

Apoio na adoção da Lei Geral de Proteção de Dados Pessoais por meio de anotações semânticas em descrições de serviços Web

Dornélio Mori Junior^I

<https://orcid.org/0009-0003-5895-0498>

Julio Cesar Nardi^{II}

<https://orcid.org/0009-0004-0644-2624>

Fabiano Borges Ruy^I

<https://orcid.org/0000-0003-2005-8288>

Giovany Frossard Teixeira^{II}

<https://orcid.org/0009-0004-8159-5313>

^I Instituto Federal do Espírito Santo, Serra, ES, Brasil;

^{II} Instituto Federal do Espírito Santo, Colatina, ES, Brasil

Resumo: Com a Lei Geral de Proteção de Dados Pessoais em vigor desde 2020 no Brasil, muitas organizações estão sendo levadas a implementar práticas de tratamento de dados. Isso tem sido foco de diferentes áreas do conhecimento como Ciência da Informação, Gestão da Informação e Tecnologia da Informação. Tais áreas tem buscado entender e incorporar os requisitos legais da Lei Geral de Proteção de Dados Pessoais no contexto dos processos organizacionais e como implementá-los por meio de sistemas de informação. Neste contexto, uma dificuldade enfrentada é a realização do tratamento de dados por meio de serviços web (p.ex., coleta, armazenamento, processamento, exclusão) disponibilizados e acessados por diversos sistemas de informação, inclusive entre distintas organizações. Identificar quais serviços realizam tratamento de dados pessoais e, com base nisso, realizar práticas em conformidade com a Lei Geral de Proteção de Dados Pessoais, é um aspecto fundamental. Neste trabalho, é proposta uma abordagem para enriquecimento semântico de descrição de serviços web REST usando anotações baseadas na OpenAPI e em uma ontologia de referência do domínio de Privacidade de Dados (OntoPrivacy). A partir dessa abordagem espera-se fornecer mais informações, por meio de meta-dados, sobre quais as operações de tratamento um serviço realiza e quais os tipos de dados sensíveis são manipulados. Em conjunto com a abordagem, foi desenvolvido um software (Privacy Finder) que dá suporte à localização automática das anotações nas APIs de serviços

descritas. A solução proposta foi aplicada à *Application Programming Interface* Pix do Banco Central do Brasil com forma de validação. Ademais, a solução foi analisada à luz de trabalhos correlatos existentes na literatura, assim como de produções técnicas cujos registros de programas de computador encontravam-se na base do Instituto Nacional da Propriedade Intelectual, o que evidenciou também uma tendência no crescimento de tais produções, principalmente, a partir do ano de publicação da referida lei.

Palavras-chave: privacidade de dados; LGPD; ontologia; anotação semântica; serviços web; meta-dados; gestão da informação; governança da informação

1 Introdução

Com a Lei Geral de Proteção de Dados Pessoais (LGPD) em vigor no Brasil desde setembro de 2020, muitas organizações necessitam implementar as práticas que a lei estabelece. A LGPD prevê que, independentemente do meio (digital ou físico), o tratamento de dados (coleta, reprodução, transmissão, processamento, armazenamento etc.) deve estar baseado em boas práticas para implementação dos requisitos legais (Brasil, 2018).

A LGPD tem sido foco de diferentes áreas do conhecimento, dentre elas: a Ciência da Informação (CI) e a Gestão da Informação (GI) (Melo; Batista; Lima, 2023; Oliveira; Gallotti; Barbosa Neto, 2024). Tais áreas tem buscado entender e incorporar os requisitos legais da LGPD no contexto dos processos organizacionais que lidam com informação sensível. Ao lado dessas áreas, encontra-se a área de Tecnologia da Informação (TI), que envolve os sistemas de informação que dão suporte aos processos organizacionais. Dessa forma, é desejável que haja um alinhamento entre os sistemas de informação e as demais áreas organizacionais que lidam com a informação sensível, de maneira que os requisitos da LGPD implementados nesses sistemas estejam em conformidade com políticas de gestão e governança da informação.

Com a ampla utilização de sistemas de informação para realização das mais diversas atividades humanas (compras, transações financeiras, entretenimento etc.) e a necessidade de integração desses sistemas, a tecnologia de serviços web ganhou destaque. Cada vez mais, organizações buscam fornecer serviços que sejam integráveis e que resulte em acesso/uso facilitado de seus sistemas e das informações neles contidas (Carey; Blevins; Takacs-Nagy, 2002

apud Tumer; Dogac; Toroslu, 2003; Tapsoba; Traore; Malo, 2023). Portanto, a publicação e o acesso à informação (em especial, informações pessoais) também se ampliou, gerando desafios como o de gerenciar conteúdo e o de garantir que as informações tenham uso adequado (Xavier, 2011; Santana, 2015).

Nesse contexto, a discussão sobre questões de privacidade no âmbito de serviços web que manipulam dados pessoais tem aumentado (Grünewald *et al.*, 2021; Hjerppe; Ruohonen; Leppänen, 2019; Netedu *et al.*, 2019; Peng; Bai, 2018; Cremaschi; Paoli, 2017). Ao acionar um serviço web, o consumidor desse serviço necessita fornecer um conjunto de informações requeridas, que pode incluir, por exemplo, dados pessoais (CPF, nome, identidade etc.). Quem consume o serviço precisa saber, por exemplo, como o provedor do serviço tratará essas informações a fim de decidir se, de fato, acionará o serviço. O provedor de serviço, por sua vez, necessita explicitar aspectos de privacidade de dados para dar segurança ao consumidor.

Assim, dentre os desafios que as organizações enfrentam ao aplicar a LGPD no âmbito de sistemas de informação estão: (i) Como identificar quais serviços web realizam tratamento de dados? (ii) Quais os tipos de dados pessoais (segundo a LGPD) são manipulados por esses serviços? E (iii) Quais operações de tratamento de dados pessoais (segundo a LGPD) são realizadas por cada serviço? Resolver questões como essas dá suporte a iniciativas de integração de sistemas de software (intra- e inter-organizacionais) permitindo, por exemplo, o alinhamento entre processos de negócio que fazem uso de dados pessoais. Assim, este trabalho apresenta uma abordagem para enriquecimento semântico de descrição de serviços web REST usando anotações baseadas na especificação OpenAPI v3 e em uma ontologia de referência desenvolvida para o domínio de Privacidade de Dados (OntoPrivacy). Além disso, é apresentada uma aplicação de software web (Privacy Finder) cujas funcionalidades tratam da identificação e da listagem de anotações em APIs (*Application Programming Interface*) de serviço feitas segundo a abordagem proposta.

