitos e de associação de sinônimos aos termos preferenciais, foram considerados os conceitos constantes do Anexo I da Instrução Normativa-TCU 84, de 22 de abril de 2020 e, como fonte secundária, a terminologia do tesouro do TCU, operado por meio da ferramenta Vocabulário de Controle Externo (VCE). No caso de conceitos ausentes nos documentos referenciais, procurou-se a definição na legislação aplicável à Administração Pública.

Já a tabela de conceito e valores foi elaborada a partir do relacionamento entre os conceitos do dicionário e seus respectivos "valores possíveis". (MENDONÇA, 2015, p. 201). Por fim, foram identificadas algumas possíveis propriedades dos conceitos ou classes. A tabela de conceitos e propriedades corresponde à tabela de conceitos e valores acrescida das propriedades identificadas nesta fase.

O dicionário de verbos também foi elaborado a partir de um produto da etapa anterior, o glossário de verbos. Esse foi elaborado acrescentando, a cada termo do glossário, uma definição do significado do verbo no domínio. Os verbos selecionados foram representados uma única vez no dicionário de verbos e associados com seus sinônimos. Por fim, foram acrescentados exemplos de uso do verbo no domínio. O dicionário de verbos, assim como ocorre com a tabela de conceitos e propriedades, contém um conjunto de termos candidatos às relações conceituais.

Os modelos conceituais correspondem ao terceiro produto da etapa de conceitualização e foram gerados a partir da ferramenta Miro<sup>6</sup>. Esta ferramenta apoiou a construção de mapas conceituais, taxonomias e outras estruturas gráficas que representam o domínio.

#### e) Etapa 4: fundamentação ontológica

Adotou-se os princípios de formalização e as entidades e relações do *Basic Formal Ontology (BFO)*, versão 2, em português<sup>7</sup>, como ontologia de alto

---

<sup>6</sup> É uma plataforma online colaborativa que auxilia na construção de mapas mentais, conceituais, diagramas e outras representações gráficas.

<sup>7</sup> Utilizou-se a versão traduzida da BFO 2.0, denominada BFO-PT, que foi elaborada por Simone Torres de Souza durante a pesquisa de doutorado que resultou na ontologia denominada

nível para fundamentar o modelo da OntoAccount. Justifica-se a opção pela BFO e pelo paradigma do realismo científico pelas características deste domínio, que para sua descrição verdadeira necessita representar fenômenos e objetos reais. Sendo assim, a OntoAccount surge como uma ontologia que busca descrições verdadeiras (ou aproximadamente verdadeiras) das entidades do domínio da tomada e prestação de contas.

A ontologia de fundamentação foi importada para o *Protégé* 5.5.0 para início da formalização. Como referência para a implementação do BFO, utilizou-se o *Basic Formal Ontology 2.0: Specification and User's Guide* (SMITH et al., 2015).

#### f) Etapa 5: formalização

Nesta fase as estruturas conceituais do domínio, tratadas anteriormente no nível textual, passaram por uma série de restrições ontológicas. A metodologia *OntoForInfoScience* considera três passos no processo de construção da taxonomia da ontologia em desenvolvimento, são eles:

- Passo 1: Classificar cada conceito do conjunto de modelos conceituais produzidos na etapa de conceitualização. No desenvolvimento da OntoAccount, os conceitos contidos nos artefatos gráficos e textuais foram organizados a fim de contrastá-los com as classes equivalentes existentes em repositórios de ontologia.
- Passo 2: Promover o reuso de classe de outras ontologias que represente o significado do conceito no domínio. Quando nenhuma classe é encontrada, deve-se criar uma classe na ontologia em desenvolvimento. Na construção da OntoAccount, o autor realizou buscas de recursos ontológicos por meio da ferramenta *Web OntoBee*<sup>8</sup>. Além dos termos da ontologia de fundamentação (BFO), foram importadas classes e axiomas associados de outras ontologias de nível médio. Na construção da OntoAccount, utilizou-se classes originalmente definidas na *Information Artifact Ontology (IAO)*, na *NCBI organismal classification (NCBITaxon)*, na *Obstetric and Neonatal Ontology*

---

Ontolegis, aplicada à informação legislativa para o Direito Médico. Disponível em: <http://mba.eci.ufmg.br/downloads/BFO-PT/BFO-PT.owl>. Acesso em: 28 nov 2021.

