

A publicação de dados abertos conectados nas instituições públicas federais brasileiras: o caso do sistema CONFEA/CREAs

The publication of open linked data in Brazilian federal public institutions:
the case of the CONFEA/CREAs system

Fabrizio Foresti

Doutor em Ciência da Informação pela Universidade Federal de Santa Catarina (UFSC). Gerente do Departamento de Arquivo do Conselho Regional de Engenharia e Agronomia de Santa Catarina (CREA-SC).
E-mail: forestifabrizio@gmail.com

Gregório Varvakis

Doutor em Manufacturing Engineering pela Loughborough University of Technology. Professor titular da Universidade Federal de Santa Catarina (UFSC).
E-mail: g.varvakis@ufsc.br

RESUMO

Os dados abertos governamentais são tendência das democracias contemporâneas. As instituições públicas estão inseridas no contexto hodierno do excesso de informação e da necessidade de transparência. Os documentos de arquivo, enquanto suportes que estruturam e coletam dados, podem entregar mais valor aos usuários através do uso das tecnologias de web semântica. O objetivo do estudo é propor um procedimento para publicação de dados abertos e apontar como os arquivos podem atuar nesse contexto. Os procedimentos metodológicos envolvem pesquisa bibliográfica nas bases de dados Scopus e Brapci para recuperar artigos publicados entre 2012-2018 com o termo 'Linked Open Data' contido no título e a abertura dos dados contidos em um documento produzido pelo Conselho Regional de Engenharia e Agronomia de Santa Catarina. Os resultados mostram a possibilidade de aplicação das tecnologias de Linked Data na rotina de gestão de documentos e apresentam um caminho possível para lidar com a questão dos dados abertos em instituições públicas federais brasileiras. Conclui que é possível se valer das técnicas de web semântica para abertura dos dados no contexto da gestão de documentos e que os arquivos são os protagonistas naturais da abertura dos dados nas instituições públicas.

Palavras-chave: Dados abertos ligados. Gestão de Documentos. Arquivos Públicos.

ABSTRACT

Open government data is a trend in contemporary democracies. Public institutions are part of today's context of excess information and the need for transparency. Archival documents, as supports that structure and collect data, can deliver more value to users using semantic web technologies. The objectives of the study are to propose a procedure for publishing open data and to pointing out how archives can act in this context. The methodological procedures involve bibliographic research in Scopus and Brapci databases to retrieve articles published between 2012-2018 with the term 'Linked Open Data' contained in the title and opening the data contained in a document produced by the Regional Council of Engineering and Agronomy of Santa Catarina. The results show the possibility of applying Linked Data technologies in the document management routine and present a possible way to deal with the issue of open data in Brazilian federal public institutions. It concludes that it is possible to use semantic web techniques to open data in the context of document management and that archives are the natural protagonists of data opening in public institutions.

Keywords: Open linked data. Record management. Public Archives.

1 INTRODUÇÃO

A necessidade de abertura dos dados é fomentada pelo cenário contemporâneo altamente tecnológico, cujas demandas sociais orbitam cada vez mais em torno do consumo de dados e informação para criar novos produtos e serviços. O Conselho Federal de Engenharia e Agronomia em conjunto com os Conselhos Regionais de Engenharia e Agronomia, formam o chamado sistema CONFEA/CREAs. São instituições que fiscalizam o exercício da profissão de engenheiro.

São autarquias públicas federais que estão inseridas no contexto contemporâneo do excesso de informação. Ao mesmo tempo, são organizações que precisam cumprir a legislação federal, abrir seus arquivos e compartilhar dados em favor da transparência pública. Inúmeros documentos produzidos por essas organizações são de grande valia para a sociedade, a exemplo do chamado receituário agrônomo, documento que regula o uso de agrotóxicos pelos Engenheiros Agrônomos. Outro documento importante é a chamada Anotação de Responsabilidade Técnica (ART), documento que controla a realização de todas as obras e serviços de Engenharia.

O processamento em tempo real adquire importância superior e as técnicas tradicionais já não atendem mais as demandas hodiernas. A compreensão da informação pelas máquinas surge como alternativa para lidar, não apenas com o excesso de informação, mas também com as demandas impostas pela da ubiquidade da informação, da tecnologia e dos usuários (pessoas e organizações).

Destacam-se neste contexto as tecnologias *Linked Data*, que permitem o processamento mecânico da informação, ao mesmo tempo em que auferem mais sentido ao conteúdo em rede e ligam as diversas fontes de informação. A rede evoluiu como um imenso documento que apenas exibe informação, eis a questão que a web semântica pretende resolver, ao tornar as máquinas capazes de compreender o que no momento apenas mostram aos usuários (BERNERS-LEE; HENDLER; LASSILA, 2001).

As tecnologias da informação já promovem mudanças no acesso e uso da informação pública, na legislação, nas instituições, e podem tornar realidade à abertura dos dados públicos no Brasil. A Lei de Acesso à Informação (BRASIL, 2011) destaca a importância de as instituições públicas compartilharem documentos em formato aberto. E os arquivos são centrais quanto ao acesso e uso da informação pública.

Deste modo, o objetivo do estudo é indicar um procedimento para publicação de dados abertos em uma instituição pública e mostrar como os arquivos

podem atuar nesse contexto. Assim, o problema da pesquisa se traduz na seguinte pergunta: como realizar a abertura dos dados nas instituições públicas federais brasileiras?

2 PROCEDIMENTOS METODOLÓGICOS

A pesquisa bibliográfica é realizada nas bases de dados Scopus e Brapci¹ para recuperar artigos publicados entre 2012-2018 com o termo "*Linked Open Data*" contido no título. A escolha da base Scopus se deve à familiaridade do autor com a ferramenta. A base de dados Brapci foi escolhida por ser a fonte brasileira que reúne a produção científica em Ciência da Informação, além de ser a fonte dos trabalhos que inspiraram a presente pesquisa.

A proposta se vale da instituição Conselho Regional de Engenharia e Agronomia de Santa Catarina - Crea-SC para exemplificação. A coleção utilizada para exemplo de abertura dos dados de um documento é a coleção de Portarias, escolhida por ser o instrumento de comunicação oficial dos atos praticados na instituição. De modo que é realizada a descrição RDF da Portaria n. 01/2018, por meio das técnicas *Linked Open Data* (LOD), para que os dados do documento se tornem legíveis semanticamente por máquinas.

A proposta se delinea da seguinte forma: inicialmente é apresentada revisão de literatura sobre os conceitos envolvidos: *Linked Open Data* e gestão de documentos. Em seguida, são apresentados os trabalhos relacionados mais importantes recuperados pela revisão de literatura. Por fim, é apresentada a proposta, seguida das conclusões.

3 LINKED OPEN DATA (LOD)

A forma como as informações estão dispostas em rede impedem e/ou dificultam o uso e reuso da informação, e é justamente a reutilização improvável da informação que contém o diferencial ofertado pela rede (BERNERS-LEE, 2006). Eis a importância do conceito de *Linked Open Data* (LOD), que emerge como alternativa para tornar os dados dispostos em rede legíveis por máquinas e assim, facilitar o uso da informação (pública ou privada). Refere à distribuição de dados de forma gratuita para uso geral.

¹ Base de dados de artigos referenciais em Ciência da Informação.

O conceito de Linked Open Data está associado ao conceito de big data. Mas o que é big data? O termo talvez não seja adequado, pois o nome faz parecer que o tamanho é o único problema explica Zikopoulos et al. (2012, p. 3). Segundo o autor, o cenário big data, diz respeito àquela “informação” que não é possível processar tecnicamente (selecionar, avaliar, classificar) pelos meios habituais; a questão gira em torno, em essência, de como “obter valor” numa “informação” cujo “formato” é inadequado, que está “crua”, “não estruturada” ou “semiestruturada”: trata-se de um “verdadeiro desafio informacional”. O foco do cenário big data é se valer da grande quantidade de dados dispostos em rede (CHIGNARD, 2013).