A solução proposta foi aplicada, como forma de demonstração e validação, à API Pix do Banco Central do Brasil. Ademais, a abordagem foi discutida à luz de trabalhos correlatos da literatura, enquanto a aplicação web

desenvolvida foi inserida no contexto de uma análise dos registros de programas de computador existentes na base do INPI (Instituto Nacional da Propriedade Intelectual). Como resultado deste trabalho, espera-se contribuir tanto para o estado da arte, por meio da abordagem proposta, visando a facilitar a identificação de aspectos de privacidade de dados nas APIs de serviços web para o atendimento de requisitos legais, quanto no estado da técnica, por meio do desenvolvimento da aplicação de software.

O restante do texto é organizado da seguinte forma: a Seção 2 apresenta o referencial teórico, abordando aspectos de gestão e governança da informação, privacidade de dados, serviços web (REST), anotação semântica e ontologias; a Seção 3 apresenta a abordagem de anotação proposta juntamente com a aplicação de software desenvolvida; a Seção 4 apresenta a aplicação da abordagem proposta na descrição da API Pix do Banco Central do Brasil; a Seção 5 aborda trabalhos correlatos à luz da solução proposta; e a Seção 6 apresenta as considerações finais.

2 Referencial teórico

Nesta seção são apresentados os conceitos fundamentais que embasam as discussões construídas nas demais seções do trabalho.

2.1 Gestão da Informação e Privacidade de Dados

A Gestão da Informação (GI) tem suas bases em áreas clássicas da organização, gestão e utilização de documentos, tais como Biblioteconomia, Documentação e Ciência da Informação (Wilson, 2002¹, *apud* Monteiro; Duarte, 2018).

Na atual Sociedade orientada a grandes volumes de dados, a GI se coloca como uma área necessária para lidar com informações de forma adequada, eficaz e estratégica, envolvendo atividades fundamentais como organização, armazenamento, preservação, recuperação e acesso (Monteiro; Duarte, 2018). Tais atividades podem ser estruturadas em inúmeros modelos de gestão (Dutra; Barbosa, 2020). Assim, com a utilização de adequadas abordagens de gestão, a informação se transforma em um recurso valioso, permitindo a geração de conhecimento e inovação.

De maneira complementar à GI, a Governança da Informação (GoI) também lida com informações organizacionais e apoia os processos fundamentais da GI (organização, armazenamento, preservação, recuperação e acesso), mas por meio de uma outra perspectiva. A GoI pode ser definida como uma ferramenta de gestão que otimiza corporativamente o valor da informação (Silva; Estrela; Araújo, 2024). Mais especificamente, a GoI busca a especificar uma estrutura de responsabilidades e permissões visando à avaliação, criação, armazenamento, uso, arquivamento e exclusão da informação.

Neste contexto, a GoI lida com processos, papéis e políticas que asseguram o uso eficaz e eficiente da informação (Information Governance, c2024) e requer a colaboração de diversas áreas dentro de uma organização, incluindo Tecnologia da Informação (TI), *Compliance*, Auditoria e Gestão de Risco, dentre outras, além do compromisso da alta administração.

No atual modelo social de uso e compartilhamento de informação, potencializados por sistemas de informação integrados, aspectos de segurança de dados e informação tem ganhado grande destaque e permeado os processos e políticas de GI e de GoI. Assim, alguns instrumentos legais tem sido desenvolvidos na tentativa de responsabilizar e penalizar indivíduos que se apropriam, indevidamente, de materiais informativos (Melo; Batista; Lima, 2023). Neste cenário, destaca-se a Lei Geral de Proteção de Dados Pessoais (LGPD).

A LGPD, nº 13.709, foi publicada no Brasil em 2018 e passou a vigorar a partir de 18 de setembro de 2020 (Brasil, 2018; Melo; Batista; Lima, 2023). Tal lei tem seu texto foi inspirado na legislação europeia *General Data Protection Regulation* (GDPR) e aborda o tratamento de dados pessoais por pessoa natural ou pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e privacidade e o livre desenvolvimento da personalidade da pessoa natural (Brasil, 2018).

Privacidade pode ser definida como o direito de indivíduos para determinar, por si mesmos, quando, como e em que medida as informações sobre eles são comunicadas a terceiros (Westin, 1968).

A LGPD define operação de tratamento de dados pessoais como toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração (Brasil, 2018).

Tal lei estabelece, ainda, papéis como titular, controlador e operador (Brasil, 2018): **titular** é a “pessoa natural a quem se referem os dados pessoais que são objetos de tratamento”; **controlador** é uma “pessoa natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais”; e **operador** é uma “pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador”. Controladores e operadores (chamados agentes de tratamento) devem adotar medidas de segurança, técnicas e administrativas aptas a proteger os dados pessoais, sem as quais, correm o risco de sofrerem sanções.

Assim, a LGPD se coloca como um aparato legal importante para dar suporte à práticas de GoI (em especial, segurança da informação e *compliance*), que, por sua vez, se alinham aos processos fundamentais de GI organizacional. Entretanto, considerando que tanto a GoI quanto a GI são suportadas por sistemas computacionais (p.ex., sistemas de informação, sensores de dados, aplicações móveis, etc.), a implementação de práticas de proteção de dados pessoais integradas a tais sistemas se coloca ainda como um desafio.

2.2 Tecnologia de serviços Web

De acordo com o *World Wide Web Consortium* (W3C), um serviço web (web service) é um sistema projetado para que aplicações de software implementadas em diferentes tecnologias possam interagir (trocando dados e mensagens) (Booth *et al.*, 2004).

Um serviço web é realizado/implementado por uma aplicação de software em benefício de outras aplicações. Portanto, tal aplicação e o serviço realizado por ela são providos por e de responsabilidade de um agente intencional (pessoa física ou organização). Um serviço web é consumido por

aplicações de software clientes que o acessam a fim de obter informações. As aplicações clientes podem ser de diversas naturezas e de diferentes origens (organizações e pessoas) e, em geral, acessam serviços web para realizarem algum processo de negócio (compras online, transações financeiras, entretenimento etc.).

Nesse contexto, as descrições de serviços web são um elemento importante, pois fornecem informações para guiar a interação entre um provedor de serviço e os consumidores. Tais descrições abordam, em geral, como usar o serviço, quais os requisitos técnicos, as restrições, políticas e mecanismos de acesso, definição de mensagens, condições legais ou comerciais para a interação, concordância sobre os termos de uso, dentre outros (OASIS, 2006). Assim, o provedor de serviços deve ser capaz de disponibilizar detalhes do serviço para os potenciais consumidores. Tais consumidores devem ser capazes de tomar conhecimento dessas informações e decidir pela utilização do serviço. Para facilitar a compreensão e o uso, serviços web são tipicamente descritos e documentados usando as especificações padronizadas. Tais descrições, geralmente, são feitas pelos desenvolvedores dos serviços diretamente no código-fonte ou em arquivos de especificação em separado.