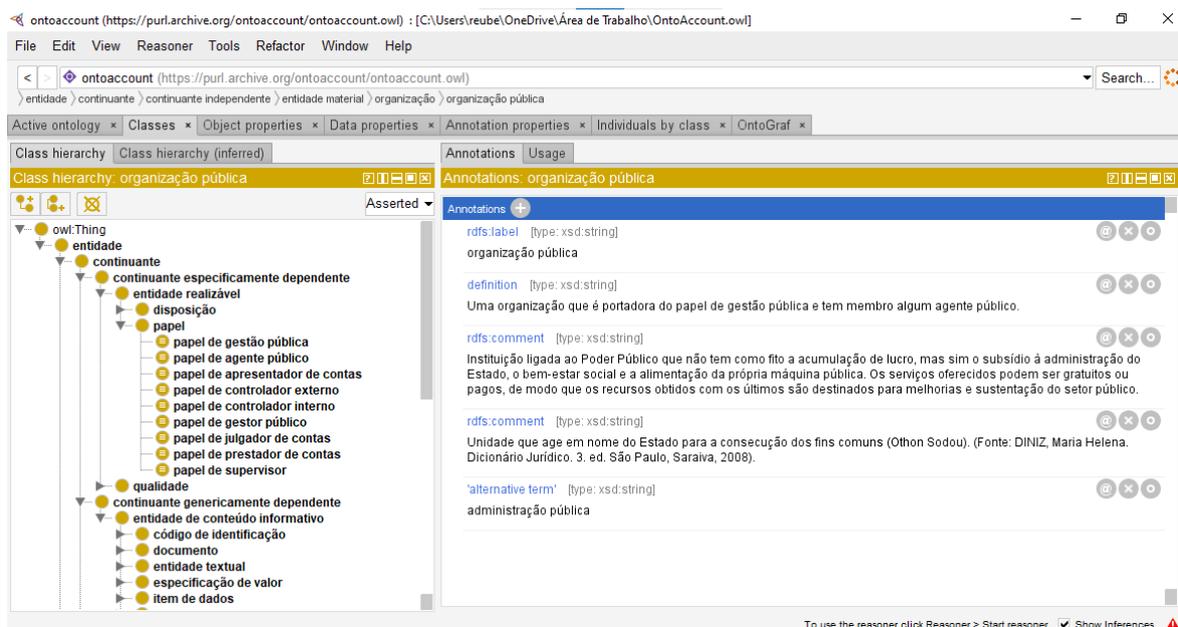
<sup>8</sup> Um servidor de dados conectados projetado para promover o reuso de ontologias. Disponível em: <http://www.ontobee.org/>. Acesso em: 18 Nov. 2021.

(*ONTONEO*), na *Ontology for Biomedical Investigations (OBI)* e na *Ontology of Medically Related Social Entities (OMRSE)*.

- Passo 3 – Uma vez criada ou identificada a classe equivalente, a metodologia *OntoForInfoScience* determina a inclusão desta classe na taxonomia da ontologia em construção. Na construção da *OntoAccount*, cada classe específica do domínio foi classificada como um tipo de classe *BFO* e inserida como uma subclasse da classe superior.

Na atividade de definição das propriedades descritivas (textuais) das classes, a metodologia *OntoForInfoScience* motiva o uso da Tabela de Conceitos e Propriedades como base desta tarefa, contudo alerta para a necessidade de se adequar as propriedades ao nível ontológico, tendo em vista se tratar de definição livre e textual (MENDONÇA, 2015). Nesse passo de construção da *OntoAccount*, foram inseridas as propriedades textuais por meio da funcionalidade “*Annotations*” do *Protégé 5.5.0*. A Figura 1 exemplifica as propriedades descritivas para a classe “organização pública”.

**Figura 1** - propriedades descritivas da classe “organização pública” no *protégé 5.5.0*.



**Fonte:** Elaborado pelo autor (2022).

A classe “organização pública” recebeu atributos ontológicos associados aos seguintes atributos do *Protégé*: nome de classe (*label*), definição (*definition*), comentários (*comments*) e termo alternativo (*alternative*

term). O preenchimento dos atributos foi realizado no formato *string* e manteve-se os acentos gráficos, inclusive no nome de classe<sup>9</sup>.

Na atividade de definição formal das classes da ontologia a *OntoForInfoScience* prescreve uma série de restrições ontológicas que objetivam viabilizar processos de inferência à base de conhecimento da ontologia. Estes procedimentos foram adotados nas classes da *OntoAccount* e um dos resultados podem ser observados no Quadro 2, a seguir.

