



ARQUITETURA DE PUBLICAÇÃO DE DADOS ABERTOS CONECTADOS GOVERNAMENTAIS DA UNIVERSIDADE DE BRASÍLIA

Publishing Architecture of Governmental Linked Open Data of The University of Brasília


Marcio de Carvalho VICTORINO

Doutor em Ciência da Informação
Universidade de Brasília, Professor Adjunto da Faculdade de Ciência da Informação, Brasília, Brasil
mcvictorino@unb.br
<https://orcid.org/0000-0003-2785-8958> 


Maristela HOLANDA


Doutora em Ciência da Computação
Universidade de Brasília, Professora Associada do Departamento de Ciência da Computação, Brasília, Brasil
mholanda@unb.br
<https://orcid.org/0000-0002-0883-2579> 

Luiz MARTINS

Mestre em Ciência da Computação
Universidade de Brasília, Departamento de Ciência da Computação, Brasília, Brasil
luizmartins@unb.br
<https://orcid.org/0000-0003-1514-1730> 

Rodrigo FONSECA

Mestrando em Ciência da Computação
Universidade de Brasília, Departamento de Ciência da Computação, Brasília, Brasil
rodrigofonseca@unb.br
<https://orcid.org/0000-0003-4965-8519> 

A lista completa com informações dos autores está no final do artigo 

RESUMO

Objetivo: apresentar uma arquitetura baseada em sistemas de organização da informação e do conhecimento, para dar suporte à publicação de dados abertos conectados.

Método: A metodologia utilizada é de caráter qualitativo, sendo a pesquisa exploratória, visto que partiu dos estudos dos dados abertos conectados governamentais, de metadados, de ontologias e de *Resource Description Framework*, para estabelecer um elo entre tais conceitos.

Resultados: Formulou-se uma arquitetura de publicação que foi verificada por meio de um estudo de caso utilizando dados abertos oriundos dos sistemas de informação da Universidade de Brasília.

Conclusões: A arquitetura proposta foi desenvolvida utilizando metadados e ontologias para conexão e representação semântica dos dados

PALAVRAS-CHAVE: Dados Abertos. Dados Conectados. Metadado. Ontologia. Transparência Pública.

ABSTRACT

Objectives: this study aims to present an architecture based on information and knowledge organization systems that offer support to the publication of linked open data.

Methods: The methodology is characterized as qualitative, exploratory research, given that it stems from studies of linked open data government, metadata, ontologies, and the Resource Description Framework to establish a link between these concepts.

Results: As a result, a publication architecture was formulated, which was verified through a case study using open data originating from the information systems of the University of Brasilia.

Conclusions: The architecture proposed was developed using metadata and ontologies for the link, and the semantic representation of data.

KEYWORDS: Open Data. Linked Data. Metadata. Ontology. Public Transparency.

1 INTRODUÇÃO

O governo brasileiro tem investido, nos últimos anos, em meios para incrementar a transparência de suas ações. Esta iniciativa foi consolidada por meio da Lei de Acesso à Informação (LAI), que regulamentou o Art. 5º, inciso XXXIII da Constituição Federal que garante o acesso a qualquer informação que possua interesse público, desde que esta informação não seja imprescindível à segurança da sociedade e do Estado (BRASIL, 1988a).

De acordo com Santarém Segundo (2013), a LAI foi criada a partir de um movimento chamado Dados Abertos em que países como Inglaterra e Estados Unidos avançam desde 2009 em um modelo de gestão que propõe aumentar a visibilidade das informações governamentais. O principal objetivo dessa lei é incentivar a sociedade a participar ativamente da gestão, contribuindo para que ela tenha eficiência e transparência.

O Brasil, em 2011, aderiu à Parceria para Governo Aberto - *Open Government Partnership* - OGP, uma iniciativa internacional criada para assegurar que os governos promovam a transparência, a participação civil, o combate à corrupção e o uso de novas tecnologias para tornar a administração mais eficaz e aberta. Desde 2012, o governo brasileiro promove ações visando incentivar a abertura de seus dados para a sociedade.

O Decreto No 8.777, de 2 de maio de 2016 (BRASIL, 1988b), instituiu a Política de Dados Abertos no Poder Executivo Federal e estabeleceu que qualquer cidadão pode solicitar a abertura de bases de dados nos termos da LAI, desde que o dado não esteja regido pelas regras de sigilo da informação e, além disso, define quais as autoridades serão responsáveis por publicar e atualizar o Plano de Dados Abertos - PDA de cada órgão. O PDA é o documento que orienta a implementação e a promoção da abertura de dados em cada órgão ou entidade da administração pública federal, estabelecendo padrões mínimos de qualidade que visem facilitar a manipulação e o reuso dos dados.

Para auxiliar na publicidade dos dados, o Governo Federal criou o Portal Brasileiro dos Dados Abertos - PBDA, ferramenta que disponibiliza o livre acesso aos dados governamentais em formato bruto para que qualquer cidadão possa utilizá-los da maneira que lhe convier (PEREIRA; MARCONDES, 2014).

Já é possível encontrar uma quantidade considerável de dados disponibilizados no PBDA oriundos de diversos órgãos, entretanto, os dados publicados estão relacionados apenas por meio de marcadores (*tags*) o que é bastante superficial. Assim, há uma

grande dificuldade de realizar integrações, comparações ou combinações de conjunto de dados de origens diferentes. Um exemplo seria utilizar dados abertos sobre a ocupação profissional do Ministério do Trabalho, integrados a dados sobre programas sociais do Ministério de Desenvolvimento Social, que poderiam gerar um novo conhecimento sobre como as ações sociais do Estado podem influenciar o mercado de trabalho.

Faz-se necessária, então, a disponibilização de uma arquitetura de publicação que subsidie um processo de publicação de dados abertos conectados provido de descrição semântica dos dados para promover a interoperabilidade efetiva entre conjuntos de dados oriundos de órgãos distintos, e dessa forma, facilitar a análise conjunta desses dados.

Este trabalho tem por objetivo apresentar uma arquitetura de publicação, baseada em sistemas de organização da informação e do conhecimento, para dar suporte à publicação, desde a extração dos dados dos sistemas fontes até o consumo pelos usuários finais, de dados abertos conectados com qualidade.

A arquitetura proposta neste trabalho foi desenvolvida utilizando ontologias e metadados para conexão e descrição das fontes de dados. Esta arquitetura foi verificada por meio de um estudo de caso utilizando dados abertos oriundos dos sistemas de informação da Universidade de Brasília - UnB que tratam de dados acadêmicos.

2 DADOS ABERTOS CONECTADOS GOVERNAMENTAIS

Segundo a definição da *Open Knowledge Foundation* (TCU, 2015), “dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los, estando sujeito, no máximo, a exigência de creditar a sua autoria e compartilhar pela mesma licença”. Eaves (2009), especialista em políticas públicas e ativista dos dados abertos, considerado um dos maiores especialistas no assunto, propôs três leis que foram adotadas pelo *World Wide Web Consortium - W3C*:

- Se o dado não pode ser encontrado e indexado na *web*, ele não existe;
- Se não estiver aberto e em formato compreensível por máquina, ele não pode ser reaproveitado; e
- Se algum dispositivo legal não permitir sua reaplicação, ele não é útil.