A especificação OpenAPI (também conhecida como Swagger, ou OAS – *OpenAPI Specification*) (OpenAPI Initiative, 2024) é um padrão proposto pela Open API Initiative (apoiada por empresas como Google, IBM, Microsoft, Oracle, PayPal) para criação de descrição de serviços web. Tal especificação define um formato padrão independente de linguagem de programação, o que permite que humanos e computadores descubram e compreendam os recursos disponibilizados por um serviço sem exigir acesso ao código-fonte. A Figura 1 mostra um fragmento de um documento de descrição de serviços em formato YAML² (*YAML Ain't Markup Language*) gerado a partir do uso de OpenAPI. Podem-se identificar, em vermelho, as *tags* (elementos descritivos) e seus respectivos valores.

A OpenAPI v3 oferece ainda a possibilidade de se ampliar o conjunto padrão de *tags* usando um mecanismo de extensão. Para tanto, novas *tags* a serem criadas devem ser padronizadas com o prefixo “x-” como, por exemplo,

“x-contato” e “x-fornecedor”. Assim, pode-se incluir novas informações na descrição de um serviço a partir de novas *tags* visando a enriquecer tal descrição tanto para humanos quanto para máquinas.

Figura 1 - Fragmento de um documento OpenAPI descrevendo um serviço web

```
25 - paths:
26 -   /pet:
27 -     put:
28 -       tags:
29 -         - pet
30 -       summary: Update an existing pet
31 -       description: Update an existing pet by Id
32 -       operationId: updatePet
33 -       requestBody:
34 -         description: Update an existent pet in the store
35 -         content:
36 -           application/json:
37 -             schema:
38 -               $ref: '#/components/schemas/Pet'
39 -           application/xml:
40 -             schema:
41 -               $ref: '#/components/schemas/Pet'
42 -           application/x-www-form-urlencoded:
43 -             schema:
44 -               $ref: '#/components/schemas/Pet'
45 -         required: true
46 -       responses:
47 -         '200':
48 -           description: Successful operation
```

Fonte: Swagger (2024).

2.3 Anotação Semântica e Ontologia

De acordo com Popov *et al.* (2003), a anotação semântica é um processo de geração de metadados específicos para possibilitar novos métodos de acesso à informação e estender os existentes. Sánchez-Fernández e Fernández-García (2005) referem-se à anotação semântica como o ato de anotar documentos com metadados. Metadado é um dado que carrega uma informação estruturada necessária para o entendimento do significado de outro dado contido em algum tipo de mídia.

Oren *et al.* (2006) consideram que as anotações podem ser realizadas de forma manual, semiautomática ou automática. A *anotação manual* pode ser realizada por uma ou mais pessoas. *Anotações semiautomáticas* são realizadas por pessoas com a assistência de ferramentas de software que fazem sugestões

automáticas. *Anotações automáticas* são criadas sem intervenção manual de pessoas, mas apenas por meio de software. Oren *et al.* (2006) afirmam ainda que a representações das anotações pode ser intrusivas ou não-intrusivas. As *intrusivas* são guardadas nos próprios documentos e a *não-intrusivas* são armazenadas em repositórios que apontam para os documentos que passaram pelo processo de anotação. Ademais, esses autores diferenciam três tipos de anotação: informal, formal e ontológica. *Anotações informais* não são lidas por máquina, pois não utilizam uma linguagem formal no processo. *Anotações formais* são interpretáveis por máquina, mas não usam termos de ontologias. *Anotações ontológicas* utilizam uma terminologia formalmente definida de um conceitualização compartilhada chamada “ontologia”.

Segundo Studer, Benjamins e Fensel (1998), uma ontologia, como um artefato, é uma especificação formal e explícita de uma conceitualização compartilhada. Nessa definição, “especificação explícita” diz respeito às definições de conceitos, instâncias, relações, restrições e axiomas; “formal” significa declarativamente definida, portanto, compreensível para pessoas e sistemas; “conceitualização” diz respeito a um modelo abstrato de parte do conhecimento; e “compartilhada” significa conhecimento consensual acerca de um domínio. Assim, ontologias são utilizadas como referência semântica para lidar com problemas de comunicação, alinhamento semântico, interoperabilidade, reuso e compartilhamento de métodos, paradigmas e linguagens (Uschold; Gruninger, 1996).

Ontologias tem sido aplicadas em diversas áreas do conhecimento como Ciência da Computação, Ciência da Informação, Filosofia, Inteligência Artificial, Linguística, Gestão de Conhecimento e Web Semântica, entre outras (Ontobras, 2024). Na Ciência da Informação, por exemplo, ontologias tem sido aplicadas como uma maneira promover o entendimento comum e a estruturação dos conceitos e relações do domínio de aplicação (Fonseca, 2007; Thomas, 2018; Smith, 2014). Assim, ela atuando como uma interlíngua entre pessoas e sistemas computacionais integrados (Nardi; Falbo; Almeida, 2013; Guizzardi, 2007). Como resultado, ontologias, podem dar suporte a atividades de captura, tratamento e uso de informações, que são, em geral, realizadas por sistemas

computacionais. Dessa forma, diz-se que ontologias são um importante artefato no apoio ao desenvolvimento de sistemas de computacionais (*ontology-driven information systems*) (Guarino, 1998; Fonseca, 2007).

3 A Abordagem proposta e seus elementos

Nesta seção, é apresentada a abordagem proposta para enriquecimento semântico de descrição de serviços web REST usando anotações baseadas na especificação OpenAPI v3. Ademais, é apresentada a ontologia OntoPrivacy, cujos conceitos são utilizados para realização das anotações. Os seguintes requisitos de *design* foram definidos:

- a) **RQ1** - a abordagem deve ser *lightweight*, permitindo anotação semântica em descrições de serviços web sem a necessidade de complexas tecnologias semânticas evitando a chamada “*semaphobia*”³ (Lanthaler; Gütl, 2011);
- b) **RQ2** - a abordagem deve fazer uso do mecanismo de extensão de *tags* da OpenAPI para criação dos metadados relativos à LGPD;
- c) **RQ3** - a abordagem deve permitir anotações semânticas formais e ontológicas, isto é, passíveis de serem processadas por máquinas e baseadas em terminologia de conceitualização compartilhada;
- d) **RQ4** - abordagem deve ser baseada em anotação manual, podendo ser realizada tanto por desenvolvedores quanto por outros atores (p.ex., jurídico, *compliance*);
- e) **RQ5** - a abordagem deve permitir tanto anotações intrusivas (implementadas diretamente no código-fonte do serviço) quanto não-intrusivas (implementadas em separado nos arquivos de descrição de serviços armazenados em repositórios próprios).