O processamento de informação no cenário big data depende de variáveis diversas e inter-relacionadas, como por exemplo, a web semântica, padrões internacionais, dados ligados, entre outros. A web semântica é área que trata dos sentidos que as pessoas dão aos objetos e acontecimentos (AQUARONE; LUZ, 2017, p. 32). Segundo Fermoso et al. (2015, p. 619) a web semântica é “uma nova forma de entender a rede, onde os dados são ajustados com significado semântico”, comumente reconhecida como “web dos dados”. A web semântica visa fazer com que as máquinas compreendam melhor o que, no momento, apenas mostram aos usuários (BERNERS-LEE; HENDLER; LASSILA, 2001). Dentre as tecnologias associadas à web semântica destaca-se o linked data.

Os dados ligados (*linked data*) são apontados como o cerne da web semântica (W3C, 2018). Bizer, Heath e Berners-Lee (2009) explicam que *linked data* é um “conjunto de boas práticas” de disseminação e interligação de informação organizada (ou “dados estruturados”) em rede. Segundo os autores é se valer da rede para gerar ligações através das mais diversas fontes de informação.

Para Fermoso et al. (2015, p. 619) diz respeito à técnica aplicada para interligar dados em rede e, ao adicionar o termo “open”, passa a se tratar da interligação de dados disseminados de maneira livre por organizações em geral. Segundo os autores, o objetivo é disseminar conteúdo passível de uso por terceiros, e assim, gerar novo valor à informação e permitir a emergência de “new citizen services”. O Quadro 1 apresenta algumas tecnologias que estão associadas com a publicação de dados abertos ligados.

Quadro 1: Conjunto de tecnologias usadas na abertura dos dados.

| Ferramenta | Função | Motivo da escolha |
|--|--|---|
| CKAN | Servir de catálogo de dados abertos ou instrumento de pesquisa dos conjuntos de dados abertos. | É o programa “mais usado para a publicação de catálogo de dados abertos ao redor do mundo” de acordo com Segundo (2015). O endereço eletrônico do projeto é ckan.org e possui vasto material explicativo e suporte aos usuários. Segundo o CKAN (2018) é “a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data”. |
| Sintaxe RDF*/XML e JSON | Modela os dados para os formatos desejados. Para usar a sintaxe é preciso usar também os vocabulários e criar links para os objetos que serão descritos. | RDF é o formato que oferta ótima oportunidade “para a interoperabilidade e processamento automático de dados governamentais” segundo a W3C (2011, p. 37). É uma “linguagem para representação de informações na Internet” que se vale de “arquivos baseados em metadados” visando à criação de “modelo de documento no formato XML com semântica” explica a W3C (2011b, p. 25-26); de uso fácil e com ordem da informação “unificada”, possui “estrutura baseada em URI” (local dos documentos na rede) e em XML (estruturação do documento). Por sua vez, a sintaxe JSON é um formato “leve e simples”, com “estrutura [...] de fácil entendimento”, que se vale de “convencões comuns em muitas linguagens de programação” conforme a W3C (2011b, p. 35), |
| Vocabulários dbpedia, dublicore, foaf, e RDFschema | Originam as ontologias que emergem das ligações entre sujeito, predicado e objeto. | Todos são complementares e se constituem em recomendação da W3C. Os vocabulários possuem referências em rede, como por exemplo: <code><https://schema.org/></code> e <code><http://xmlns.com/foaf/spec/></code> , que facilitam o uso com explicações e exemplos de aplicação. |
| Linguagem de consulta ao banco de dados de triplas RDF* SPARQL | Realizar consultas nos dados convertidos por meio das triplas RDF | Após a estruturação dos dados, o uso de “linguagens para consulta” ajuda a “pesquisar ou referenciar dados contidos em documentos” em rede, dentre as “linguagens” destaca-se a “SPARQL” (<i>Protocol and RDF Query Language</i>) aponta a W3C (2011b, p. 25). |
| Formatos XML e CSV | Compartilhar as bases de dados selecionadas para abertura. | Em conformidade com a W3C (W3C, 2011b, p. 26) mais do que dispor “arquivos em formato RDF, é interessante que eles também sejam disponibilizados em outros formatos de fácil manipulação por humanos, como o CSV”. O formato XML é muito utilizado ao manter a “estrutura dos dados” nas diversas “operações”, explica a W3C (2011, p. 36-37); igualmente, o formato CSV é compacto e simples, também amplamente difundido, mas precisa ser bem documentado para não impossibilitar seu uso. |
| Validator | Validar a serialização | Pode ser usada ferramenta <code><https://www.w3.org/RDF/Validator/></code> para identificar erros de sintaxe. É fundamental checar a qualidade do serviço antes de disseminar os documentos. A ferramenta é de grande valor para localizar erros de serialização. |
| Dreamweaver | Interface para realizar a serialização RDF nos itens documentais. Pode ser substituído por qualquer outro programa de desenvolvimento web. | Trata-se de interface de programação e desenvolvimento de web sites da marca Adobe, amplamente disseminada. O manual da W3C para desenvolvedores (2011) mostra que é possível fazer a serialização RDF através das páginas web. Assim, a ferramenta é adequada para a tarefa, ainda que não se constitua numa API. |
| RDF Translator | Converter os dados dos documentos nos mais diversos formatos. Por exemplo, de RDF/XML para JSON. | Segundo o RDF Translator ² trata-se de “ferramenta de conversão [...] para marcação estruturada” que “permite conversões acionadas por URI ou por entrada de texto direta”. |

Fonte: Elaborado pelo autor. * *Resource Description Framework*.

² <http://rdf-translator.appspot.com>

Mas o que significa um documento aberto? Documento aberto segundo Bittencourt e Isotani (2018) é àquele “documento” posto em rede com licenças abertas (por exemplo, nos formatos “PDF ou PNG”), ou seja, para ser aberto (“dado aberto”) basta estar em rede com licenças abertas.

E qual o significado de dados abertos? Conforme a Open Knowledge International (2018) “open data and content can be freely used, modified, and shared by anyone for any purpose”. Nos dados abertos, a criação do valor não reside em grandes volumes, mas na capacidade de compartilhar e disponibilizar em rede dados aos usuários, observando critérios técnicos, econômicos e legais, de forma gratuita e em formatos que possibilitem o reuso (CHIGNARD, 2013).

Os cenários “big” e “open data” possuem o mesmo objeto (dados), contudo, o foco do cenário big data é explorar as possibilidades advindas do grande volume de dados em crescimento contínuo; nos dados abertos à criação do valor reside na capacidade de compartilhar (CHIGNARD, 2013).

O conceito BOLD (Big and Open Linked Data) de acordo com Janssen e Hoven (2015, p. 363) significa a “junção dos três conceitos anteriores: big data, dados abertos e dados ligados”. Segundo os autores “possui grande influência na forma como os governos operam e interagem com o público”. Contudo, as pesquisas sobre as técnicas de BOLD para uso dos governantes ainda são incipientes (JANSSEN; KUK, 2016, p. 3).

A questão BOLD ganha caráter especial quando se trata de instituições públicas. Jorente, Pádua e Segundo (2017, p. 4) verificam que as “tecnologias Linked Data estão sendo utilizadas para compartilhar dados que abrangem muitos assuntos” e que o tópico “governos” abarca parcela significativa das aplicações desenvolvidas. O BOLD está ligado ao chamado Governo Aberto. Os dados abertos governamentais (Open Government Data - OGD) possuem princípios que são sintetizados na Figura 1.

Figura 1: Princípios do Governo Aberto.



Fonte: Adaptada pelo autor de *Open Government Data* (2018).

Em relação ao Open Government Data, uma iniciativa em especial se destaca. A “Open Government Partnership” (OGP, 2018) é iniciativa que se consolida no ano de 2011, é liderada pelo Brasil, Estados Unidos, Indonésia, México, Filipinas, África do Sul e Reino Unido; se trata de “iniciativa multilateral” em busca do “comprometimento dos governos nacionais para promoção da transparência”, do “empoderamento dos cidadãos”, ao mesmo tempo em que corrobora o combate à “corrupção” e usa as “novas tecnologias para fortalecer” a democracia.

A Declaração do Open Government Partnership (OGP, 2018) mostra o compromisso de ampliar o acesso à informação pública, fortalecer a participação social e o uso de tecnologias da informação na abertura das informações. A iniciativa chama a atenção dos governos para a gestão da informação e para os resultados positivos advindos da abertura dos dados, explicam Janssen et al. (2017, p. 189). Segundo os autores o uso dos dados (ou “data-driven innovation”) pode “transformar o serviço público” e ser “usado por governantes” e pelos cidadãos para planejar ações, “compreender as implicações dos atos políticos” e servir de subsídio à tomada de decisão.