Desde a criação da *web*, um enorme volume de dados tem sido publicado diariamente de forma desorganizada o que dificulta o acesso, entendimento e processamento desses dados. Nesse contexto, Berners-Lee, Hendler e Lassila (2001)

apresentaram o que seria a *web* semântica. Trata-se de uma evolução da própria *World Wide Web* - *www* - na qual os dados seriam publicados utilizando-se padrões estabelecidos pelo *World Wide Web Consortium* – *W3C*, dentre eles, *eXtensible Markup Language* - *XML*, *Resource Description Framework* - *RDF*, *SPARQL Protocol and RDF Query Language* - *SPARQL*, *Uniform Resource Locator* – *URL*, *Web Ontology Language* – *OWL*.

Dessa forma, os dados publicados estariam acessíveis e processáveis por máquinas e passíveis de processos de organização que podem facilitar a apresentação desses dados, a geração de novos dados, a ligação com outros grupos de dados e o aumento do conhecimento para o suporte à decisão.

Embora não haja um relacionamento direto entre os conceitos de "*open data*" e de "*semantic web*", quando se agrega recursos da *web* semântica à publicação de dados abertos, vislumbra-se uma melhoria significativa da organização e do acesso a esses dados. Para disponibilizar dados abertos numa estrutura semântica é necessário levar em conta a estrutura de camadas proposta por Berners-Lee, Hendler e Lassila (2001) para a *web* semântica. Nessa estrutura, destaca-se o *RDF*, o uso de metadados e principalmente o uso de ontologias.

Um dos principais objetivos do *RDF* é construir uma rede de informações distribuídas onde os nós estejam semanticamente ligados, formando um grande grafo global, com informações advindas de várias fontes diferentes ao redor do planeta.

Por outro lado, ontologia é o recurso utilizado para se construir uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando o processo de interpretação dos dados mais eficiente e facilitando a recuperação da informação pelas ferramentas computacionais.

Segundo Berners-Lee (2006), esse grafo global composto por nós semanticamente ligados e estruturado por meio do uso de tecnologias previstas nas camadas da *web* semântica, entre elas, *RDF* e ontologias, denomina-se "*linked data*" (dados conectados). Heath e Bizer (2011) afirmam que dados conectados é um conjunto de melhores práticas para publicação e ligação de dados estruturados na *web*, permitindo estabelecer *links* entre itens de diferentes fontes de dados para formar um único espaço de dados global. Segundo Berners-Lee (2006), *Linked Open Data* - *LOD* (Dados Abertos Conectados) são dados conectados disponibilizados sob licença aberta.

Berners-Lee (2006) afirma que a construção do dado aberto e do dado aberto conectado está baseada em quatro princípios:

- 1) Use URIs como nome para coisas;

- 2) Use HTTP URIs para que as pessoas possam procurar por esses nomes;
- 3) Quando alguém procurar por uma URI, forneça informações úteis, usando os padrões RDF e SPARQL;
- 4) Inclua *links* para outros URIs, para que as pessoas possam descobrir mais coisas.

Existe uma categoria de dados abertos oriundos de órgãos do governo, denominada "dados abertos governamentais". O W3C define dados abertos governamentais como a publicação e divulgação de informações públicas na *web*, compartilhadas em um formato aberto, logicamente compreensíveis, permitindo a reutilização de informações em aplicativos digitais desenvolvidos pela sociedade.

Com base na relevância da interoperabilidade de dados governamentais abertos, Berners-Lee (2006) propôs o sistema de cinco estrelas, que classifica o grau de abertura de dados - quanto mais aberto, maior será o número de estrelas e maior será a facilidade de enriquecer os dados. As cinco estrelas de dados abertos conectados são:

- Uma Estrela: disponível na internet (em qualquer formato; por exemplo, PDF), desde que com licença aberta, para que seja considerado dado aberto;
- Duas Estrelas: disponível na internet de modo estruturado (por exemplo, em planilhas MS-Excel);
- Três Estrelas: disponível na internet de modo estruturado e em formato não proprietário (em uma planilha OpenOffice.org ou *Comma Separated Values* – CSV em vez de MS-Excel);
- Quatro Estrelas: seguindo todas as regras anteriores, mas dentro dos padrões estabelecidos pelo W3C (*Resource Description Framework* - RDF e *SPARQL Protocol and RDF Query Language* - SPARQL), uso de *Uniform Resource Locator* – URL para a identificação de coisas e propriedades, de modo que todos possam direcionar para suas publicações; e
- Cinco Estrelas: todas as regras anteriores e mais a conexão de seus dados a outros dados, fornecendo um contexto.

Uma análise dos dados de órgãos da administração pública do governo brasileiro disponibilizados na internet mostrou que esses dados massivos podem ser classificados com o grau de abertura de três estrelas, pois estão disponíveis na Internet de forma estruturada, não proprietária, formato CSV, e qualquer pessoa ou entidade pode baixá-los. O principal objetivo do presente estudo é proporcionar subsídios para elevar o grau de abertura desses dados.

Na bibliografia atual, é possível encontrar várias iniciativas com o objetivo de publicar dados abertos com qualidade para atender às mais variadas demandas. Rautenberg, Burda e Souza (2018) propõem um workflow para compartilhamento de dados científicos primários como dados abertos conectados. Por outro lado, Marcondes (2012) verifica o uso de dados abertos conectados para que arquivos, bibliotecas e museus disponibilizem e tornem seus acervos na *web* interoperáveis.

Isotani e Bittencourt (2015) ressaltam que várias empresas já criaram aplicações baseadas em dados abertos depois da proposta de Berners-Lee (2006). O Google anunciou a utilização do formato de serialização recomendado pelo W3C para serialização de dados conectados baseada no JSON para o Gmail; a IBM anunciou que o Banco de Dados DB2 se tornaria um servidor de Dados Conectados; o Facebook expôs os Dados Conectados via Graph API; e a BBC usou Dados Conectados para gerar páginas *web* para alguns de seus produtos. Os autores também citam iniciativas governamentais como a do governo britânico que disponibilizou várias de suas fontes de dados em formato RDF.

No entanto, o projeto precursor nessa área foi o *The Linked Open Data (LOD) Project*. O objetivo do projeto, coordenado pela W3C, é estender a *web* com dados comuns, publicando vários conjuntos de dados abertos como grafos RDF e definir *links* RDF entre itens de dados de diferentes fontes de dados. Já foram publicados pelo projeto, até o momento, 570 conjuntos de dados conectados por 2909 conjuntos de links (LOD PROJECT, 2011).

Do exposto, pode-se observar que tanto no ambiente privado quanto no público o dado aberto conectado vem se destacando como um dos recursos mais eficazes para a interoperabilidade e localização da informação descrita semanticamente.

3 SISTEMAS DE ORGANIZAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO

Resumidamente, pode-se afirmar que o objetivo da organização da informação e do conhecimento é dar suporte ao fluxo do tratamento e da recuperação dos objetos informacionais estruturados, semiestruturados e não estruturados nas organizações.

De acordo com Hjørland (2017), as ontologias são tipos amplamente reconhecidos de sistemas de organização do conhecimento. Segundo o autor, o objetivo da

organização do conhecimento é descobrir ou interpretar algum tipo de configuração de ordem. Smiraglia (2017) afirma que "a Organização do Conhecimento é a ciência da ordem do conhecimento, baseada na unidade central do conceito".