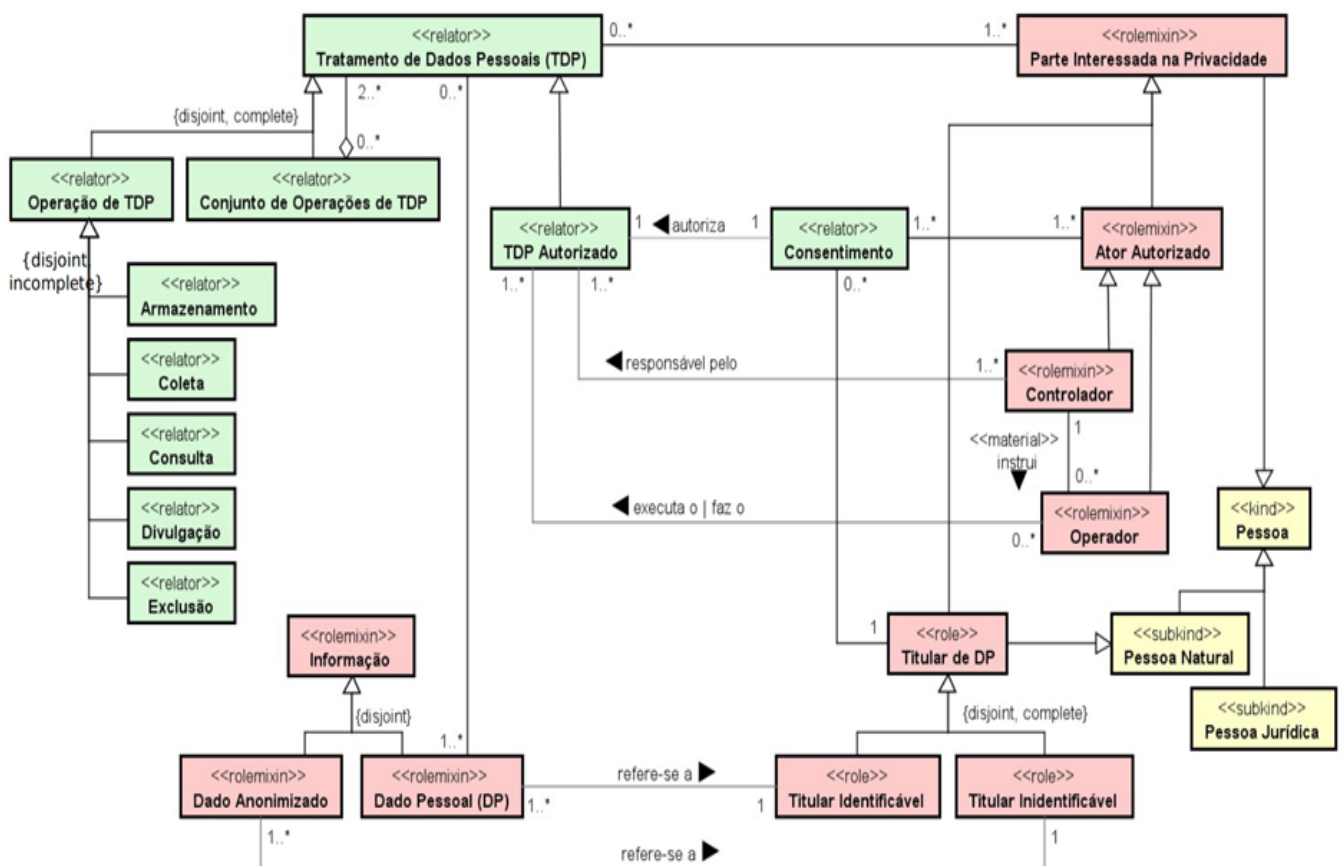
3.1 Uma Ontologia de referência para privacidade de dados

OntoPrivacy é uma ontologia do domínio de Privacidade de Dados desenvolvida com base na norma NBR ISO/IEC 29100 (Associação Brasileira de Normas Técnicas, 2020), na Lei nº 13.709/2018 – Lei Geral de Proteção de Dados Pessoais (LGPD) (Brasil, 2018) e em modelos relacionados, tais como (El

Ghosh; Abdulrab, 2021; Gharib; Giorgini; Mylopoulos, 2021; Wu *et al.*, 2021; Palmirani *et al.*, 2018; Pandit *et al.*, 2019).

Visando a maior expressividade e rigor na representação, OntoPrivacy é fundamentada em UFO (*Unified Foundation Ontology*) (Guizzardi, 2005) e especificada em OntoUML. UFO é uma ontologia de fundamentação que define distinções ontológicas úteis para compreender e representar um determinado domínio de interesse, tais como sortais rígidos (*kind*) e anti-rígidos (*role* e *phase*), mediadores em relações materiais (*relator*), além de não-sortais para generalizações (*category* e *rolemixin*). OntoUML captura tais distinções em uma extensão da UML (*Unified Modeling Language*), e tem sido usada na construção de modelos ontológicos em diversos domínios. A Figura 2 apresenta um fragmento de OntoPrivacy focado no tratamento de dados pessoais e atores envolvidos.

Figura 2 - Fragmento da Ontologia de Privacidade (OntoPrivacy)



Fonte: Elaborado pelos autores.

Segundo o fragmento da ontologia, **Dado Pessoal (DP)** é qualquer **Informação** relacionada a uma **Pessoa Natural** que desempenha o papel (*role*) de **Titular Identificável** desse dado. **Titular Inidentificável** é uma **Pessoa Natural** que não pode ser identificado a partir de dados pessoais por estes estarem anonimizados.

Operações de **Tratamento de Dados Pessoais (TDP)** podem ocorrer de forma simples/isolada (**Operação de TDP**), ou combinadas (**Conjunto de Operações de TDP**). As operações podem ser dos tipos **Armazenamento, Coleta, Consulta, Divulgação e Exclusão**, entre outros.