Mas como pode ocorrer a abertura dos dados? Bittencourt e Isotani (2018) explicam que a “abertura dos dados” e a transmutação de meros “dados abertos” para a condição de “dados abertos conectados” se dá através de um conjunto de “transformações intermediárias”, representadas pelo “esquema das cinco estrelas”.

O esquema cinco estrelas de Berners-Lee (2006) é uma orientação de boas práticas de publicação de dados abertos conectados. Dentre os princípios de Berners-Lee (2006) destaca-se o uso de dos padrões RDF e SPARQL, bem como o uso de links para dar nome aos objetos. Segundo o 5stardata (2018), o esquema cinco estrelas traz benesses para quem publica os dados e também para os usuários.

De modo que a questão BOLD se mostra como tendência que pode favorecer as unidades de informação, destacadamente os arquivos públicos. São conceitos que transformam a forma de gerir e compartilhar documentos públicos.

4 A GESTÃO DE DOCUMENTOS

Os princípios do *Open Government Partnership* apresentados na seção anterior são estritamente ligados com a razão de ser dos arquivos, que zelam pela prova, pelo acesso, pela transparência da administração pública. Arquivos públicos é o “conjunto de documentos produzidos” por organizações públicas conforme Paes (1997, p. 24). Os conjuntos de documentos públicos formam os arquivos públicos (BRASIL, 1991). Também representam as unidades de informação que realizam a gestão dos documentos nas organizações.

Documento é o “suporte da informação” explicam Castro, Castro e Gasparian (1985, p. 19) ou mesmo o “simples registro de uma informação” para além do seu “suporte” físico explica Paes (1997, p. 26). Contudo, segundo a autora, documento de arquivo é àquele tipo de documento produzido por uma organização, seja ela pública ou privada, no “exercício de suas atividades”, e possui caráter de “prova” ou “informação”. Ainda segundo a autora, documento público é àquele documento produzido por organizações públicas ao longo de suas atividades.

É o documento que origina e alimenta as bases de dados, que estruturam os dados em sua coleta e também na apresentação inteligível da informação aos usuários. E o documento enquanto suporte da informação pode ser digital. O dicionário Houaiss (2001) mostra que documento também é entendido como um “conjunto de dados digitalizados”.

Todas as informações produzidas pelas organizações públicas ao longo de suas atividades se constituem em documentos de arquivo, inclusive as bases de dados. É preciso destacar que a digitalização tem transformado o conceito de documento, inclusive o conceito de documento de arquivo (CASTRO; CASTRO; GASPARIAN, 1985).

A gestão dos documentos de arquivo compõe o arsenal metodológico da Ciência da Informação. De acordo com a Lei nº 8159 (BRASIL, 1991) “considera-se gestão de documentos o conjunto de procedimentos e operações técnicas à sua produção, tramitação, uso, avaliação e arquivamento em fase corrente e intermediária, visando a sua eliminação ou recolhimento para guarda permanente”.

O arquivo público é o responsável natural para lidar com a questão da abertura dos dados nas organizações públicas, e que o arquivista pode ser o grande facilitador do acesso aos dados públicos. Jardim e Salomão Neto (2016) reconhecem que são claros os “contornos arquivísticos do governo aberto” e a “participação” dos arquivos como atores é central, e é necessário para o *Open Government Partnership* que os arquivos busquem pautar a abertura dos dados em seu cotidiano. Segundo os autores o “sucesso das iniciativas de governo e dados abertos passam necessariamente” pela “gestão de documentos”, seja nos planos teóricos, práticos, políticos ou estratégicos; apenas assim, por meio deste reconhecimento ou “entendimento”, é que serão estabelecidos os instrumentos necessários ao “governo aberto”.

Na legislação brasileira se destacam o Decreto 8777 (BRASIL, 2016) estabelece a “Política de Dados Abertos do Poder Executivo Federal” em busca do aprimoramento da “cultura de transparência pública”. Em seu Art. 3º, aponta, entre outras questões, a necessidade de criar um Plano de Dados Abertos (PDA), documento orientador dos atos ligados à “abertura de dados” de cada organização “pública federal”, atendendo “padrões mínimos de qualidade” para o “entendimento e a reutilização das informações”. A Lei n. 12557/2011, amplamente conhecida como a Lei de Acesso à informação (BRASIL, 2011), também destaca a questão dos dados abertos.

Assim, buscou-se mostrar a forte inter-relação entre a gestão de documentos e a publicação *Linked Open Data* em instituições públicas. Observa-se a necessidade de compreensão dos fenômenos contemporâneos de acesso e uso da informação. Da mesma forma que os documentos, as unidades de informação precisam se transformar, acompanhar a sociedade hodierna e o novo paradigma produtivo emergente da digitalização da economia (FORESTI; VARVAKIS; GODOY VIERA, 2019). A seguir, são apresentadas as pesquisas mais pertinentes relacionadas ao tema da pesquisa, recuperadas em conformidade com os procedimentos metodológicos.

5 PESQUISAS RELACIONADAS

O Quadro 3 mostra os temas ligados ao *Linked Open Data* na base de dados Brapci e evidencia que o tema documentação é o mais pesquisado, seguido do assunto governos e *Open Government Partnership*. Jardim e Salomão Neto (2016) apontam as pesquisas nacionais sobre o tema do “governo aberto” como uma “embrionária literatura arquivística nacional”.

Quadro 3: Publicações recuperadas na Brapci com o tema *Linked Open Data* no título³.

| Temas | Quant. |
|--|--------|
| Acervos digitais, arquivos, documentos | 9 |
| Dados abertos governamentais, Governos | 6 |
| Bibliometria, Cientometria | 6 |
| E-science, Pesquisa, Ciência | 5 |
| Ontologias | 3 |
| Bibliotecas, catálogo | 2 |
| Avaliação, Uso | 2 |
| Catalogação, Descrição | 2 |
| Recuperação da Informação | 1 |
| Análise de redes sociais | 1 |
| Organização | 1 |
| Tesouro | 1 |
| Outros | 4 |

Fonte: Dados da pesquisa.

O Quadro 4 mostra os estudos selecionados. Marcondes (2016, p. 78-79) traz reflexões importantes sobre a questão Linked Open Data, mas não aponta caminho para realizar a abertura dos dados em unidades de informação e foca no aspecto teórico da semântica. Marcondes (2017) sugere espécie de guia para realizar a tarefa de publicar dados Linked Open Data, contudo, foca em bibliotecas e na aplicação da tecnologia Linked Open Data nos catálogos de bibliotecas.

Quadro 4: Publicações recuperadas na Brapci com o tema *linked open data* no título.

| Autor/Data | Título |
|------------------------------|---|
| Silva e Pinheiro (2018) | Métricas para dados governamentais abertos |
| Marcondes (2017) | Publicando e interligando acervos digitais na web através das tecnologias de dados abertos ligados |
| Monteiro e Jacynto (2016) | Emprego dos princípios <i>Linked Data</i> para gestão semântica de documentos digitalizados |
| Jardim e Salomão Neto (2016) | A perspectiva arquivística do governo aberto: um estudo comparativo da participação de Brasil, EUA, México e Reino Unido na OGP |
| Marcondes (2016) | Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos ligados |
| Segundo (2015) | Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. |

Fonte: Elaborado pelo autor.

Silva e Pinheiro (2018, p. 329) mostram diferentes métricas para avaliar os dados abertos e afirmam que as “métricas de avaliação” variam em conformidade com o modelo, muitas “recentes e ainda pouco utilizadas”. Os autores explicam que o esquema cinco

³ Quando o artigo cita arquivos, bibliotecas e museus, optou-se por classificar na seção de arquivo. O quadro aponta apenas um assunto para cada material.

estrelas e os oito princípios Linked Open Data são apenas o começo, que é preciso avaliar a abertura dos dados através de modelos diversos; destacam ainda a iniciativa OPENGOVB, “modelo” que lida com “indicadores relacionados ao governo aberto e que utiliza diferentes critérios de avaliação”.