Guarino (1998) define ontologias como um artefato de engenharia constituído de um vocabulário específico, utilizado para descrever uma determinada realidade e um conjunto de suposições explícitas, relacionadas ao significado intencional das palavras do vocabulário.

Por outro lado, metadado pode ser definido como dado ou informação sobre o dado. Normalmente, é utilizado para armazenar informações úteis à recuperação ou acesso à informação, devendo ser capaz de descrever ou servir de sumário para o conteúdo de determinada informação. O termo surgiu em 1995, por ocasião de um simpósio realizado em Dublin, Ohio, que deu origem à *Dublin Core Metadata Initiative* – DCMI (DUBLIN CORE METADATA INICIATIVE, 2012).

Ontologias e metadados podem ser consideradas ferramentas extremamente úteis para organizar o conhecimento. Neste estudo, que aborda a publicação de dados abertos governamentais, as ontologias e metadados são usadas para contextualizar e descrever os dados abertos e relacionar os vários conjuntos de dados disponibilizados, transformando-os em dados abertos conectados, proporcionando o incremento do nível de abertura dos dados de três para cinco estrelas, conforme definido por Berners-Lee (2006).

4 METODOLOGIA

A metodologia utilizada neste trabalho é de caráter qualitativo, sendo uma pesquisa exploratória, que partiu de estudos dos dados abertos conectados governamentais, de sistemas de organização da informação e do conhecimento e de Resource Description Framework, para estabelecer um elo entre tais conceitos.

Para o desenvolvimento dessa pesquisa, primeiramente recorreu-se à literatura sobre dados abertos conectados governamentais, para fornecer pressupostos teóricos das principais características, buscando traçar um panorama da sua relevância para o governo brasileiro.

Na sequência, buscaram-se subsídios em metadados, ontologias e Resource Description Framework para entender como esses recursos poderiam ser utilizados para

a descrição das fontes de dados durante o processo de publicação dos dados abertos conectados. Após esse entendimento, formulou-se uma arquitetura da informação composta pelos sistemas de organização da informação e do conhecimento estudados e RDF.

Em seguida, foi proposta uma arquitetura de software para dar suporte à arquitetura da informação formulada. A arquitetura de software foi materializada por meio da implementação de uma ferramenta denominada UnB *Government Linked Open Data* (UnBGOLD).

Por fim, arquitetura da informação e a ferramenta UnBGOLD foram verificadas por meio de um estudo de caso utilizando dados oriundos dos sistemas de informação operativos da Universidade de Brasília que tratam de dados acadêmicos.

5 MODELO LÓGICO DA ARQUITETURA DE PUBLICAÇÃO DE DADOS ABERTOS CONECTADOS

A arquitetura de publicação de dados abertos conectados proposta visa subsidiar o gerenciamento da publicação desses dados, proporcionando o incremento da qualidade dos dados abertos publicados. Como parâmetro de qualidade, utilizou-se a definição proposta por Berners-Lee (2006) a partir do índice cinco estrelas.

O incremento da qualidade é realizado através do enriquecimento semântico dos dados obtido por meio da adoção de metadados e ontologias que proporcionem a descrição e ofereçam semântica, possibilitando a conexão de vários conjuntos de dados distintos.

Essa arquitetura de publicação é composta por duas outras: Arquitetura da Informação (AI) e Arquitetura de Software.

Rosenfeld e Morville (2002) afirmam que a AI consiste na combinação da organização, dicionarização e esquemas de navegação em sistemas de informação.

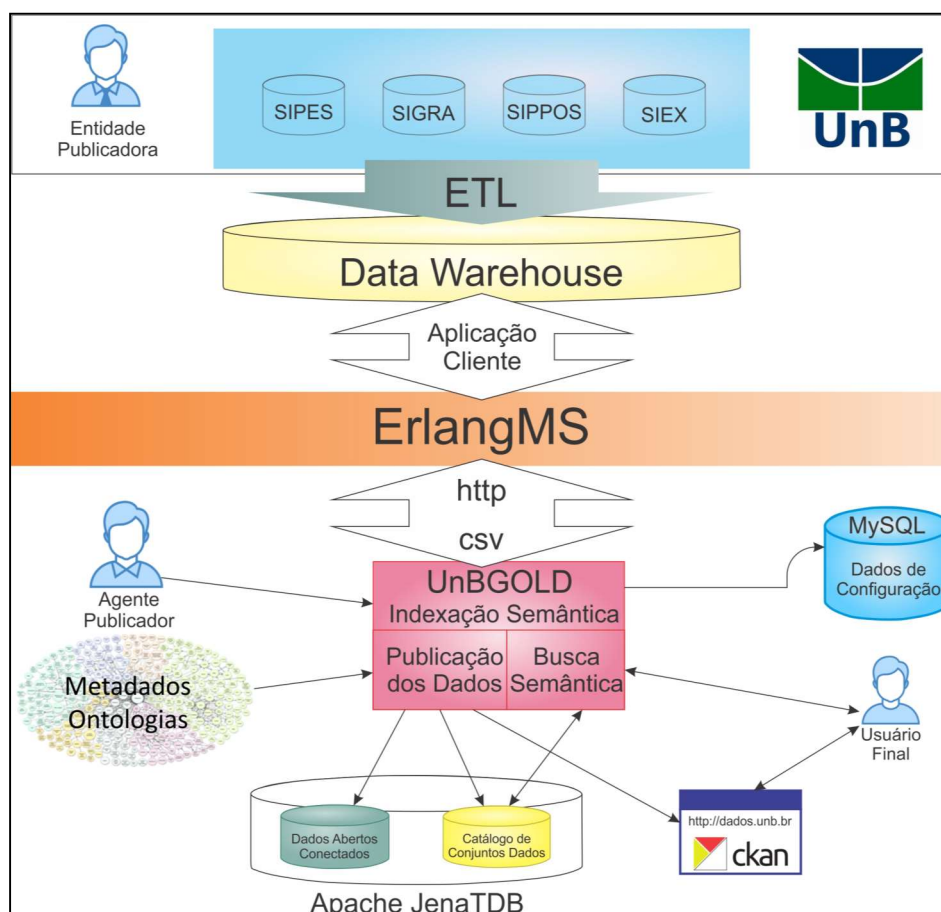
Segundo Bailey (2003), AI é a arte e a ciência de organizar os sistemas de informação para auxiliar o usuário a alcançar seus objetivos. Os arquitetos da informação organizam o conteúdo e projetam sistemas de navegação para auxiliar os usuários a encontrar e gerenciar informação.

Por outro lado, Arquitetura de Software refere-se à organização geral do software e aos modos pelos quais proporciona integridade ao sistema (SHAW; GARLAN, 1995). A

arquitetura de um sistema de software pode ser entendida como a sua organização ou estrutura de componentes significativos interagindo através de interfaces.

A Figura 1 apresenta o modelo lógico da arquitetura de publicação proposta. Essa arquitetura foi verificada por meio do uso de dados dos sistemas de informação operativos da UnB.

Figura 1 – Modelo lógico da arquitetura de publicação de dados abertos conectados governamentais.



Fonte: Elaborado pelos autores.

Observa-se na parte superior da Figura 1 que o processo de publicação se inicia com as Entidades Publicadoras que são instituições da administração pública que devem respeitar a legislação vigente e publicar os dados dos seus sistemas de informação em formato aberto para comunidade.