**Quadro 2** – exemplo da definição formal da classe “organização” da *Ontoaccount*

Termo (Organização Pública)	
<b>Definição textual</b>	Uma organização que é portadora do papel de gestão pública e tem membro algum agente público.
<b>Anotação</b>	Uma <u>organização</u> [OBI/ONTOACCOUNT: organização] que é <u>portadora do papel</u> de gestão pública [RO/ONTOACCOUNT: tem papel], e <u>tem membro</u> algum agente público. [BFO/ONTOACCOUNT: tem membro].
<b>Definição formal (OWL-DL)</b>	<i>Organização pública</i> é um <i>organização</i> E ( <b>tem papel</b> SOME <i>papel de gestão pública</i> ) E ( <b>tem membro</b> SOME <i>agente público</i> )

Fonte: Elaborado pelo autor (2022).

No que tange à definição de propriedades de dados das classes, algumas restrições foram aplicadas. Uma amostra do resultado é apresentada a seguir (Quadro 3).

**Quadro 3** – parte das propriedades de dados da *ontoaccount*

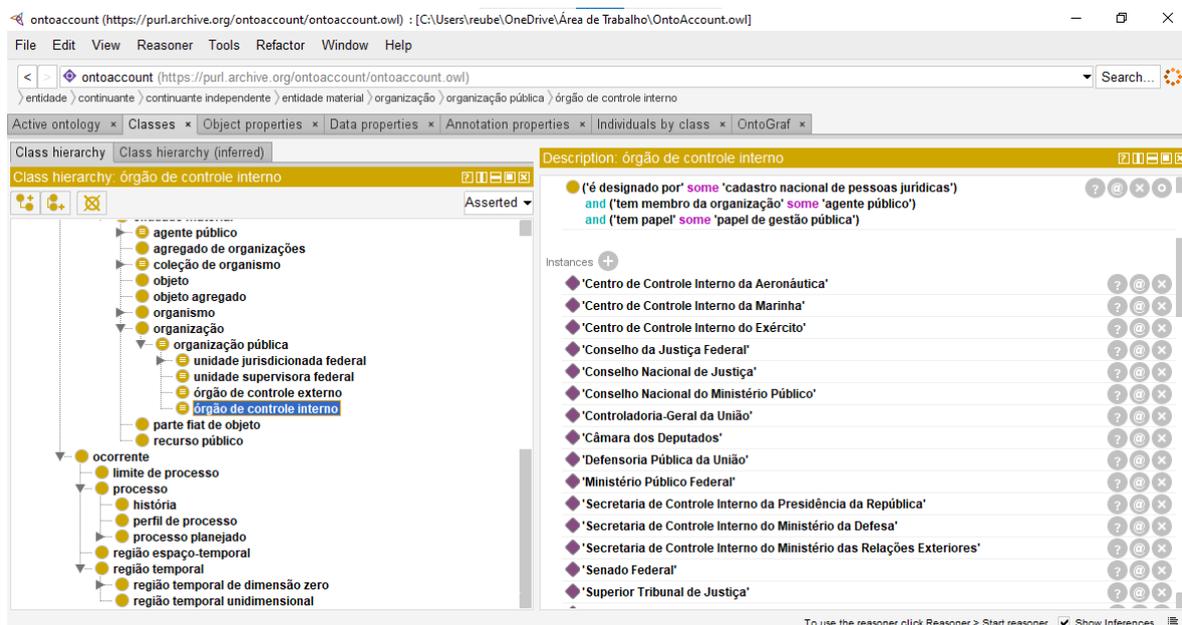
ID	Propriedades de dados	Valor
147	'tem ano de atuação'	Value = integer
134	'tem cargo'	Value = string
164	'tem data sessão'	Value = dateTime
403	'tem dirigente UCI'	Value = string
118	'tem entidade'	Value = string
131	'tem início de exercício'	Value = dateTime
115	'tem nome próprio'	Value = string
149	'tem relator'	Value = string
396	'tem tipo'	Value = string

Fonte: Elaborado pelo autor (2022).

<sup>9</sup> A *OntoForInfoScience* orienta a utilizar no nome de classe (*label*) apenas letras minúsculas e sem acentos gráficos. Contudo, não se verificou incompatibilidade na manutenção de acentos gráficos em nomes de classes originados da língua portuguesa.

A criação de instâncias das classes é mais um procedimento previsto na etapa de formalização da ontologia, a OntoAccount foi instanciada com dados abertos estruturados<sup>10</sup>, que representam objetos específicos do domínio da tomada e prestação de contas. Entre eles, destacam-se, como exemplo, os particulares da classe 'órgão de controle interno' visualizados na Figura 2.