Segundo (2015) apresentou trabalho importante no XVI ENANCIB (Encontro Nacional de Pesquisa em Ciência da Informação), no Grupo de Trabalho (GT 8) sobre os “os desafios do Brasil frente às iniciativas internacionais”. O autor afirma que no Brasil existe “tendência positiva” em relação à abertura dos dados, mas que ainda é algo embrionário, se comparada aos países desenvolvidos. Ainda segundo o autor, a “representatividade” do Brasil é “quase nula no Linked Open Data”, pois é usual a disseminação dos dados nas formas HTML e CSV, enquanto a questão semântica é inexistente ou “simplesmente” ignorada “pela grande maioria das bases de dados”.

O XVII ENANCIB no GT 5 sobre Política e Economia da Informação apresentou estudo importante para a proposta. Denominado “Perspectiva arquivística do governo aberto” de Jardim e Salomão Neto (2016), o estudo compara a participação de algumas nações na iniciativa Open Government Patnership. Os autores verificam a “presença” do tema arquivo nas discussões e planejamento das nações sobre dados abertos, e algo ainda mais importante: de que o tema Open Government Patnership marginaliza as unidades de informação (os arquivos, gestores da informação por excelência nas instituições públicas), ou seja, não existe o entendimento de que as bases de dados são documentos de arquivo. Mas isso acontece apenas nos países mais atrasados economicamente.

Jardim e Salomão Neto (2016) verificam que “o entendimento das relações entre governo aberto, dados abertos e arquivos pouco se expressa nos planos de ação da Parceria nos casos de Brasil e México”, enquanto no cenário das nações desenvolvidas (como Reino Unido e Estados Unidos) a situação se mostra diversa. Os resultados da pesquisa de Jardim e Salomão Neto (2016) corroboram e valoram grandemente o presente estudo e sua proposição: a chamada da responsabilidade dos arquivos públicos na publicação de dados abertos.

Nenhum estudo recuperado aborda a questão Linked Open Data nas instituições públicas federais brasileiras. Outros estudos são ofuscados por questões teóricas em demasia e outros ainda, são focados nos fluxos de abertura de dados sem nenhuma contextualização. Todos eles se abstêm do conhecimento da Ciência dos Arquivos. Verifica-se assim, que existe muita discussão sobre as potencialidades Linked Open Data

e pouca orientação voltada para aplicações práticas. Para preencher essa lacuna é apresentada a proposta que segue.

6 A PROPOSTA

A proposta mostrar, com fins didáticos e de modo teórico, como pode se dar a abertura de um conjunto de documentos específico: as Portarias produzidas pelo Conselho Regional de Engenharia e Agronomia de Santa Catarina (Crea-SC), por meio dos itens documentais. De forma que é realizada a abertura de uma unidade documental. O Quadro 5 mostra os conceitos e referências da proposta apresentada. Trata-se de junção entre legislação nacional, conceitos de gestão de documentos e informação e referências internacionais sobre a questão da publicação *Linked Open Data*.

Quadro 5: Referências centrais para a elaboração da proposta metodológica.

| | |
|---|---|
| Gestão de documentos (BRASIL, 1991) | A gestão de documentos e o fluxo da gestão de documentos são entendidos como uma série de etapas que transcorrem linearmente no tempo, da criação do documento até o seu uso, tramitação, avaliação, arquivamento, descarte e ou/preservação. |
| Conceito de Fluxo de informação de Barreto (2002) e Floridi (2010 <i>apud</i> Bittencourt e Isotani, 2018). | Abordagem que entende fluxo de informação como uma série de etapas envolvidas no processamento da informação, da origem ao usuário final até a reutilização dos dados. |
| Aspecto legal envolvendo essencialmente a Lei de Acesso à Informação (BRASIL, 2011) e o Decreto 8777 (BRASIL, 2016) | Abertura obrigatória de todas as informações de interesse social. Observar a questão da informação classificada ou sigilosa, da informação pessoal, do processamento, do acesso, da ubiquidade, confiabilidade, integridade e primariedade. |
| Princípios do <i>Open Government Data</i> (2018) | São princípios <i>Open Government Data</i> são a “completude”, “primariedade”, temporalidade, “acessibilidade”, processabilidade, “indiscriminatório”, “não proprietário”, “sem direitos autorais”. Os princípios adicionais são “estar em rede e ser de livre acesso”, “de forma permanente”, “confiável”, legal ou deve “cumprir a lei”, “ser documentado”, “seguro” e planejado para atender as necessidades do cidadão. |
| <i>5C functions</i> (conexão, <i>cloud</i> , customização, comunidade e conteúdo) para gerir informação sob o novo paradigma produtivo da quarta revolução industrial ou I4.0 (LEE et al. 2013) | É visão que norteia qualquer ação de gestão da informação na atualidade. Dá ênfase ao tempo real, ao acesso sem restrição espaço-temporal, bem como atenção devida às redes sociais. A questão da “correlação e do sentido” acontece através da devida conversão semântica. |
| Esquema das cinco estrelas (BERNERS-LEE, 2006) | Esquema clássico de boas práticas ligado à publicação <i>Linked Open Data</i> . Para alcançar as três estrelas é preciso compartilhar os documentos em formato CSV; para quatro estrelas é preciso realizar a conversão ou serialização RDF; para alcançar as cinco estrelas é preciso a ligação com os dados de outras instituições. |

| | |
|--|---|
| Os quatro princípios dos dados ligados (BERNERS-LEE, 2006) | São eles: se valer de URIs ⁴ em forma de nomes para os objetos, de URIs HTTP para o usuário consultar os objetos, nas consultas, prover informação RDF útil e por fim, se valer de sentenças RDF com links para outras URIs, para possibilitar a visualização de inter-relações. |
| Etapas de publicação <i>Linked Open Data</i> (BAUER; KALTENBOCK, 2012) | As etapas se constituem na “análise dos dados”, “limpeza”, “modelagem”, “escolha do vocabulário”, “especificação das licenças”, “conversão em RDF”, “vinculação a outros dados”, “publicação e promoção”. |

Fonte: Elaborado pelo autor.

Tendo em vista os resultados da revisão de literatura, se pode afirmar que se trata de uma proposta metodológica original para publicação de dados abertos em instituições públicas. O Quadro 6 mostra as etapas da aplicação da proposta. Não são tratadas questões sobre a extração dos dados ou reutilização, tema que exprime, necessariamente, outra abordagem. O foco da proposta é a abertura e a exposição dos dados abertos.

Quadro 6: Etapas da proposta metodológica.

| Etapa | Em que consiste | Breve explicação |
|--------------|--|---|
| 1 | Modelagem da publicação de dados numa empresa pública federal. | Identificação do fluxo de informação associado e visão geral do processo, identificação dos níveis administrativos e demais etapas envolvidas no processo. Integração dos fluxos de informação, de publicação <i>Linked Open Data</i> e gestão de documentos. |
| 2 | Estruturação e exposição dos dados em formato aberto | Para alcançar três estrelas em conformidade com o esquema cinco estrelas de Berners-Lee (2006), ou seja, abertura da base de dados nos formatos recomendados pela W3C. |
| 3 | Etapa de serialização ou conversão semântica | Eis o cerne da questão semântica. Neste momento são aplicadas as tecnologias associadas para dar sentido dos documentos e permitir sua leitura pelas máquinas. |

Fonte: Elaborado pelo autor.