O tratamento de dados é realizado por um ou mais **Atores Autorizados**, sendo o **Controlador** responsável por tratar os dados, podendo ou não instruir **Operadores** a executar tal tratamento. O **Tratamento de Dados Pessoais (TDP)** envolve as **Partes Interessadas na Privacidade**. Um tratamento de dados autorizado (**TDP Autorizado**) surge apenas a partir de consentimento do titular dos dados. O **Consentimento** estabelece que o **Titular de DP** autorizou um ou mais atores (**Ator Autorizado**) a tratar um conjunto de seus dados. Por exemplo, João (**Pessoa Natural**) pode preencher seu nome, e-mail e endereço (**Dados Pessoais**) em um site de comércio eletrônico e fornecer um **Consentimento** autorizando o responsável pelo site (**Controlador**) a realizar **Operações de TDP de Armazenamento, Coleta e Consulta** sobre seus dados, mas não de **Divulgação**.

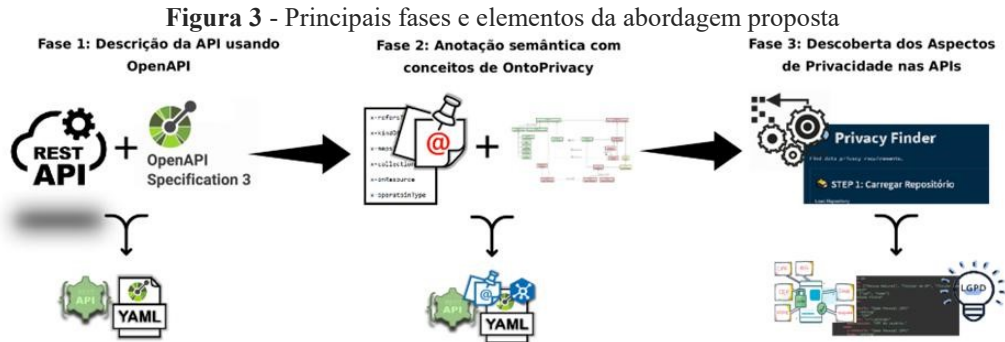
Os conceitos e relações estabelecidos pela ontologia são utilizados no contexto da abordagem proposta para realizar marcações semânticas nas descrições dos serviços web. Mais especificamente, serão inseridos nas descrições a fim de buscar ampliar a caracterização dos dados e operações realizadas por cada serviço web, conforme detalha a seção a seguir.

3.2 Fases da abordagem proposta: da descrição à descoberta

A Figura 3 ilustra os principais elementos que compõem a abordagem e como esses elementos se articulam ao longo de três fases, a saber:

- a) **Fase 1** - descrição da API de serviço usando OpenAPI;

- b) **Fase 2** - anotação semântica da API com conceitos de OntoPrivacy;
- c) **Fase 3** - descoberta dos aspectos de privacidade anotados em descrições de APIs.



Fonte: Elaborado pelos autores.

Como resultado da Fase 1, ao descrever a APIs de serviço usando-se a especificação OpenAPI, tem-se como resultado descrição(ões) em .YAML, as quais abordam as operações dos serviços, os parâmetros de entrada e saída, dentre outros. Em seguida, passa-se para a Fase 2, quando essas descrições são enriquecidas com anotações semânticas. Tais anotações são feitas utilizando-se as propriedades definidas por Karavisileiou *et al.* (2021) como mecanismos de extensão da OpenAPI. Tais propriedades referenciam conceitos e relações da ontologia OntoPrivacy. O Quadro 1 sumariza as propriedades. Como resultado da Fase 2, tem-se, então, arquivo(s) .YAML já anotado semanticamente.

Quadro 1 - Propriedades de extensão adotadas

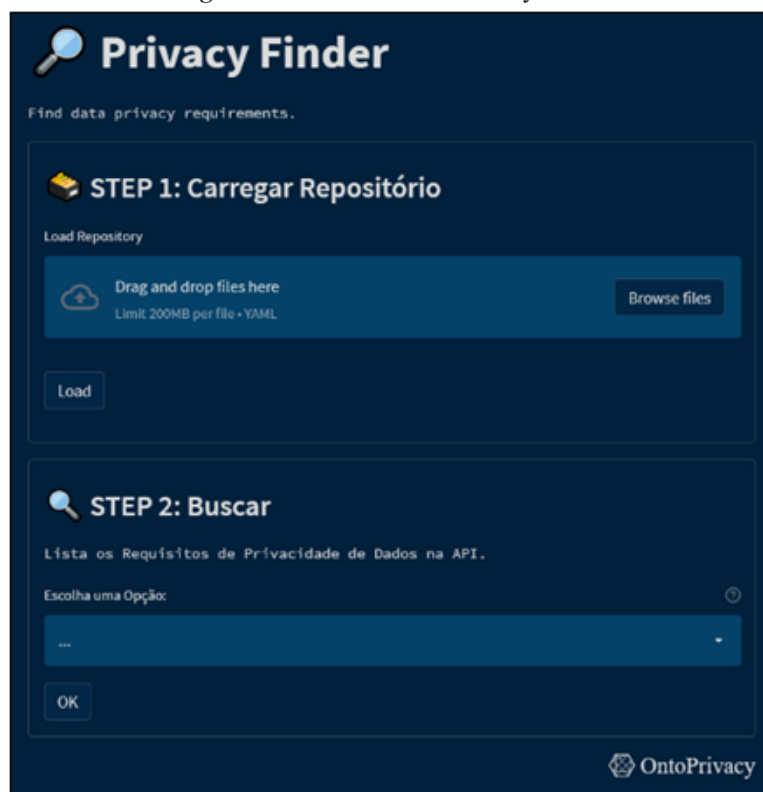
Propriedade	Aplica-se a	Descrição	Exemplo
x-refersTo	Schema	Relaciona um elemento da descrição de serviço a um conceito da ontologia.	x-refersTo: "Dado Pessoal (DP)"
x-kindOf	Schema	Especialização entre um elemento da descrição de serviço (<i>subkind</i>) e um conceito da ontologia (<i>superkind</i>). Esta relação é mais restrita que x-refersTo. É usada quando não há uma equivalência direta entre elementos.	x-kindOf: "Operador"
x-collectionOn	Schema	Relaciona uma coleção de elementos a um conceito da ontologia, que representa os tipos dos itens da coleção.	x-collectionOf: "Pessoa"
x-operationType	Operation	Relaciona uma determinada operação da descrição de serviço às operações de tratamento de dados de OntoPrivacy.	x-operationType: "Coleta"

Fonte: Baseado em Karavisileiou *et al.* (2021).

Na prática, as anotações realizadas nas fases um e dois podem ser feitas tanto pela equipe de desenvolvimento que implementam as APIs dos serviços quanto por outros departamentos (p.ex., jurídico e/ou *compliance*) envolvidos em operações de tratamento de dados pessoais.