No caso da UnB, os dados são extraídos das bases de dados dos sistemas operativos que informatizam seus processos, como, por exemplo, o Sistema de Pessoal (SIPES), Sistema de Graduação (SIGRA), Sistema de Pós-Graduação (SIPPOS) e Sistema de Extensão e Pós-graduação (SIEX). Os dados são extraídos, transformados e

carregados em um *Data Warehouse* (DW) onde ficam disponíveis para o público interno da UNB realizar consultas integradas.

Depois de carregados no DW, os dados podem ser exportados em vários formatos por meio de requisição HTTP ao barramento de serviço ErlangMS (AGILAR, 2016) que consiste de um software livre desenvolvido na UnB que envia a solicitação para uma aplicação cliente que extrai os dados do DW de acordo com as regras de negócio estabelecidas pela Entidade Publicadora e retorna para o solicitante o conjunto de dados em formato já estruturado e aberto, neste caso optou-se pelo formato CSV.

Após a geração dos conjuntos de dados no formato CSV, o Agente Publicador preparará os dados para consumo dos usuários finais externos à UnB por meio da ferramenta livre UnB *Government Linked Open Data* (UnBGOLD) (MARTINS, 2018). A ferramenta UnBGOLD foi desenvolvida no contexto da pesquisa descrita neste trabalho para auxiliar o processo de publicação dos dados abertos conectados tornando-se a camada intermediária entre a extração dos dados nas suas fontes e a interface de consumo pelo usuário final, conforme apresentado na Figura 1.

Na verdade, a ferramenta UnBGOLD é utilizada pelo Agente Publicador para configurar os parâmetros de publicação, gerar metadados, definir as ontologias utilizadas para conectar os conjuntos de dados e gerar as triplas RDF disponibilizando os dados abertos conectados.

Essa ferramenta, depois de configurada, proporciona a automação das republicações dos dados, gerencia o Catálogo de Conjunto de Dados e oferece aos usuários finais uma interface para a busca semântica aos dados baseada na linguagem SPARQL.

O Catálogo de Conjunto de Dados é um banco de dados onde são armazenados os metadados que descrevem as características dos conjuntos de dados publicados visando facilitar a recuperação desses dados. A especificação DCMI foi selecionada inicialmente como conjunto de metadados utilizado para identificar as características dos conjuntos de dados, porém outros padrões de metadados podem ser utilizados conjuntamente.

Para realizar a descrição semântica, o Agente Publicador seleciona Metadados e Ontologias que possam descrever os dados. Neste processo os dados são descritos em formato de triplas RDF e são armazenados em um banco de dados Apache Jena TDB.

Também é gerado um banco de dados com os metadados selecionados para descrever as principais características dos conjuntos de dados, o Catálogo de Conjunto

de Dados. Por fim, configuram-se os parâmetros que proporcionarão que os dados sejam publicados automaticamente no portal de dados abertos da UnB que é uma instância da plataforma CKAN. Enquanto os dados conectados são armazenados em um banco de dados TDB, os dados de configuração são armazenados em um banco de dados relacional MySQL.

Após a publicação, o Usuário Final poderá consumir os dados abertos conectados diretamente no portal dos dados abertos da UnB, onde poderá navegar ou baixar os dados para usá-los como bem entender. Também é disponibilizada uma interface de busca no UnBGOLD que recupera informações, por meio de consultas SPARQL ao banco de dados das triplas RDF.

6 ESTUDO DE CASO

O UnBGOLD, conforme apresentado na Figura 1, é uma ferramenta desenvolvida para auxiliar a publicação de dados abertos. Essa ferramenta está localizada na camada intermediária entre a extração de dados, a partir de diversas fontes, e a interface de consumo utilizada pelo usuário final.

O processo de publicação dos dados pela ferramenta está organizado em quatro etapas sequenciais:

1. Publicação e Automatização;
2. Coleta de Informações Sobre os Dados;
3. Escolha do Vocabulário Controlado; e
4. Representação Semântica.

6.1 PUBLICAÇÃO E AUTOMAÇÃO

A primeira etapa é a Publicação e Automatização. Nessa etapa, o Agente Publicador preenche algumas informações sobre a publicação dos conjuntos de dados, conforme é apresentado na Figura 2.

Inicialmente o Agente Publicador pode optar por automatizar a publicação, caso contrário, poderá apenas realizar o download do arquivo RDF gerado pela ferramenta referente à representação semântica ao final do processo. Para configurar a

automatização, é necessário selecionar qual instância CKAN será utilizada e a frequência da publicação.

A lista de instâncias CKAN disponíveis é referente às Entidades Publicadoras que o Agente Publicador tem vinculação, essas instâncias devem ser previamente cadastradas. Já a frequência da publicação poderá ser diária, semanal, bimestral, semestral ou anual. O publicador também pode escolher o horário em que os dados serão publicados.

Figura 2 – Etapa Publicação e Automatização.

InBGOLD - UnB Governanmet Open Linked Data

Opções

- Buscas
- Conjunto de Dados
- Entidades Publicadoras
- Usuários/Agentes Publicadores
- Instâncias CKAN
- Vocabulários

1 Publicação e Automati... 2 Informações Sobre os ... 3 Vocabulário Contr... 4 Indexação Semã... 5 Concl...

Configuração de Publicação

Automatizar Publicação?

Instância CKAN
Dados Abertos da UnB

Frequência de Publicação
Semanal

Dia da Publicação
Domingo

Horário de Publicação
05:00

Formato de arquivo para publicação

RDF CSV JSON

Próximo >

Fonte: Elaborado pelos autores.

6.2 COLETA DE INFORMAÇÕES SOBRE OS DADOS

Na etapa Coleta de Informações sobre os Dados, são preenchidas as informações que descrevem os dados, isto é feito informando os metadados referentes ao conjunto de dados. A Figura 3 apresenta a interface da ferramenta.

A origem dos dados deve ser informada por meio do preenchimento do campo “Fonte de Dados”, é preciso definir a origem dos dados por meio de uma URL que proporcionará o download do arquivo que deverá estar disponível previamente. Também será possível informar parâmetros que serão utilizados na requisição HTTP. Eles poderão ser fixos, deixando explícito o valor que se deseja para o parâmetro, ou temporal, o qual o

Agente Publicador deve informar uma das opções de valor temporal (dia, mês, semana, mês, semestre, ano, etc.).

O UnBGOLD faz um teste de conexão acessando o conjunto de dados, caso funcione como o esperado, será apresentada uma mensagem de sucesso informando algumas características do conjunto de dados acessado.

Figura 3 – Etapa Informações Sobre os Dados.

The screenshot shows a web interface with a progress bar at the top containing five steps: 1. Publicação e Automa... (selected), 2. Informações Sobre os ... (active), 3. Vocabulário Contr..., 4. Indexação Semã..., and 5. Concl... Below the progress bar, the title 'Extração dos Dados' is displayed. Underneath, the 'Fonte dos Dados' is listed as 'http://servicosssi.unb.br/dadosabertos/orgaos'. A 'Filtros:' section contains a table with columns for 'Parâmetro', 'Tipo de Parametros', 'Valor', and an 'Incluir' button. Below this is a table with columns for 'Nome', 'Tipo', 'Valor', and 'Ação'. The table contains two rows: one for 'ano' (Parâmetro Temporal, Anual) with an 'Excluir' button, and one for 'departamento' (Parâmetro Fixo, Departamento de Administração) with an 'Excluir' button. The bottom section, 'Metadados de Identificação', includes fields for 'Titulo' (Conjunto de Dados de Órgãos), 'Descrição' (Dados sobre os órgãos da administração pública federal), and 'Orgão Responsável' (Universidade de Brasília).