A primeira etapa diz respeito ao mapeamento do fluxo das atividades relacionadas à publicação *Linked Open Data* numa instituição federal. Atenta-se aos três níveis administrativos, em conformidade com Chiavenato (1985). Segundo o autor, são três os níveis administrativos, o primeiro, “institucional”, diz respeito ao “planejamento estratégico”, são decisões de longo prazo que afetam a empresa em sua totalidade e que cabem à alta Direção; o nível “intermediário” refere ao “planejamento tático”, de médio prazo e envolve áreas específicas da organização (como um ou mais de um departamento,

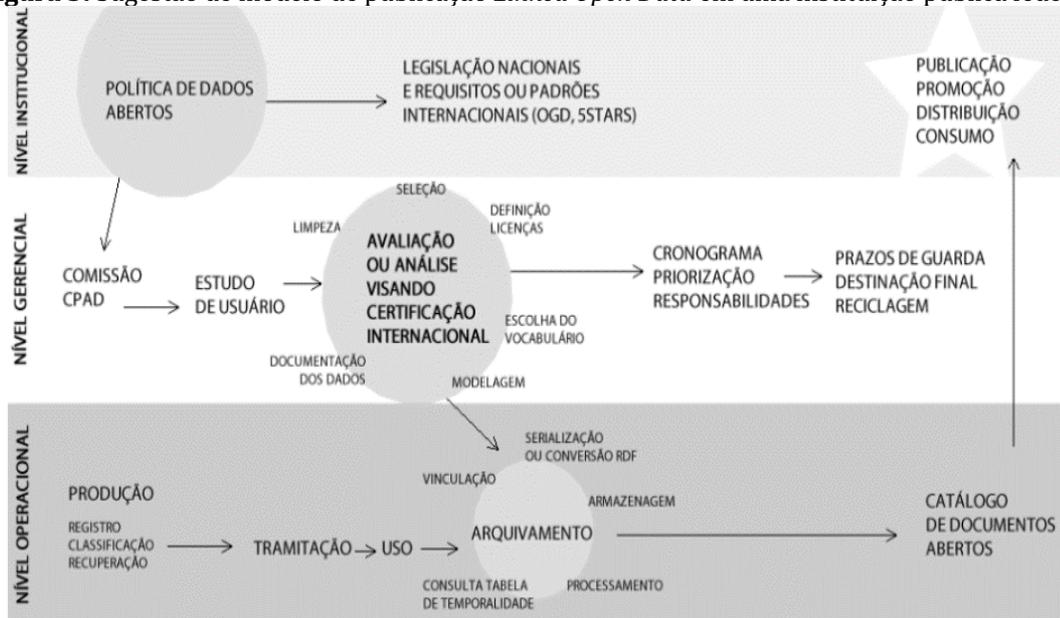
⁴ Uniform Resource Identifier.

por exemplo) e as decisões relacionadas dizem respeito à gerência; por fim, o nível “operacional” trata do “planejamento operacional” e é relacionado a tarefas específicas, de curto prazo, envolvendo supervisores e colaboradores em geral.

Em essência, a abordagem dos níveis administrativos busca elucidar o que compete à alta direção, à gerência e aos colaboradores. As etapas do fluxo ou do ciclo de abertura dos dados (BAUER; KALTENBOCK, 2012; JANSSEN et al., 2017) contemplam decisões e tarefas para todos os níveis administrativos. A Figura 3 apresenta as etapas da publicação dos dados abertos e as tarefas que competem a cada nível administrativo da organização.

A decisão de abertura dos dados e a própria Política de Dados Abertos é apontada pela Figura 3 como de nível institucional, e cabe a Diretoria da organização aprovar a política proposta. A alta direção precisa decidir acerca da abertura total ou mínima: a Lei de Acesso à Informação (BRASIL, 2011) aponta o que minimamente deve ser compartilhado, ao mesmo tempo, deixa claro que o sigilo é a exceção. Mas esse mínimo apontado pela legislação abre espaço para que o sigilo não seja tratado como exceção e se transfigure em regra. Uma das muitas incongruências legais do Brasil.

Figura 3: Sugestão de modelo de publicação *Linked Open Data* em uma instituição pública federal.



Fonte: Elaborada pelo autor.

A avaliação é a etapa que, talvez, seja a mais importante. É na avaliação que se decide que tipo de documento/conjunto de dados será compartilhado. Para tanto, é preciso tirar proveito de dois recursos existentes nas instituições públicas federais: a Tabela de

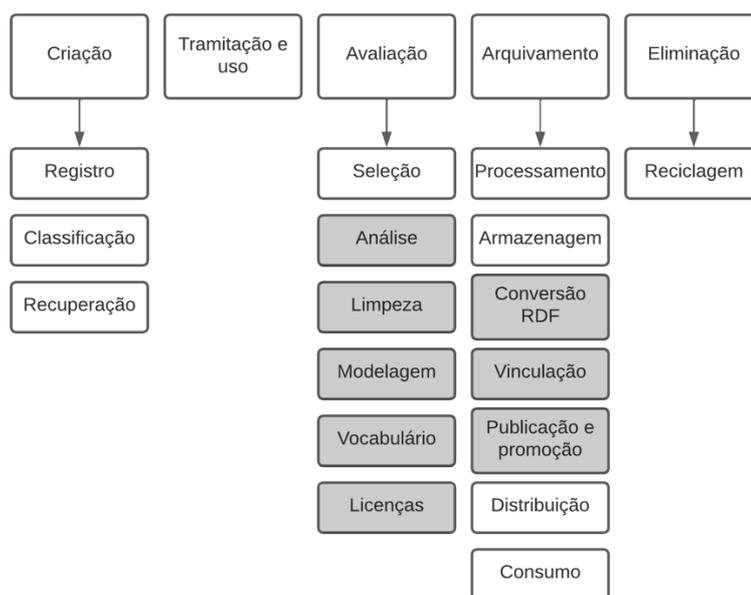
Temporalidade dos Documentos (BRASIL, 2020) e as Comissões Permanentes de Avaliação de Documentos.

O uso da Tabela de Temporalidade (BRASIL, 2020) como instrumento de abertura dos dados, exprime a participação ativa das Comissões Permanentes de Avaliação de Documentos, responsáveis pela aplicação da Tabela de Temporalidade da documentação (BRASIL, 2020) e dos descartes e preservação de documentos nas organizações públicas. Sugere-se que a referida comissão seja o ator central de etapas ligadas a publicação *Linked Open Data*, a exemplo da seleção dos dados que serão abertos, divulgação do cronograma de abertura, documentação dos dados, escolha dos vocabulários, entre outras etapas de nível gerencial.

Não menos importante é o que acontece em nível operacional. Momento em que é realizada a serialização RDF/XML. Nessa etapa é preciso capacitar os colaboradores e criar interface amigável, pré-programar os cabeçalhos de páginas internas e externas com códigos comuns a todos os documentos.

Em nível operacional, o documento é usado, movimentado, registrado, classificado e arquivado. Sugere-se que a serialização seja realizada no ato do arquivamento dos documentos, momento em que seriam inseridos e compartilhados no catálogo. Contudo, a serialização pode ser realizada no ato do registro, e assim, prover dados em tempo real.

Figura 4: O fluxo de gestão de documentos como elemento integrador da publicação de dados abertos conectados.



Fonte: Elaborada pelo autor com base nos conceitos apresentados no Quadro.

Esclarecido teoricamente o processo de publicação *Linked Open Data*, é oportuno contextualizar a questão no fluxo de gestão de documentos. A Figura 4 mostra a inter-relação existente entre os fluxos de informação, de publicação *Linked Open Data* e de gestão de documentos, sendo o último fluxo o elemento central, ou seja, a gestão de documentos se mostra como elemento agregador dos fluxos de informação e de publicação *Linked Open Data*.

O fluxo de Barreto (2002) complementa os demais por contemplar o “registro” e a “classificação” dos documentos, etapas importantes do processamento técnico. O fluxo de Floridi (2010 *apud* BITTENCOURT; ISOTANI, 2018) complementa os demais pela questão da “reciclagem” das informações advindas dos documentos. Mas é o fluxo de publicação *Linked Open Data*, evidenciado pelos quadros mais escuros da Figura 4, que diz respeito a real novidade e desafio para a gestão de documentos.

Até o momento, muitas etapas da proposta são associadas à avaliação e teoricamente já foram realizadas pela Comissão Permanente de Avaliação de Documentos, em conformidade com a metodologia proposta. A referida comissão decidiu acerca dos conjuntos de dados que devem ser abertos, e pode assim, expressar e documentar essa decisão na Tabela de Temporalidade (BRASIL, 2020), instrumento de consulta e referência da organização e da sociedade em geral. Dessa maneira, observa-se até o momento que a metodologia proposta aponta alguns preceitos importantes:

1. Abertura dos arquivos e bases de dados (o sigilo é a exceção);
2. As bases de dados são documentos de arquivo;
3. O tempo real é a meta de compartilhamento ideal;
4. A Tabela de Temporalidade é o instrumento de abertura dos dados;
5. A Comissão Permanente de Avaliação de Documentos é o espaço de discussão sobre quais os conjuntos de dados contidos nos documentos das empresas públicas federais devem ser abertos;
6. Os arquivos devem ser os vetores centrais da abertura dos dados.