**Figura 2** – exemplos de instâncias da classe 'órgão de controle interno' da ontoaccount



Fonte: Elaborado pelo autor (2022).

No escopo da especificação de relações ontológicas, foram incorporadas na OntoAccount algumas relações ontológicas específicas do domínio da tomada e prestação de contas, conforme o Quadro 4, a seguir.

<sup>10</sup> Os dados inseridos na ontologia, para efeito de testes e avaliação, são dados não sensíveis, de origem pública e não requerem autorização prévia para exibição conforme artigo 7º, parágrafos 3º, 4º e 7º da Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais - LGPD), resguardados os direitos do titular e os princípios previstos nesta Lei. Como forma de proteção, os dados do Cadastro de Pessoas Físicas (CPF) foram inseridos num formato anonimizado. Por exemplo, o CPF 123.456.789-00, se tornou \*\*\*.456.789-\*\*. Já o nome próprio dos gestores públicos foi inserido no mesmo formato tornado público pelo próprio titular, em documentos de prestação de contas, ou pela Administração Pública, tendo em vista os princípios da publicidade e da transparência, além do direito de informação dos cidadãos.

**Quadro 4** – parte das relações específicas do domínio na Ontoaccount

Relação	Origem	Definição informal
<b>C julga C1</b>	OntoAccount	Relação ontológica identificada entre dois continuantes, C e C1, sendo que C é um órgão de controle externo que julga o processo de contas que corresponde a uma organização pública C1.
<b>C é julgado por C1</b>	OntoAccount	Relação ontológica (inversa) identificada entre dois continuantes, C e C1, sendo que C é uma organização pública que tem as contas julgadas por um órgão de controle externo C1.
<b>C é apresentadora de contas C1</b>	OntoAccount	Relação ontológica identificada entre dois continuantes, C e C1, sendo que C é uma unidade apresentadora de contas que apresenta a prestação de contas correspondente a uma unidade prestadora de contas C1.
<b>C tem apresentadora de contas C1</b>	OntoAccount	Relação ontológica (inversa) identificada entre dois continuantes, C e C1, sendo que C é uma unidade prestadora de contas que tem sua prestação de contas corresponde apresentada por uma unidade apresentadora de contas C1.
<b>C tem responsável C1</b>	OntoAccount	Relação ontológica identificada entre dois continuantes, C e C1, sendo que C é uma unidade prestadora de contas que tem como responsável um gestor público C1.
<b>C é responsável por C1</b>	OntoAccount	Relação ontológica (inversa) identificada entre dois continuantes, C e C1, sendo que C é um gestor público responsável pela prestação de contas corresponde a uma unidade prestadora de contas C1.

Fonte: Elaborado pelo autor (2022).

Por fim, as propriedades das relações ontológicas foram estabelecidas no *Protégé* 5.5.0. O conjunto completo de relações e propriedades da ontologia compõe a documentação da ontologia.

#### g) Etapa 6: avaliação

O sexto estágio envolve a avaliação da ontologia com os objetivos de validação (adequação ao domínio) como de verificação ontológica (adequação interna).

Na metodologia proposta, um conjunto de critérios avaliativos, consistente e suficiente, para avaliar o conteúdo da ontologia, tanto do ponto de vista da validação quanto da verificação ontológicas foi sugerido. As bases destes critérios remetem aos princípios ontológicos e às regras de avaliação de outras metodologias tradicionais de construção de ontologias (por exemplo, *NeOn*, *Methontology* e o 101).

O resultado da avaliação da ontologia *OntoAccount*, utilizando-se tais parâmetros e critérios, também compõe a documentação formal da ontologia.

#### h) Etapa 7: documentação

Nesta etapa foi concatenada a documentação formal da ontologia produzida ao longo de todo o processo. O documento formal inclui todos os produtos das etapas anteriores e é oferecido em linguagem natural. Ficaram de fora da documentação formal os artefatos produzidos com representações preliminares, isto é, aqueles que sofreram transformações ao longo do processo, tais como: a) o pré-glossário de termos; b) os conjuntos de candidatos à ontologia; e c) a tabela de conceitos e valores.