Nome	Tipo	Valor	Ação
ano	Parâmetro Temporal	Anual	Excluir
departamento	Parâmetro Fixo	Departamento de Administração	Excluir

Metadados de Identificação

Titulo
Conjunto de Dados de Órgãos

Descrição
Dados sobre os órgãos da administração pública federal

Orgão Responsável
Universidade de Brasília

Fonte: Elaborado pelos autores.

A Figura 3 apresenta a interface que contém o serviço de requisição dos dados com dois parâmetros configurados de exemplo: um temporal e um fixo.

Em seguida são fornecidas as principais informações que descrevem o conjunto de dados. Estes dados são padronizados a partir dos metadados definidos pela "Cartilha Técnica de Publicação de Dados Abertos do Governo Federal" (BRASIL, 2011). Eles podem ser obrigatórios ou opcionais, sendo que alguns destes metadados são gerados automaticamente.

Os metadados são utilizados para descrever as fontes de dados. Dentre os metadados obrigatórios pode-se destacar o "Vocabulário Controlado do Governo Eletrônico" (VCGE) (BRASIL, 2019), que é um vocabulário controlado com o objetivo de

indexar informações (documentos, bases de dados, sites, etc.) no âmbito do governo federal que possui uma lista de termos sobre diversos assuntos no qual o governo atua. O uso destes metadados tem a finalidade de categorizar o conjunto de dados por assunto e auxiliar no agrupamento dos dados.

6.3 ESCOLHA DO VOCABULÁRIO CONTROLADO

Na etapa Escolha do Vocabulário Controlado, são definidos os metadados e ontologias que serão utilizados para representar semanticamente os dados. O vocabulário utilizado para o catálogo e para os conjuntos de dados podem ser diferentes.

Conforme é apresentado na Figura 4, a ferramenta apresenta uma lista de metadados e ontologias cadastradas previamente. É importante ressaltar que o Agente Publicador deve conhecer o conjunto de dados e também o vocabulário que deseja utilizar, assim, ele irá selecionar o vocabulário que melhor se adéqua para a representação semântica dos dados.

Figura 4 – Etapa Escolha do Vocabulário Controlado.

<input type="checkbox"/>	Ontologia	Prefixo	Endereço
<input checked="" type="checkbox"/>	Academic Institution Internal Structure Ontology	aiiso	http://purl.org/vocab/aiiso/schema
<input type="checkbox"/>	Dublin Core Metadata Element Set, Version 1.1	dc	http://purl.org/dc/elements/1.1
<input checked="" type="checkbox"/>	Estruturas Organizacionais Governamentais Brasileiras	siorg	http://vocab.e.gov.br/2011/09/org
<input type="checkbox"/>	Friend of a Friend	foaf	http://xmlns.com/foaf/0.1/
<input checked="" type="checkbox"/>	UnB Vocabulário	uvoc	http://dadosabertos.unb.br/images/UnBVocabulario.owl
<input type="checkbox"/>	Univ-bench Ontology	lubm	http://swat.cse.lehigh.edu/onto/univ-bench.owl
<input checked="" type="checkbox"/>	DCMI Metadata Terms	dcterms	http://purl.org/dc/terms
<input type="checkbox"/>	Metadata terms related to the DCMI Abstract Model	dcam	http://purl.org/dc/dcam
<input type="checkbox"/>	DCMI Type Vocabulary	dcmitype	http://purl.org/dc/dcmitype

[< Voltar](#) [Próximo >](#)

Fonte: Elaborado pelos autores.

6.4 REPRESENTAÇÃO SEMÂNTICA

A última etapa para publicação dos dados é a Representação Semântica. Esta etapa consiste em fornecer semântica aos dados por meio dos termos existentes nas ontologias selecionadas e, para isso, é necessário transformar o conjunto de dados estruturados em triplas RDF.

Considerando que os dados a serem publicados estejam em formato CSV, formato mais comum para a publicação de dados abertos, a transformação desses dados em dados conectados é realizada por meio da criação de triplas em RDF. Essas triplas RDF são formadas de modo que cada linha do arquivo CSV é considerada um recurso com uma URI única formando um sujeito e os dados contidos em cada coluna da linha são os objetos. Um termo oriundo de um vocabulário controlado deve ser escolhido para ser o predicado da tripla RDF a ser formada.

A Figura 5 apresenta um exemplo de publicação de triplas RDF criadas a partir de dados oriundos de quatro conjuntos de dados em CSV extraídos das bases de dados da UnB, são eles:

- Conjunto de dados de Ofertas de Matérias: os dados das ofertas são referentes às disciplinas que são oferecidas pelos departamentos em um determinado período letivo, obrigatoriamente elas possuem um professor vinculado.
- Conjunto de dados de Professores: dados referentes ao corpo docente da UnB. Este conjunto de dados apresenta apenas informações que já são de domínio público, não apresentando dados pessoais que configurem quebra de privacidade.
- Conjunto de dados de Matérias: dados referentes às disciplinas que são ofertadas pelos departamentos.
- Conjunto de dados de Departamentos: dados referentes aos departamentos internos da UnB.

Os quadros 1, 2, 3 e 4 apresentam os campos referentes a cada conjunto de dados com uma breve descrição de suas características.

Quadro 1 – Conjunto de dados de Ofertas de Matérias

Campos	Descrição	Tipo de Campo
COD_OFERTA	Código identificador do Oferta	Numérico Inteiro
TURMA	Nome da Turma	Carácter
VAGAS_OFERECIDAS	Quantidade de vagas oferecidas	Numérico Inteiro
PERIODO	Período letivo em que a matéria é oferecida	Numérico Inteiro
MATRICULA_PROFESSOR	Matrícula do professor que ministra a matéria	Numérico Inteiro
COD_MATERIA	Código identificador da matéria oferecida	Numérico Inteiro
COD_DEPARTAMENTO	Código identificador do departamento	Numérico Inteiro

Fonte: Elaborado pelos autores.

Quadro 2 – Conjunto de dados de Professores

Campos	Descrição	Tipo de Campo
MATRICULA	Matrícula do Professor	Numérico Inteiro
NOME_PROFESSOR	Nome do Professor	Carácter

Fonte: Elaborado pelos autores.

Quadro 3 – Conjunto de dados de Matérias

Campos	Descrição	Tipo de Campo
COD_MATERIA	Código identificador da Matéria	Numérico Inteiro
NOME_MATERIA	Nome da Matéria	Carácter
COD_DEPARTAMENTO	Código identificador do departamento da matéria	Numérico Inteiro
SIGLA_DEPARTAMENTO	Sigla do departamento da matéria	Carácter
NOME_DEPARTAMENTO	Nome do Departamento da matéria	Carácter
CREDITOS	Quantidade de crédito que possui a matéria	Numérico Inteiro

Fonte: Elaborado pelos autores.