A segunda etapa da proposta metodológica é resultado da Política de Dados Abertos da organização, das decisões da Comissão Permanente de Avaliação de Documentos e da aplicação da Tabela de Temporalidade (BRASIL, 2020) enquanto instrumento de abertura dos dados contidos nos documentos. Os conjuntos de dados

selecionados para abertura e compartilhamento são processados para que se alcancem as três estrelas de abertura dos dados, conforme esquema de Berners-Lee (2006).

Para tanto, é preciso estabelecer um Catálogo de Dados Abertos, usar licenças abertas e dispor os dados em formatos específicos (XML⁵ e CSV⁶). Ao mesmo tempo, as unidades arquivísticas são convertidas aos formatos RDF⁷/XML no ato do arquivamento, preferencialmente de forma automatizada.

O Apêndice I mostra como os dados são dispostos nos diferentes formatos ao usuário. Na prática, a conversão das bases de dados para os formatos XML e CSV é algo muito simples operacionalmente, afinal, basta exportar e compartilhar as bases ou o conjunto de documentos selecionados. O desafio reside na avaliação, na tomada de decisão e nos procedimentos associados de limpar, de documentar, entre outros.

O desafio maior da abertura dos dados reside na transformação semântica dos documentos por meio das chamadas triplas RDF. O fundamento do RDF segundo a W3C (2011b, p. 40) é a “tripla que define o sujeito, o predicado e o objeto utilizados para descrever uma informação na web e o relacionamento entre elas”, sendo o RDF definido da seguinte forma: um grupo de “dados e seus relacionamentos em forma de tripla”.

O Apêndice II mostra o modelo teórico usado para criar as triplas RDF e respectivas associações entre sujeito (a portaria 01/2018), predicado (com o uso dos vocabulários) e os objetos (advindos do conteúdo do documento). O Apêndice II apresenta a serialização RDF. O modelo de dados dessa serialização, em forma de triplas ou “*Triples of the Data Model*”, é apontado no Apêndice III.

Por meio do uso de vocabulários específicos e da serialização realizada na interface de programação do programa Dreamweaver - a título de exemplo - foi processada/serializada/convertida a Portaria nº 01/2018 para o formato RDF/XML. A sintaxe aplicada, conforme Figura 5 acima, retrata o esquema teórico apresentado no Apêndice II.

O Apêndice IV mostra como é possível ir além da informação do documento e enriquecer e ofertar mais contexto e sentido, por meio do uso das tecnologias semânticas. A partir da serialização inicial, é possível converter para os mais diversos formatos, usando a ferramenta RDF Translator⁸: microdata, N3, Ntriples, JSON.

⁵ Extensible Markup Language.

⁶ Comma Separated Values.

⁷ Resource Description Framework.

⁸ <http://rdf-translator.appspot.com/>.

O conjunto de tecnologias utilizadas na abertura dos dados foi explicitado anteriormente no Quadro 1. Trata-se de conjunto de tecnologias que permite estruturar e expor os dados produzidos pelo Crea-SC de forma aberta aos usuários, implicando os mais diversos conjuntos documentais e respectivas bases de dados.

São tecnologias que permitem realizar a conversão semântica e estabelecer um catálogo de dados abertos. Desta forma, através do conjunto de conceitos e tecnologias apresentadas, se pretende realizar a abertura dos dados no Crea-SC de forma ampla, em sintonia com a alta direção, com o corpo gerencial e demais colaboradores, envolvendo assim, todos os níveis administrativos no processo de abertura dos dados.

7 CONCLUSÕES

O estudo contribui com as pesquisas sobre o tema dos dados abertos e oferece aos gestores um caminho possível para lidar com a questão nas instituições públicas federais brasileiras. Os resultados evidenciam que a publicação de dados abertos nas instituições públicas passa necessariamente, por uma série de critérios legais, administrativos e técnicos que envolvem todos os níveis administrativos. Contudo, de forma especial pelas unidades de informação e pelos documentos de arquivo.

De modo que se verifica que os objetivos foram alcançados. A proposta, além de orientar a publicação de dados abertos no contexto público, realizou a abertura dos dados contidos em um documento produzido pelo Crea-SC. O estudo mostrou a estreita associação entre os fluxos de gestão de documentos, de informação e de publicação *Linked Open Data*, apontados como complementares entre si.

Os resultados mostram que é possível, natural, e viável, integrar a abertura dos dados à rotina dos arquivos, suas estruturas e instrumentos usuais, a exemplo das Comissões Permanentes de Avaliação de Documentos e do instrumento de destinação documental, a Tabela de Temporalidade. O estudo deixa claro que as bases de dados se constituem em documentos de arquivo advindos dos conjuntos documentais produzidos pelas organizações. Assim, afirma a possibilidade de atuação do profissional da informação e a importância dos arquivos na publicação de dados abertos.

O estudo também mostra que as pesquisas sobre *Linked Open Data* em Ciência da Informação são realidade incipiente. Contudo, embora a literatura utilizada corrobore a participação dos arquivos na abertura dos dados nas organizações, ela mostra que ainda não existe um protagonismo dos arquivos no Brasil - seja pela legislação, falta de atuação

profissional, falta de conhecimento e/ou baixa percepção de valor por parte dos gestores, ou simplesmente pela mera novidade do tema.

É importante salientar que o objetivo final da abertura dos dados em uma organização é permitir o uso dos dados por terceiros, seja pessoa física ou jurídica. Por isso, o mais importante é que as organizações atentem ao valor real que pode emergir do uso social dos dados públicos. Que atentem acerca da existência dos inúmeros benefícios envolvidos, inclusive financeiros.

As grandes organizações como o Crea-SC precisam realizar a abertura paulatina dos dados sob a sua custódia. Precisam divulgar a abertura dos dados por meio da realização de concursos que fomentem o uso dos dados abertos - por exemplo, o melhor aplicativo utilizando os dados sobre receituário agrônomo ou anotação de responsabilidade técnica. Os resultados podem ser imprevisíveis e surpreendentes, tais quais os impactos sociais, políticos e econômicos que emergem a abertura dos dados.

É preciso destacar ainda que a Lei Geral de Proteção de Dados pessoais (LGPD) ou a Lei n. 13709/2018 (BRASIL, 2018) que entrou em vigor no Brasil recentemente, no final do ano de 2020, não deve ser compreendida como um empecilho à abertura dos dados, ao contrário, é uma legislação que surge justamente para garantir a abertura dos dados públicos de forma ordeira, natural e segura.

Existem muitas questões em aberto e muitas limitações no presente estudo. Não foram elucidadas as questões acerca da estruturação do *web service*, da extração e do reuso dos dados, da criação da *triplestore*, da realização das pesquisas usando a linguagem SPARQL e dos procedimentos para a criação do catálogo de dados abertos utilizando o CKAN. São muitas as lacunas que precisam ser exploradas em estudos futuros. Estudos empíricos também se fazem necessários para testar os procedimentos explicitados na proposta.

REFERÊNCIAS

5 STAR Open Data. 2018. Disponível em: <http://5stardata.info/pt-BR/>. Acesso em: 12 mai. 2018.

AQUARONE, D. C. B.; LUZ, C. S. A conectividade e a organização da informação: uma abordagem entre a internet das coisas e a web semântica. **Ciência da Informação em Revista**, v. 4, n. 2, 2017.

BARRETO, Aldo de Albuquerque. O tempo e o espaço da Ciência da Informação. **Transinformação**, Campinas, vol.14, n.1, p.17-24. jan. /jun., 2002.

BAUER, Florian; KALTENBOCK, Martin. **Linked Open Data: a quick start guide for decision makers**. Austria: DGS, 2012, 62 p. Disponível em: https://semantic-web.com/wp-content/uploads/2017/05/LOD-the-Essentials_0.pdf. Acesso em: 22 mai. 2018.

BERNERS-LEE, T. HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, EUA, p. 29-37, May 2001. Disponível em: https://www.sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf. Acesso em: 08/05/2018.