Uma vez anotada semanticamente a API, a Fase 3 trata da descoberta / localização de aspectos de privacidade de dados incorporados na descrição da API por meio das anotações. A descoberta/localização pode ser realizada para verificar, por exemplo, quais operações dos serviços fazem uso de dados sensíveis, quais dados sensíveis são utilizados ou mesmo quais operações de tratamento são realizadas sobre esses dados. Essa fase pode ser realizada tanto por membros da organização que disponibiliza os serviços web, quanto por membros de equipes externas à organização e que estão interessados em acessar os serviços e, portanto, em entender melhor o tratamento de dados realizado antes de vir a utilizá-los.

Esta fase pode ser realizada de maneira manual, mas o uso de ferramentas computacionais facilita bastante esse trabalho. Assim, com o propósito de facilitar a utilização da abordagem proposta, foi desenvolvida, no contexto deste trabalho, uma ferramenta de software web, chamada Privacy Finder (Privacy Finder, 2021). Seu objetivo é apoiar o usuário (desenvolvedores, *compliance* etc.) na tarefa de descobrir/localizar anotações semânticas criadas por meio da abordagem, apresentando-as de forma mais facilitada. A Figura 4 apresenta a tela inicial de Privacy Finder. Seu uso se dá em dois passos: “Carregar Repositório” e “Buscar”.

Figura 4 - Tela inicial de Privacy Finder

Fonte: Elaborado pelos autores.

“Carregar Repositório” permite ao usuário carregar para a ferramenta especificações OpenAPI (arquivos .YAML) nas quais ele se deseja realizar buscas/varreduras pelas anotações semânticas. “Buscar” permite que o usuário escolha entre três opções de busca:

“CONCEITO - Busca as Anotações da API por Conceitos”: realiza varredura sintática nas descrições de serviço carregadas no repositório buscando um conceito indicado pelo usuário. Os conceitos passíveis de seleção pelo usuário são aqueles de OntoPrivacy (p.ex., “Dado Pessoal (DP)”, “Titular de DP” e “Autor Autorizado”). Assim, retorna todas as ocorrências de um dado conceito indicado pelo usuário, listando as anotações em cada descrição .YAML.

“ALL - Carrega todas as Anotações da API”: realiza uma varredura em todas as descrições de API carregadas no repositório buscando por anotações criadas por meio da abordagem. Como resultado, retorna uma lista de anotações por descrição .YAML.

“VIEW - Visualiza a(s) API(s) carregadas”: lista, na íntegra, as descrições de todas as APIs carregadas no repositório sem realizar uma busca específica.

4 Aplicação da abordagem na API PIX do Banco Central

Como forma de validar a abordagem proposta, esta foi aplicada no contexto da API Pix disponibilizada pelo Banco Central do Brasil (BC). Tal API padroniza o acesso aos serviços web oferecidos pelo Provedor de Serviço de Pagamento (PSP) do BC permitindo acesso a operações como: criação de cobrança, verificação de Pix recebidos, devoluções e consultas. Os serviços disponibilizados pelo PSP possibilitam que outras organizações, por meio de seus sistemas, acessem softwares do BC.

4.1 Anotando semanticamente a API PIX do BC

Ao analisar o documento OpenAPI da API Pix (v2.6.1) do BC pode-se identificar que o serviço oferecido faz uso de diferentes tipos de informação, dentre elas, informações de dados pessoais (p.ex., nome e CPF). Entretanto, identificar os pontos da API Pix em que se dá algum tipo de tratamento de dados pessoais não é uma tarefa fácil. A Figura 5 apresenta um fragmento da API Pix cujo foco é a descrição do schema “PessoaFisica”, em que constam informações como “cpf” e “nome”. Pode-se notar que não há detalhamentos explícitos sobre aspectos de privacidade.

Figura 5 - Fragmento 1: schema “PessoaFisica” da API Pix (v2.6.1)

```
openapi: 3.0.0
info:
  title: API Pix
  version: "2.6.1"
  license: 
  contact: 
  description: |
  servers: 
  tags: 
  paths: 
  components:
    securitySchemes: 
    examples: 
    requestBodies: 
    schemas:
      PessoaFisica:
        type: "object"
        required: ["cpf", "nome"]
        title: "Pessoa Física"
        properties:
          cpf:
            type: "string"
            title: "CPF"
            pattern: "^\\d{11}$/"
            description: "CPF do usuário."
          nome:
            type: "string"
            title: "Nome"
            description: "Nome do usuário."
            maxLength: 200
```

Fonte: Banco Central (2024).

No caso de descrições de operações, a situação fica mais evidente. A Figura 6 apresenta outro fragmento da API, no qual se tem uma operação HTTP POST “/cob”, que usa informações de dados pessoais, mas não é possível identificar explicitamente aspectos sobre tratamento de dados. Ademais, o schema de dados utilizado na operação (\$ref:”#/components/requestBodies/CobBody”) é composto por uma cadeia de *schemas*, sendo um destes o schema “PessoaFisica”. Isso não fica explícito ao se ler diretamente a definição da operação

Figura 6 - Fragmento 2: operações HTTP com dados pessoais da API Pix (v2.6.1)

```
openapi: 3.0.0
info:
  title: API Pix
  version: "2.6.1"
  license:
  contact:
  description: |-
servers:
tags:
paths:
  "/cob":
    post:
      tags:
        - "Cob"
      summary: "Criar cobrança imediata."
      security:
        - OAuth2: [cob.write]
      description: |-
      requestBody:
        $ref: "#/components/requestBodies/CobBody"
      responses:
        "201":
          description: "Cobrança imediata criada"
          content:
            "application/json":
              schema:
                $ref: "#/components/schemas/CobGerada"
              examples:
        "400":
        "403":
        "503":
```

Fonte: Banco Central (2024).

Ao adotar a abordagem proposta e anotar semanticamente a descrição da API Pix com conceitos de OntoPrivacy, pode-se fornecer informações sobre aspectos de privacidade de dados, e, com isso, dar maior suporte à utilização da API à luz da LGPD. As figuras 7 e 8 replicam, respectivamente, os fragmentos da descrição da API Pix ora tratados nas figuras 5 e 6, agora, com as anotações semânticas da abordagem proposta.

Na Figura 7 pode-se observar o schema “PessoaFisica” anotado com conceitos da ontologia indicando que se refere a (“x-refers-To”) **Pessoa Natural, Titular de DP e Titular Identificável**. Os elementos “cpf” e “nome” também foram identificados como **Dado Pessoal (DP)**.