Quadro 4 – Conjunto de dados de Departamentos

Campos	Descrição	Tipo de Campo
COD_DEPARTAMENTO	Código identificador do Órgão	Numérico Inteiro
NOME_DEPARTAMENTO	Nome do Órgão	Carácter
SIGLA	Sigla do Cargo	Carácter
COD_ORGAO	Código identificador do Órgão	Numérico Inteiro

Fonte: Elaborado pelos autores.

Após o levantamento de metadados e ontologias relacionados a dados acadêmicos, a ontologia *Lehigh University Benchmark* (LUBM) (GUO; PAN; HEFLIN, 2005) mostrou-se aderente ao objetivo desta pesquisa. Esse vocabulário visa descrever estruturas organizacionais das universidades com seus departamentos, cursos, faculdades e professores, porém não cobriu todo o domínio observado. Foi necessária a criação de uma ontologia leve, para preencher as lacunas existentes, denominada *UnB Vocabulary* (UVOC), cuja especificação está disponível em (MARTINS, 2018). Também foram selecionados para este estudo de caso a ontologia *Friend of a Friend* (FOAF) e o conjunto de metadados Dublin Core Metadata Element (DC).

A seguir são apresentadas as informações dos conjuntos de dados utilizados para criar as triplas RDF:

- **Ofertas de Matérias:** os campos `cod_oferta`, `turma`, `vagas_oferecidas` e `periodo` são literais com os predicados `dc:identifier`, `uvoc:Turma`, `uvoc:NumeroVagas` e `uvoc:periodo`. Os campos `matricula_professor`, `cod_materia` e `cod_departamento` são utilizados para integração com os conjuntos de dados e Professores, Matérias e Departamentos.

- **Professores:** este conjunto de dados terá seus dois campos publicados como objetos literais, sendo que `matricula` utilizará o predicado `uvoc:MatriculaProfessor` e `nome_professor`, o predicado `foaf:name`.

- **Matérias:** no conjunto de dados de Matérias, os objetos literais serão `cod_materia`, `nome_materia` e `créditos` com os predicados `dc:identifier`, `foaf:name` e `uvoc:QuantidadeCredito`. O campo `cod_departamento` fará ligação com o conjunto de dados de Departamento.

- **Departamentos:** decidiu-se publicar todos os campos, sendo que os campos `cod_departamento` e `nome_departamento` serão tratados como objetos literais utilizamos os predicados `dc:identifier` e `foaf:name`.

A conexão dos conjuntos de dados é uma das funcionalidades mais importantes do UnBGOLD, desta maneira é possível garantir a interoperabilidade entre conjuntos de dados publicados distintos. A Figura 5 apresenta a interface da ferramenta utilizada para a publicação dos dados conectados relativos ao conjunto de dados de Oferta de Matérias e os conjuntos de dados relacionados. Como exemplo, são apresentados no Quadro 5 os dados referentes a uma linha do conjunto de dados de Oferta.

Figura 5 – Publicação de triplas RDF.

Campo	Publicar	Predicado	Objeto	Complemento
cod_oferta	<input checked="" type="checkbox"/>	Selecione o Vocabulário dc:identfier	Selecione o Objeto Objeto Literal	Selecione
turma	<input checked="" type="checkbox"/>	Selecione o Vocabulário uvoc:turma	Selecione o Objeto Objeto Literal	Selecione
vagas_oferecidas	<input checked="" type="checkbox"/>	Selecione o Vocabulário uvoc:numero_vagas	Selecione o Objeto Objeto Literal	Selecione
periodo	<input checked="" type="checkbox"/>	Selecione o Vocabulário uvoc:periodo	Selecione o Objeto Objeto Literal	Selecione
matricula_professor	<input checked="" type="checkbox"/>	Selecione o Vocabulário lubm:Professor	Selecione o Objeto Conjunto de dados de professores	Selecione matricula_professor
cod_materia	<input checked="" type="checkbox"/>	Selecione o Vocabulário lubm:Course	Selecione o Objeto Conjunto de dados de Matérias	Selecione cod_materia
sigla_departamento	<input checked="" type="checkbox"/>	Selecione o Vocabulário lubm:Departament	Selecione o Objeto Conjunto de dados de departamento	Selecione sigla_departamento

Fonte: Elaborado pelos autores.

Quadro 5 – Exemplo de Dados do Conjunto de Ofertas de Matérias.

Campos	Valor
COD_OFERTA	659991
TURMA	A
VAGAS_OFERECIDAS	40
PERIODO	20181
MATRICULA_PROFESSOR	149055
COD_MATERIA	185221
COD_DEPARTAMENTO	588

Fonte: Elaborado pelos autores.

Cada linha do conjunto de dados Ofertas de Matérias será transformada em uma tripla RDF durante o processo de publicação. O valor atribuído ao Sujeito de cada tripla será formado pelo prefixo “http://dadosabertos.unb.br/ofertadisciplinas/” concatenado com o atributo identificador de cada linha do conjunto de dados de Ofertas de Matérias, neste caso o campo “cod_oferta” que é “659991” (Quadro 5).

O prefixo é formado por uma URL do domínio da UnB “<http://dadosabertos.unb.br/>” concatenada com um nome que represente o conjunto de dados, por exemplo, “ofertadisciplinas/”. Essa estratégia garante que cada conjunto de dados possua um prefixo específico.

Ao concatenar o prefixo de um conjunto de dados com o seu atributo identificador para formar o sujeito da tripla, é possível garantir que cada linha do conjunto de dados seja mapeada para uma tripla RDF com um Sujeito exclusivo.

Os Predicados das triplas são formados pelos termos escolhidos a partir dos vocabulários cadastrados na ferramenta, pode ser usado qualquer termo de um vocabulário cadastrado na ferramenta, essa atividade fica por conta do Agente Publicador.

Finalmente, os Objetos das triplas são formados pelos atributos dos conjuntos de dados. Esses objetos poderão ser do tipo “Objeto Literal”, que representa um literal, ou um recurso, que representa um Sujeito de uma outra tripla RDF já publicada, instanciando, assim, uma conexão entre conjuntos de dados distintos.

Na Figura 5, as quatro primeiras linhas da interface apresentada, destacadas em azul, representam os atributos do conjunto de dados de Oferta de Matérias que são configuradas como literais, devido ao tipo de objeto selecionado na ferramenta ser “Objeto Literal”. Nesse caso, cada atributo do conjunto de dados será transformado em um Objeto que representa um literal em uma tripla RDF.

A relação do Sujeito gerado com o Objeto será descrita pelo Predicado, que é um termo vindo dos vocabulários escolhidos.

As demais linhas, que estão destacadas em vermelho, são usadas como ponte para a ligação a outros conjuntos de dados já publicados, devido ao tipo de objeto selecionado na ferramenta ser algum conjunto de dados já publicado, como por exemplo, “Conjunto de dados de Professores”. Nesse caso, o conjunto de dados escolhido já está representado por triplas RDF.

Na Figura 5, pode-se observar também que para o campo “matricula_professor” foi selecionado o vocabulário “[lubm:Professor](#)”, devido ao fato desse vocabulário estar cadastrado no UnBGOLD.