BERNERS-LEE, T. **Linked Data**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 12/05/2018.

BITTENCOURT, I. I.; ISOTANI, S. **Curso online dados abertos conectados**. Escola de Políticas Públicas. Disponível em: <http://cursos.ep.org.br/>. Acesso em: 19/04/2018.

BIZER, C; HEATH, T; BERNERS-LEE, T. **Linked Data - The Story So Far**. 2009. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 02 ago. 2018.

BRASIL. Casa Civil. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei no 8.112, de 11 de dezembro de 1990; revoga a Lei no 11.111, de 5 de maio de 2005, e dispositivos da Lei no 8.159, de 8 de janeiro de 1991; e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 19/04/2018.

BRASIL. **Lei nº 8.159 de 8 de janeiro de 1991**. Dispõe sobre a política nacional de arquivos públicos e privados e dá outras providências. Disponível em: <http://www2.camara.leg.br/legin/fed/lei/1991/lei-8159-8-janeiro-1991-322180-publicacaooriginal-1-pl.html>. Acesso em: 11/05/2018.

BRASIL. **Decreto nº 8.777/2016**: 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: 22 mai. 2018.

BRASIL. Arquivo Nacional. **Código de classificação e tabela de temporalidade e destinação de documentos relativos às atividades-meio do Poder Executivo Federal**. Rio de Janeiro: Arquivo Nacional, 2020.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 17 jan. 2021.

CASTRO, A. de M. e; CASTRO, A. de M. e; GASPARIAN, D. de M. e C. **Arquivística - técnica, arquivologia - ciência**. Brasília, DF: ABDF, 1985- v.

CHIAVENATO, Idalberto. **Administração: teoria, processo e prática**. São Paulo: McGraw-Hill do Brasil, 1985.

CHIGNARD, S. **A Brief History of Open Data**. Paris Tech Review. 2013. Disponível em: <http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data>. Acesso em: 19/04/2018.

CKAN. 2018. Disponível em: <http://www.ckan.org/>. Acesso em: 02 ago. 2018.

DOCUMENTO. (verbete). HOUAISS, Antônio; Villar, Mauro de Salles. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2001, p. 1096.

FERMOSO, A. M. et al. Open linked data and mobile devices as e-tourism tools. A practical approach to collaborative e-learning. **Computers in Human Behavior**. v. 51, 2015. p. 618-626. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0747563215001363>. Acesso em: 02 ago. 2018.

FORESTI, F.; VARVAKIS, GODOY VIERA, A. F. A biblioteca e o novo paradigma produtivo da Indústria 4.0. **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v. 24, n. 3, p. 513-535, jul./out., 2019.

JANSSEN, M; HOVEN, J. V. D. Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? **Government Information Quarterly**. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0740624X15001069>. Acesso em: 02 ago. 2018.

JANSSEN, M; KUK, G. Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*. Disponível em: <https://www.tandfonline.com/doi/pdf/10.1080/10919392.2015.1124005?needAccess=true>. Acesso em: 02 ago. 2018.

JANSSEN, M et al. Driving public sector innovation using big and open linked data (BOLD). **Inf. Syst. Front**, v. 19. 2017. p. 189-195. Disponível em: <https://link.springer.com/article/10.1007/s10796-017-9746-2>. Acesso em: 02 ago. 2018.

JARDIM, J. M.; SALOMAO NETO, N. A. A perspectiva arquivística do governo aberto: um estudo comparativo da participação de Brasil, Estados Unidos, México e Reino Unido na Open Government Partnership. **Encontro Nacional de Pesquisa em Ciência da Informação**, v. 17, 2016.

JORENTE, M. J. V.; PADUA, M. C.; SEGUNDO, J. E. S. Criação de padrões na web semântica: perspectivas e desafios. **Em Questão**, v. 23, n. 3, p. 158-180, 2017. Disponível em: <http://www.brapci.inf.br/v/a/23125>. Acesso em: 08 mai. 2018.

LEE et al. Recent advances and trends in predictive manufacturing systems in big data environment. **Manufacturing Letters**, v. 1, p. 38-41, 2013.

MARCONDES, C. H. Publicando e interligando acervos digitais na web através das tecnologias de dados abertos interligados. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, 2017. Disponível em: <http://www.brapci.inf.br/v/a/28400>. Acesso em: 23 mai. 2018.

MARCONDES, Carlos Henrique. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 21, n. 2, p. 61-83. jun. 2016.

MONTEIRO, L. L. P; JACYNTO, M. D. de A. Emprego dos princípios Linked Data para gestão semântica de documentos digitalizados. **TransInformação**, Campinas, v. 28, n. 2, mai./ago. 2016. p. 241-251.

OPEN Knowledge International. **Open definition**. Disponível em: <https://opendefinition.org>. Acesso em: 19/04/2018.

OPEN Government Partenership. 2018. Disponível em: <https://www.opengovpartnership.org/>. Acesso em: 12 mai. 2018.

OPEN Goverment Data. 2018. Disponível em: <https://opengovdata.org/>. Acesso em: 12 mai. 2018.

PAES, M. L. **Arquivo**: teoria e prática. 3 ed. Rio de Janeiro: Fundação Getúlio Vargas, 1997.

SEGUNDO, J. E. S. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. **Encontro Nacional de Pesquisa em Ciência da Informação**, v. 16, 2015.

SILVA, P. N; PINHEIRO, M. M. K. Métricas Para Dados Governamentais Abertos. **Revista Ibero-americana de Ciência da Informação**. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/25726>. Acesso em: 02 ago. 2018.

W3C. 2018. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 12 mai. 2018.

W3C Brasil. **Publicação de dados em formato aberto**. Disponível em: <https://www.ep.org.br/dados-em-formato-aberto/>. Acesso em: 19/04/2018.

W3C Wold Web Wide Consortuim. Escritório Brasil. **Manual dos dados abertos**: governo. 2011. 54 p. Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf . Acesso em: 20/06/2018.

W3C Wold Web Wide Consortuim. Escritório Brasil. **Manual dos dados abertos: desenvolvedores**. 2011b. 146 p. Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/manual_dados_abertos_desenvolvedores_web.pdf. Acesso em: 20/06/2018.

ZIKOPOULOS, Paul. et al. **Understanding Big Data**: Analytics for Enterprise Class Hadoop and Streaming Data. New York: McGraw-Hill, 2012.

Recebido em: 16 de abril de 2022
Aprovado em: 14 de outubro de 2022
Publicado em: 22 de outubro de 2022

APÊNCIDE I: Dados da Portaria 01/2018 em XML e CSV.

Apresentação da informação em XML.

```
<portarias>
<id>9229</id>
<numero_portaria>1</numero_portaria>
<data>02/01/2018</data>
<ementa> Nomeia Coordenador do CDR - Colégio de Diretores Regionais
do Crea-SC</ementa>
<nome>Dourado, Álvaro Antônio Ribas;</nome>
<assunto>Nomeações; Coordenadores;</assunto>
<ano>2018</ano>
<link> http://www.crea-
sc.org.br/sites/ementario/portarias/2018/1.pdf?codigo=9229</link>
<cidade>Lages</cidade>
</portarias>
```