Figura 7 - Fragmento 1 anotado: schema “PessoaFisica” da API Pix (v2.6.1)

```
openapi: 3.0.0
info:
  title: API Pix
  version: "2.6.1"
  license:
  contact:
  description: "API Pix."
servers:
tags:
paths:
  ## ANOTADO ##
  "/cob":
    ## ANOTADO ##
    post:
      x-operationType: ["Operacao de TDP", "TDP Autorizado", "Coleta",
      tags:
        - "Cob"
      summary: "Criar cobrança imediata."
      security:
        - OAuth2: [cob.write]
      description:
      requestBody:
        $ref: "#/components/requestBodies/CobBody"
      responses:
        "201":
          description: "Cobrança imediata criada."
          content:
            application/json:
              schema:
                $ref: "#/components/schemas/CobGerada"
              examples:
        "400":
        "403":
        "503":
```

Fonte: Elaborado pelos autores.

Na Figura 8 observa-se a anotação na operação “/cob”. Com a anotação pode-se identificar que a operação é dos seguintes tipos (“x-operationType”) de tratamento de dados: **TDP Autorizado** que faz a **Coleta** e **Armazenamento**.

Figura 8 - Fragmento 2 anotado: operações HTTP com dados pessoais da API Pix (v2.6.1)

```

title: API Pix
version: "2.6.1"
> license:
> contact:
> description: "API Pix."
> servers:
> tags:
> paths:
> components:
> securitySchemes:
> examples:
> requestBodies:
## ANOTADO ##
> schemas:
  ## ANOTADO ##
  PessoaFisica:
    x-refersTo: ["Pessoa Natural", "Titular de DP",
    type: "object"
    required: ["cpf", "nome"]
    title: "Pessoa Física"
    properties:
      cpf:
        x-refersTo: "Dado Pessoal (DP)"
        type: "string"
        title: "CPF"
        pattern: "/^\d{11}$/"
        description: "CPF do usuário."
      nome:
        x-refersTo: "Dado Pessoal (DP)"
        type: "string"
        title: "Nome"
        description: "Nome do usuário."
        maxLength: 200

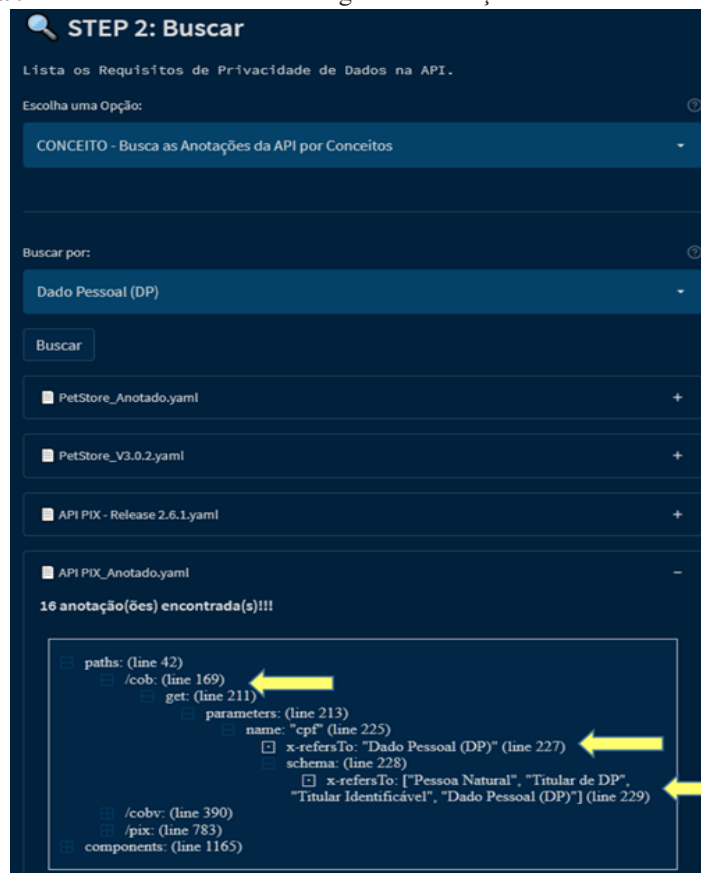
```

Fonte: Elaborado pelos autores.

4.2 Localizando anotações na API PIX do BC

Considere, pois, o uso da ferramenta de software Privacy Finder sobre a descrição da API Pix do BC já anotada por meio da abordagem proposta. Ao se utilizar a opção “CONCEITO - Busca as Anotações da API por Conceitos” informando, por exemplo, como conceito “**Dado Pessoal (DP)**”, a ferramenta faz uma varredura sintática na descrição de serviço da API Pix do BC buscando o conceito indicado. Como resultado, conforme exibido pela Figura 9, foram localizadas referências ao conceito **Dado Pessoal (DP)** nas linhas 227 e 229 da especificação da API (em destaque). Pode-se observar, então, na listagem que a operação “/cob” (linha 169) da API faz o uso do dado “cpf”, referenciado como “**Dado Pessoal (DP)**”, e de um schema de dados (linha 228), que foi caracterizado pelas anotações “**Pessoa Natural**”, “**Titular de DP**” e “**Dado Pessoal (DP)**”, dentre outras. Isso permite ao usuário uma análise, por exemplo, dos aspectos de privacidade de dados inerentes à operação.

Figura 9 - Resultados de busca e listagem de anotações na API Pix BC anotada



Fonte: Elaborado pelos autores.

5 Trabalho correlatos

Podem-se encontrar na literatura propostas correlatas à abordagem apresentada neste trabalho, assim como, produções técnicas, mais especificamente softwares, cujo propósito é dar suporte à implantação da LGPD. Portanto, são apresentados e discutidos trabalhos tanto da literatura quanto produções técnicas correlatas.

Diversos trabalhos na literatura podem ser apresentados, os quais abordam soluções para descrição semântica de serviços web, como OWL-S (Martin *et al.*, 2004), SAWSDL (Kopecký *et al.*, 2007), WSMO-Lite (Vitvar *et al.*, 2008), SA-REST (Gomadam; Ranabahu; Sheth, 2010), RESTdesc e Hydra (Verborgh *et al.*, 2011; Lanthaler; Gütl, 2013; Peng; Bai, 2018). Tais soluções apresentam em comum o fato de requererem amplo conhecimento sobre tecnologias semânticas, o que causa relutância na aplicação, levando ao que é chamado por Lanthaler e Gütl (2011) de “semaphobia”. Assim, abordagens

alternativas mais leves (*lightweight approaches*) tem sido desenvolvidas e defendidas para dar suporte à anotação semântica de serviço como SemREST (Peng; Bai, 2018), Hydra (Lanthaler; Gütl, 2013), SEREDAS (Lanthaler; Gütl, 2012). A abordagem proposta neste trabalho, pois, se alinha às *lightweight approaches*, visando a uma maior aplicabilidade.