O conjunto de dados de Ofertas de Matérias utiliza o mesmo vocabulário que o conjunto de dados de Professores que já foi publicado, ou seja, está representado por triplas RDF. Assim, o Agente Publicador pode selecionar o conjunto de dados de

Professores no campo Objeto da interface da ferramenta para realizar a conexão entre triplas RDF.

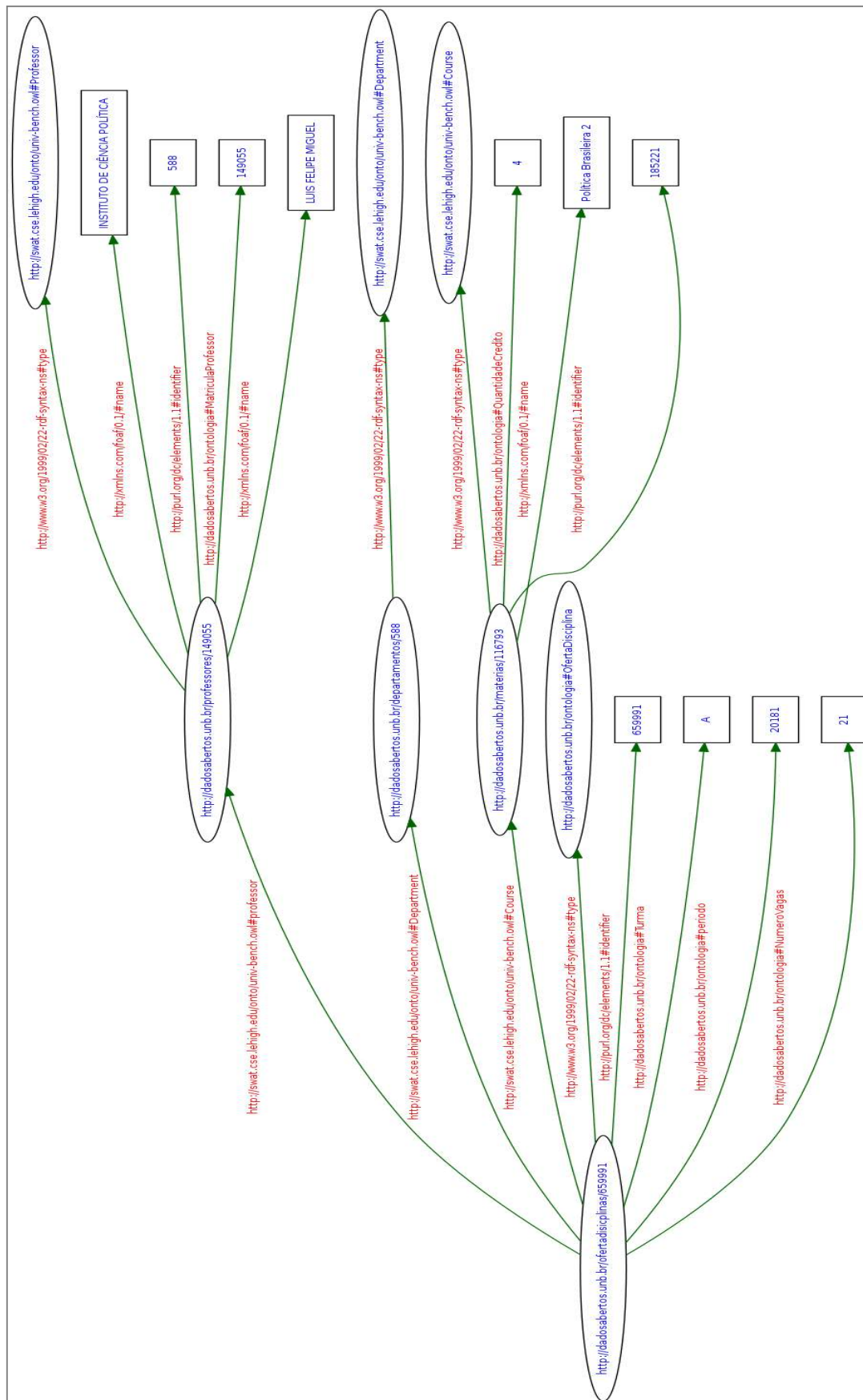
Nesse caso, a opção selecionada foi “Conjunto de Dados de Professores”, e no campo “Complemento” da respectiva linha é habilitado uma lista dos nomes das colunas que o conjunto possui, na qual o Agente Publicador irá selecionar o campo identificador para ser usado na a ligação por meio da comparação entre os valores. Neste caso o Agente Publicador selecionou “matricula_professor” na opção complemento. Este processo funciona semelhante ao uso de chaves primárias e estrangeiras em tabelas relacionais.

No exemplo, o valor referente à “matricula_professor” que está em no conjunto de dados relativo à oferta é “149055”. Então, o Agente Publicador irá procurar a coluna “matricula_professor” do conjunto de dados de Professores escolhido para ligação na tripla em que o valor seja exatamente igual, no caso “149055”.

Neste exemplo, foi encontrada uma tripla do tipo (rdf:type) Professor, cujo nome (foaf:name) é “LUIS FELIPE MIGUEL” e matrícula “14905”, ou seja, essa tripla possui a relação do conjunto de dados de Professor com o de Ofertas de Matérias. A Figura 6 apresenta o resultado da ligação.

A publicação dos campos “cod_departamento” e “cod_materia” do conjunto de dados Oferta de Matérias seguirá a mesma lógica para a conexão com o conjuntos de dados de Departamentos e de Matérias, respectivamente.

Figura 6 – Grafos das triplas do conjunto de dados de Ofertas de Matérias.



Fonte: Elaborado pelos autores.

7 CONSIDERAÇÕES FINAIS

A publicação de dados abertos conectados vem se tornando um dos recursos mais utilizados por organizações públicas e privadas para a interoperabilidade e disponibilização da informação ao público em geral.

A publicação de dados abertos pelas instituições públicas brasileiras é uma realidade ao qual os órgãos públicos precisam adaptar-se, mesmo que ainda não haja um processo amplamente definido para que isso ocorra. Pode-se inferir que a divulgação de dados governamentais de forma aberta e conectada incrementa a transparência da administração pública e pode proporcionar inúmeros benefícios aos governos e cidadãos.

A proposta de arquitetura de publicação de dados abertos conectados apresentada neste trabalho busca contemplar as necessidades da UnB, estabelecendo o emprego de diversos recursos oriundos da Ciência da Informação e da Ciência da Computação utilizados para otimizar as atividades de publicação de conjuntos de dados incrementando a qualidade dos dados publicados por meio da representação semântica. No entanto, todos os softwares utilizados são livres, o que torna essa arquitetura possível de ser utilizada por qualquer órgão público que tenha interesse em publicar seus dados de forma conectada.

Verificou-se que com o subsídio à automação da publicação de dados abertos conectados proporcionado pela arquitetura apresentada é possível que os dados possam ser publicados com qualidade em intervalos de tempo menores, garantindo que o dado seja o mais atual possível. Por outro lado, a representação semântica viabiliza que os dados não sejam publicados isoladamente, oportunizando análises mais sofisticadas por meio da integração de conjuntos de dados gerados por órgãos distintos.

Do exposto, espera-se que a arquitetura apresentada estimule os órgãos públicos a publicarem seus dados abertos de forma conectada, proporcione ao gestor público realizar análises de conjuntos de dados gerados pelos mais diversos órgãos da administração pública de forma integrada e incentive pesquisadores e cidadãos comuns a fiscalizar as ações governamentais por meio da análise de dados abertos conectados governamentais publicados com qualidade e celeridade.