#### i) Etapa 8: disponibilização

A última versão da OntoAccount foi disponibilizada na Web em linguagem de lógica descritiva, no formato OWL/XML, para que o seu conteúdo ontológico possa ser interpretado por máquinas de inferências e para que seus recursos possam ser consumidos na Web. A escolha por este formato se deu pela familiaridade do autor com o editor *Protégé* 5.5.0, sendo este o formato padrão do arquivo criado nesta ferramenta. Essa escolha também foi motivada pela aparente popularidade deste formato em ontologias de domínio. Por meio do *Protégé* 5.5.0, foi registrado um identificador de recurso à ontologia, um *Internationalized Resource Identifier* (IRI)<sup>11</sup>, por meio do qual os usuários podem fazer referências ao conteúdo da ontologia.

A OntoAccount também recebeu um endereço Web<sup>12</sup> e uma homepage com interface de busca ao conteúdo ontológico, que permite conhecer seus elementos (classes, propriedades e indivíduos) a fim de permitir a reutilização de seus termos em outros projetos. O documento de especificação, inserido no Quadro 5, a seguir, reúne as dimensões básicas da arquitetura e do propósito da OntoAccount.

**Quadro 5** – especificação da ontologia Ontoaccount

1	Namespace	<a href="https://purl.archive.org/ontoaccount/">https://purl.archive.org/ontoaccount/</a>
2	IRI	<a href="https://purl.archive.org/ontoaccount/ontoaccount.owl">https://purl.archive.org/ontoaccount/ontoaccount.owl</a>
3	Identificador local/ <i>Local Identifier</i>	<a href="https://purl.archive.org/ontoaccount/&lt;PREFIX&gt;_&lt;99999999&gt;">https://purl.archive.org/ontoaccount/&lt;PREFIX&gt;_&lt;99999999&gt;</a>
4	PREFIX or IDSPACE	ONTOACCOUNT
5	Website	<a href="https://ontoaccount.org/">https://ontoaccount.org/</a>
6	Idioma padrão/ <i>Default language</i>	Português

<sup>11</sup> Disponível em: <https://purl.archive.org/purl/ontoaccount/ontoaccount.owl>. Acesso em: 12 dez 2021.

<sup>12</sup> Disponível em: <https://ontoaccount.org/>. Acesso em: 12 dez 2021.

7	Idioma alternativo/ <i>Alternative language</i>	<i>English</i>
8	Licença / License	<i>Creative Commons CC-BY license version 4.0</i>
9	Ontologia de alto nível base / <i>Top-level ontology</i>	<i>Basic Formal Ontology version 2.0</i>
10	Domínio e Escopo geral/ <i>Domain and General Scope</i>	A <i>Ontology of Accountability (OntoAccount)</i> é uma ontologia de domínio que representa o conhecimento relativo à tomada e à prestação de contas anual dos gestores públicos federais. Seu escopo geral abrange os elementos constituintes do processo tomada e de prestação de contas anual dos gestores públicos federais.
11	Propósito Geral/ <i>General Purpose</i>	A <i>OntoAccount</i> tem como propósito oferecer suporte ao desenvolvimento de aplicações cívicas sobre o domínio da tomada e da prestação de contas anual dos gestores públicos federais. Esta ação poderá auxiliar no aumento da transparência e na maior participação do cidadão na atividade de controle social.
12	Classes de usuários/ <i>User classes</i>	A <i>OntoAccount</i> destina-se a desenvolvedores de aplicações cívicas e profissionais que trabalham em processos de abertura de dados, transparência ativa e controle da gestão pública.
13	Uso pretendido/ <i>Intended use</i>	O uso geral pretendido com a <i>OntoAccount</i> é no suporte à modelagem de conhecimento sobre o domínio da tomada e da prestação de contas anual dos gestores públicos federais.
14	Tipo da ontologia/ <i>Ontology type</i>	A <i>OntoAccount</i> é classificada como ontologia de domínio
15	Grau de formalidade/ <i>Degree of formality</i>	A <i>OntoAccount</i> está representada na linguagem OWL-DL, considerada de médio rigor formal.

Fonte: elaborado pelo autor (2022).

## CONSIDERAÇÕES FINAIS

Este estudo se baseia na premissa de que a ausência de modelagem semântica formal e explícita do domínio da tomada e da prestação de contas anual dos gestores públicos federais prejudica o consumo dos dados públicos por agentes que queiram assimilar o valor público gerado. Esta ausência gera barreiras ao avanço da participação e do controle social.

Deste modo, esta pesquisa buscou contribuir para a publicação de dados abertos estruturados e semânticos deste domínio por meio de uma solução ontológica, nomeada *OntoAccount*, que atribuiu sentido e significado ao processo e ao conteúdo dos relatórios da amostra. Esta iniciativa oportuniza o reuso de recursos e a troca de informação e conhecimento entre pessoas e aplicações.