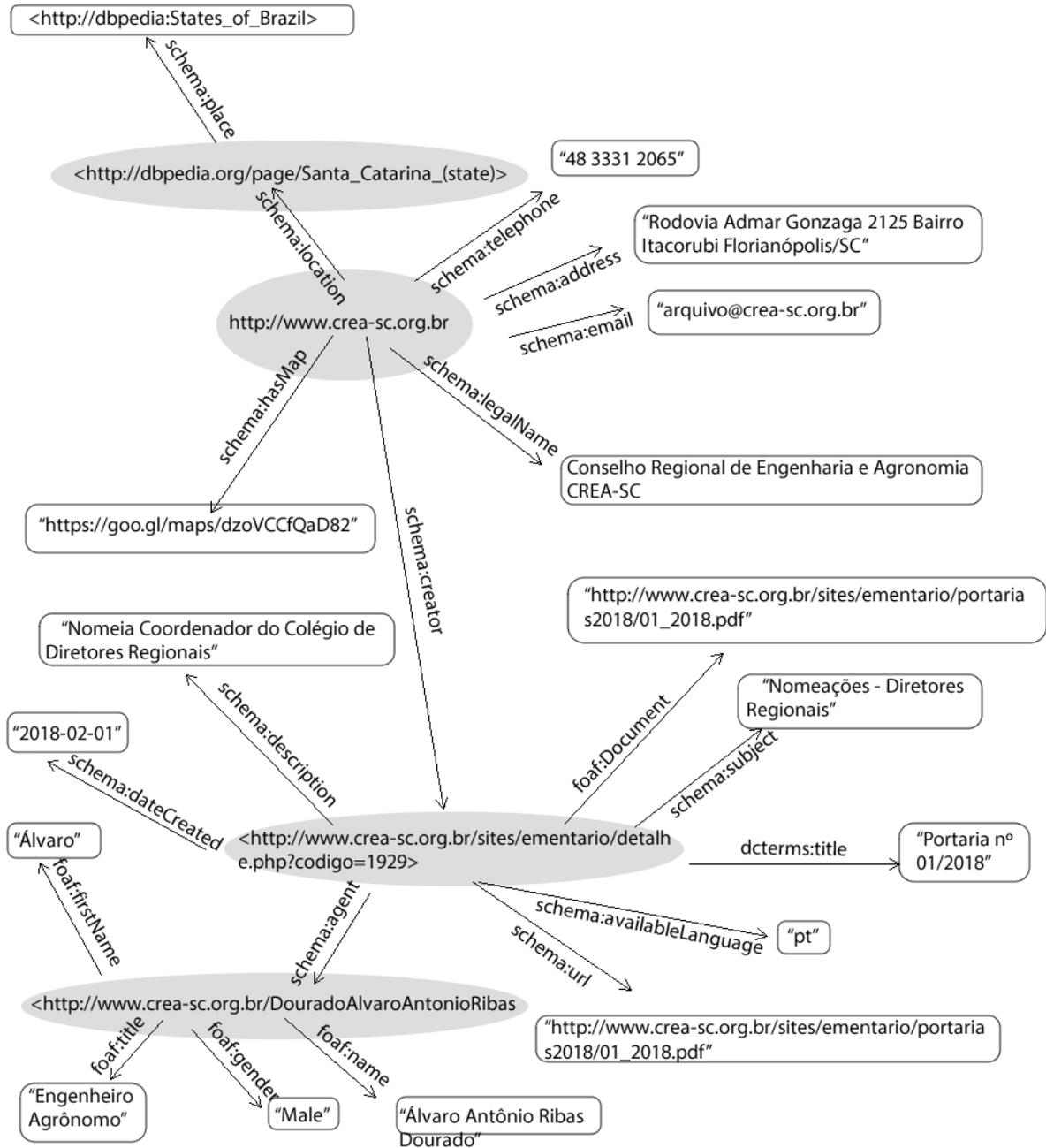
Fonte: Elaborado pelo autor.

Apresentação da informação em CSV

```
"9229";"1"; "02/01/2018"; "Nomeia Coordenador do CDR - Colégio de Diretores Regionais do Crea-SC";
"Dourado, Álvaro Antônio Ribas"; "Nomeações; Coordenadores;"; "2018"; "http://www.crea-
sc.org.br/sites/ementario/portarias/2018/1.pdf?codigo=9229"; "Lages";
```

Fonte: Elaborado pelo autor.

Apêndice II: Esquema teórico das Triplas RDF da Portaria 01/2018.



Fonte: Elaborado pelo autor.

APÊNDICE III: *Triples of the Data Model da Portaria 01/2018*

| N ^o | Subject | Predicate | Object |
|----------------|---|---|---|
| 1 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://www.w3.org/1999/02/22-rdf-syntax-ns#type | http://xmlns.com/foaf/0.1/Document |
| 2 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://dublincore.org/documents/2012/06/14/dcmi-terms/creator | http://www.crea-sc.org.br |
| 3 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://xmlns.com/foaf/0.1/agent | "Álvaro Antônio Ribas Dourado"@pt |
| 4 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://xmlns.com/foaf/0.1/img | http://www.crea-sc.org.br/sites/ementario/portarias2018/01_2018.pdf |
| 5 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://dublincore.org/documents/2012/06/14/dcmi-terms/subject | "Nomeações Coordenadores"@pt |
| 6 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://dublincore.org/documents/2012/06/14/dcmi-terms/title | "Portaria 01/2018"@pt |
| 7 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://dublincore.org/documents/2012/06/14/dcmi-terms/abstract | "Nomeia Coordenador do Colégio de Diretores Regionais"@pt |
| 8 | http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929 | http://dublincore.org/documents/2012/06/14/dcmi-terms/date | "2018-01-02"@pt |

Fonte: Elaborado pelo autor.

APÊNDICE IV: Serialização RDF da Portaria n. 01/2018 do CREA-SC.

```

<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:admin="http://webns.net/mvcb/"
  xmlns:dc="http://dublincore.org/documents/2012/06/14/dcmi-terms/"

  <foaf:Document xml:lang="pt" rdf:about="http://www.crea-sc.org.br/sites/ementario/detalhe.php?codigo=1929">
    <dc:creator rdf:resource="http://www.crea-sc.org.br"/>
    <foaf:agent>Álvaro Antonio Ribas Dourado</foaf:agent>
    <foaf:img rdf:resource="http://www.crea-sc.org.br/sites/ementario/portarias2018/01_2018.pdf"/>
    <dc:subject>Nomeações Coordenadores</dc:subject>
    <dc:title>Portaria 01/2018</dc:title>
    <dc:abstract>Nomeia Coordenador do Colégio de Diretores Regionais</dc:abstract>
    <dc:date>"2018-01-02"</dc:date>
  </foaf:Document>

  <foaf:Organization xml:lang="pt" rdf:about="http://www.crea-sc.org.br">
    <foaf:logo rdf:resource="http://www.crea-sc.org.br/sites/ementario/logo.jpg"/>
    <foaf:name>CREA-SC Conselho Regional de Engenharia e Agronomia de Santa Catarina</foaf:name>
    <foaf:homepage rdf:resource="http://www.crea-sc.org.br"/>
    <foaf:mbox> creasc@creasc.org.br</foaf:mbox>
    <foaf:based_near rdf:resource="https://goo.gVmaps/dzoVCCfQa082"/>
    <foaf:phone>4833312000</foaf:phone>
    <foaf:topic>Fiscalização</foaf:topic>
    <foaf:birthday>1958-17-03</foaf:birthday>
    <foaf:fundedBy>Celso Ramos Filho</foaf:fundedBy>
  </foaf:Organization>

  <foaf:Person xml:lang="pt" rdf:about="http://www.crea-sc.org.br/pessoas/DouradoAlvaroAntonioRibas">
    <foaf:name>Álvaro Antônio Ribas Dourado</foaf:name>
    <foaf:title>Engenheiro Agrônomo</foaf:title>
    <foaf:firstname>Álvaro</foaf:firstname>
    <foaf:family_name>Dourado</foaf:family_name>
    <foaf:birthday>xxxx-xx-xx</foaf:birthday>
    <foaf:Document>CREA-SC xxxxxxx</foaf:Document>
    <foaf:Group>xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx</foaf:Group>
    <foaf:gender>male</foaf:gender>
    <foaf:based_near>Caçador</foaf:based_near>
    <foaf:workplaceHomepage rdf:resource="http://www.xxxx.xx.xxx.br" />
    <foaf:schoolHomepage rdf:resource="http://www.xxx.br" />
    <dc:location rdf:resource="http://dbpedia.org/page/Santa_catarina_(state)"/>
    <foaf:phone>xxxxxxxx</foaf:phone>
    <rdf:type> schema:email="xxxxx@hotmailcom"</rdf:type>
    <rdf:type> schema:telephone="xxxxxxx" </rdf:type>
    <rdf:type> schema:email="xxxxx@xxxxxx.sc.gov.br"</rdf:type>
    <rdf:type> schema:postalCode="xxxxxxx"</rdf:type>
    <rdf:type> schema:address="Rua, Bairro, Cidade"</rdf:type>
    <rdf:type> schema:nationality="br"</rdf:type>
  </foaf:Person>
</rdf:RDF>

```

Fonte: Elaborado pelo autor.