REFERÊNCIAS

AGILAR, Everton de Vargas. **Uma abordagem orientada a serviços para a modernização de sistemas legados**. 2016. Dissertação (Mestrado) — Programa de Pós-Graduação em Computação Aplicada, Departamento de Ciência da Computação, Universidade de Brasília. Brasília Disponível em: <https://core.ac.uk/download/pdf/80746649.pdf>. Acesso em: 10 abr. 2019.

BAILEY, S. **Information architecture: a brief introduction**. 2003. Disponível em: <https://www.iainstitute.org/sites/default/files/bailey-iaintro.pdf> Acesso em: 19 novembro 2019.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. **Scientific american**, v. 284, n. 5, p. 28-37, 2001. Disponível em: [https://kask.eti.pg.gda.pl/redmine/projects/sova/repository/revisions/master/entry/doc/Master%20Thesis%20\(In%20Polish\)/materials/10.1.1.115.9584.pdf](https://kask.eti.pg.gda.pl/redmine/projects/sova/repository/revisions/master/entry/doc/Master%20Thesis%20(In%20Polish)/materials/10.1.1.115.9584.pdf). Acesso em: 15 jan. 2019.

BERNERS-LEE, Tim, Tim. **Linked data-design issues**. W3C, 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 15 jan. 2019.

BRASIL. Constituição (1988a). Lei 12.527, de 18 de novembro de 2011. Dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso a informações. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm. Acesso em: 12 abr. 2019.

BRASIL. Constituição (1988b). Decreto 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm. Acesso em: 10 abr. 2019.

BRASIL. Governo Federal. **Cartilha técnica para publicação de dados abertos no Brasil**. Brasília: Governo Federal, 2011; 1.0. Disponível em: <http://wiki.dados.gov.br/GetFile.aspx?Page=Tecnologia&File=Cartilha%20T%c3%a9cnica%20para%20Publica%c3%a7%c3%a3o%20de%20Dados%20Abertos%20no%20Brasil%20v1.pdf>. Acesso em: 25 mai. 2019.

BRASIL. Governo Federal. **Vocabulário Controlado do Governo Eletrônico**. Brasília: Governo Federal, 2019. Disponível em: <https://www.gov.br/governodigital/pt-br/governanca-de-dados/vocabulario-controlado-do-governo-eletronico>. Acesso em: 21 dez. 2019.

DUBLIN CORE METADATA INITIATIVE – DMCI. **Dublin core metadata element set**, version 1.1. 2012. Disponível em: <http://dublincore.org/documents/dces/>. Acesso em: 15 jul. 2018.

EAVES, David. 2009. **The Three Laws of Open Government Data**. eaves.ca (blog), September 30. Disponível em: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/> Acesso em: 13 fev. 2019.

GUARINO, N. Formal Ontology and Information Systems. In: **PROCEEDINGS OF FOIS'98**. Amsterdam, Netherlands: IOS Press, p.3-15, jun. 1998.

GUO, Yuanbo; PAN, Zhengxiang; HEFLIN, Jeff. LUBM: A benchmark for OWL knowledge base systems. **Journal of Web Semantics**, v. 3, n. 2-3, p. 158-182, 2005. Disponível em: <http://swat.cse.lehigh.edu/pubs/guo05a.pdf> Acesso em: 25 abr. 2019.

HEATH, Tom; BIZER, Christian. **Linked Data: Evolving the Web into a Global Data Space**. Synthesis Lectures on the Semantic Web 1. [San Rafael, CA]: Morgan & Claypool. 2011.

HJØRLAND, Birger. Classification. **KO KNOWLEDGE ORGANIZATION**, v. 44, p. 97-128, 2017.

ISOTANI, Seiji.; BITTENCOURT, Ig Ibert. Dados Abertos Conectados: **Em busca da Web do Conhecimento**. São Paulo: Novatec, 2015.

LOD PROJECT - *Linking Open Data* (LOD) Project. 2010. Disponível em: https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#Project_Description. Acesso em: 15 dez. 2019.

MARCONDES, Carlos Henrique. “Linked data”—dados interligados—e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 34, p. 171-192, 2012. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/24410>. Acesso em: 12 abr. 2018.

MARTINS, Luiz Carlos Barbosa. **Proposta de arquitetura de publicação automatizada de dados abertos conectados utilizando meta-dados e ontologias**. 2018. Dissertação (Mestrado) — Programa de Pós-Graduação em Computação Aplicada, Departamento de Ciência da Computação, Universidade de Brasília. Brasília Disponível em: <https://repositorio.unb.br/handle/10482/34816>. Acesso em: 10 abr. 2019.

PEREIRA, Durval Vieira; MARCONDES, Carlos Henrique. Modelagem e Representação Semântica de Dados Governamentais Abertos da Previdência Social Brasileira. **ENANCIB**, Belo Horizonte - MG, October 2014.

RAUTENBERG, Sandro; BURDA, Alessandra Cassiana; DE SOUZA, Lucélia. Um workflow para compartilhamento de dados científicos primários baseado em dados abertos conectados. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 23, n. 53, p. 110-123, 2018. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p110>. Acesso em: 20 fev. 2019.

SANTARÉM SEGUNDO, José Eduardo. Tecnologias de informação e comunicação para disponibilização de dados abertos em formato semântico. **Ibersid**, v. 7, 2013. Disponível em: <https://www.ibersid.eu/ojs/index.php/ibersid/article/view/4075/3744>. Acesso em: 18 mai. 2019.

SHAW, Mary; GARLAN, David. Formulations and formalisms in software architecture. In: **Computer Science Today**. Springer, Berlin, Heidelberg, p. 307-323, 1995.

SMIRAGLIA, Richard P. Replication and Accumulation in Knowledge Organization—An Editorial. **KO KNOWLEDGE ORGANIZATION**, v. 44, n. 5, p. 315-317, 2017.

TRIBUNAL DE CONTAS DA UNIÃO – TCU. Secretaria de Fiscalização de Tecnologia da Informação. **5 motivos para a abertura de dados na Administração Pública**. Brasília, 2015. Disponível em:

<https://portal.tcu.gov.br/lumis/portal/file/fileDownload.jsp?fileId=8A8182A24F0A728E014F0B36E7016F34>. Acesso em: 23 mai. 2019.

Notas

AGRADECIMENTOS

Não se aplica

CONTRIBUIÇÃO DE AUTORIA

Concepção e elaboração do manuscrito: M. C. Victorino, L. Martins, M. Holanda e R. Fonseca.

Software: L. Martins e M. C. Victorino,

Discussão dos resultados: M. C. Victorino, L. Martins, M. Holanda e R. Fonseca.

CONJUNTO DE DADOS DE PESQUISA

- 1) O conjunto de dados que dá suporte aos resultados deste estudo não está disponível publicamente.

FINANCIAMENTO

Não se aplica

CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica

CONFLITO DE INTERESSES

Não se aplica

LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHER

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Enrique Muriel-Torrado, Edgar Bisset Alvarez, Camila Barros.

HISTÓRICO

Recebido em: 19/09/2019 – Aprovado em: 04/03/2020 – Publicado em: 08/05/2020