Como se verificou na metodologia apresentada, a ontologia foi desenvolvida em etapas estruturadas que se somam e se apoiam para garantir a recuperação e a reutilização de conhecimento do domínio. Recomenda-se para trabalhos futuros a exploração de soluções de extração

automática de dados de documentos do domínio para contribuir para a publicação de dados abertos estruturados que podem instanciar a estrutura básica da ontologia de modo a permitir seu uso como base de conhecimento.

## REFERÊNCIAS

ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 242-258, 2014. Disponível em: <https://brapci.inf.br/index.php/res/v/35762>. Acesso em: 03 nov. 2020.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic Web. **Scientific american**, [s.l.], v. 284, n. 5, p. 34-43, 2001. Disponível em: <https://www.jstor.org/stable/26059207>. Acesso em: 04 set. 2018.

BERNERS-LEE, T. Linked Data. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 04 set. 2018.

BRÄSCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. **Passeios no bosque da informação: Estudos sobre Representação e Organização da Informação e do Conhecimento**. Brasília (DF): IBICT, 2010. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso: 30 ago. 2019.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília (DF), v. 33, n. 1, 2004. Disponível em: <http://revista.ibict.br/ciinf/article/view/1064>. Acesso: 11 ago. 2020.

GRUBER, T. R. A translational approach to portable ontologies. **Knowledge Acquisition**, [s.l.], v. 5, n. 2, p. 199-229, 1993. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1042814383710083?np=y>. Acesso: 13 ago. 2020

GUARINO, N. Formal Ontology and Information Systems. In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS (FOIS98), 1., 1998, Trento. **Proceedings** [...]. Trento: [s.n.], 1998.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**: em busca da Web do Conhecimento. [s.l.]: Novatec Editora, 2015. Disponível em: <https://ceWeb.br/livros/dados-abertos-conectados/>. Acesso em: 09 jul. 2018.

JÄSKE, M.; ERTIÖ, T. The democratic potential of civic applications. **Information Polity**, [s.l.], v. 24, n. 1, p. 21-39, 2019. Disponível em: <https://content.iospress.com/articles/information-polity/ip180105>. Acesso em: 18 jul. 2020.

KIM, J. G.; HAUSENBLAS, M. **5 Star Open Data**. 2012. Disponível em: <https://5stardata.info/en/>. Acesso em: 05 maio 2019.

LEE, M.; ALMIRALL, E.; WAREHAM, J. Open data and civic apps: first-generation failures, second-generation improvements. **Communications of the ACM**, [s.l.], v. 59, n. 1, p. 82-89, 2015. Disponível em: <https://dl.acm.org/doi/fullHtml/10.1145/2756542>. Acesso em: 23 nov. 2020.

MENDONÇA, F. M. *OntoForInfoScience*: metodologia para construção de ontologias pelos cientistas da informação-Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). 2015.

MENDONÇA, F. M.; SOARES, A. L. Construindo ontologias com a metodologia *OntoForInfoScience*: uma abordagem detalhada das atividades do desenvolvimento ontológico. **Ciência da Informação**, Brasília (DF), v. 46, n. 1, 2017.

OPEN GOVERNMENT PARTNERSHIP [OGP]. Declaração de governo aberto. set. 2011. Disponível em: [www.opengovpartnership.org/open-government-declaration](http://www.opengovpartnership.org/open-government-declaration). Acesso em: 21 jul. 2020.

OPEN KNOWLEDGE FOUNDATION [OKF]. What is open?. 2020. Disponível em: <https://okfn.org/opendata/>. Acesso em: 21 jul. 2020.

O'REILLY, T. Government as a Platform. **Innovations: Technology, Governance, Globalization**, v. 6, n. 1, p. 13-40, 2011. Disponível em: [https://www.mitpressjournals.org/doi/abs/10.1162/INOV\\_a\\_00056](https://www.mitpressjournals.org/doi/abs/10.1162/INOV_a_00056). Acesso em: 23 ago. 2020.

SMITH, B. *et al.* **Basic formal ontology 2.0**: specification and users guide. [s.l.]: [s.n.] 2015.

USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. **Technical Report-University of Edinburgh Artificial Intelligence Applications Institute AIAI TR**, 1996. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.5917&rep=rep1&type=pdf>. Acesso em: 23 jun. 2020.