Em (Peng; Bai, 2018) é proposto um método, SemREST (*Semantic Resource Tagging*), para adicionar anotações semânticas à descrição do serviço REST a fim de tornar a interação entre o serviço e seu cliente mais genérica e autônoma. A abordagem visa a evitar a necessidade de conhecimento especializado sobre a Web Semântica (e suas tecnologias), adicionando anotações semânticas baseadas em tags à descrição OpenAPI do serviço. Embora foque no uso de tags para a realizar as anotações, vale destacar que o uso de tags é limitado, em OpenAPI, ao objeto *Operation*. As *tags* podem ser usadas para agrupamento lógico de operações. Nossa abordagem utiliza o mecanismo de extensão da OpenAPI (^x-), que pode ser utilizado em/por toda a OpenAPI, não se limitando aos elementos do tipo *Operation*.

Em (Hjerppe; Ruohonen; Leppänen, 2019) é proposta a utilização de três anotações para a análise de código-fonte. Em essência, a anotação @PersonalData (A1) deve ser usada para documentar todas as classes que contenham dados pessoais. Essa anotação dá o contexto necessário para separar os tipos de dados pessoais de outros tipos, mas não muito mais além disso. Também serve como documentação do código fonte. As outras duas anotações, @PersonalDataHandler (A2) e @PersonalDataEndpoint (A3), são usadas para ampliar a documentação da classe onde os dados pessoais são processados. Essa proposta aborda a utilização de anotações, porém de forma rígida (limitando-se ao código-fonte) e com pouca diversidade de anotações, o que difere da abordagem proposta neste trabalho, que tende a ser mais flexível (por poder ser utilizada tanto diretamente no código-fonte quanto somente na descrição do serviço (arquivo .yaml)), e a anotação é rica em conceitos diretamente ligadas ao domínio graças à OntoPrivacy.

Em (Grünewald *et al.*, 2021) é apresentado o TIRA, um ToolBox visando alcançar a transparência relacionada ao GDPR em arquiteturas de

aplicativos RESTful. Compreende uma extensão OpenAPI e respectivos vocabulários especificamente adaptados aos requisitos de transparência do GDPR. Essa abordagem se assemelha à abordagem aqui apresentada, porém TIRA possui um vocabulário próprio para realizar as anotações (x-tira), que é apenas uma forma de indicar dados pessoais (PD Indicator). O restante das informações são adicionadas à anotação de forma livre, o que difere da nossa abordagem, em que as anotações (exemplo, x-refersTo) são alimentadas com informações/conceitos estruturados e definidos pela Ontologia de Privacidade.

O Quadro 2 apresenta uma sumarização dos aspectos comparativos analisados entre a abordagem proposta neste artigo e os trabalhos correlatos. Com base na sumarização, é possível notar que a abordagem proposta se destaca pela proposição de um ontologia de referência – OntoPrivacy – e de seu uso sistemático no processo de anotação.

Quadro 2 - Sumarização da análise dos trabalhos correlatos

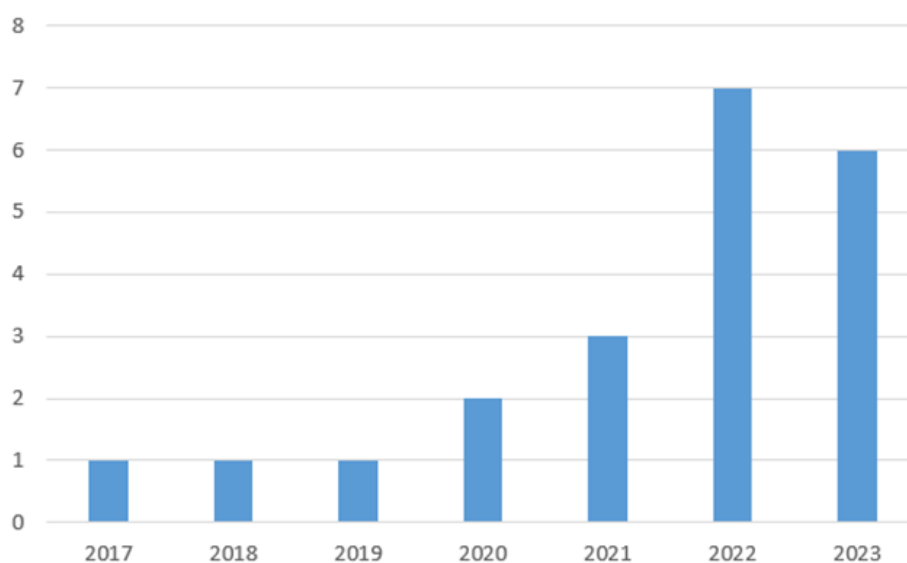
Trabalho	Domínio	Uso de Anotações	Tags	Uso de Ontologias
(Peng; Bai, 2018)	Geral	Tags da OpenAPI	Da própria OpenAPI	Sugerido.
(Hjerppe; Ruohonen; Leppänen, 2019)	LGPD	Anotações próprias em código-fonte	@PersonalData @PersonalDataHandler @PersonalDataEndpoint	Não identificado.
(Grünwald <i>et al.</i> , 2021)	LGPD	Extensões da OpenAPI	x-tira x-tira-ignore	Não identificado.
Abordagem deste trabalho	LGPD	Extensões da OpenAPI	x-refersTo x-kindOf x-collectionOn x-operationType Vide Quadro 1.	Sim, como parte da abordagem. OntoPrivacy.

Fonte: Elaborado pelos autores.

No que tange às produções técnicas de software, consideraram-se, a título de caracterização, os registros de programas de computador existentes na base do INPI. Foram realizadas buscas e análise dos registros existentes. As buscas consideraram: (i) as palavras-chave “proteção” e “dados”, ambas

existentes no título do programa, e (ii) as palavras-chave “privacidade”, “privacy” e “lgpd”, com qualquer delas existente no título do programa. Como resultado final, obtiveram-se 21 (vinte e um) registros, já eliminando-se duas duplicações nas buscas. A Figura 10 apresenta a distribuição temporal dos 21 (vinte e um) registros de software selecionados na base do INPI.

Figura 10 - Registros de software obtidos na base de dados do INPI



Fonte: Elaborado pelos autores.

Pela distribuição, pode-se notar uma tendência de crescimento com ligeira estabilização de 2022 para 2023. Embora em 2017, 2018 e 2019 havia certa preocupação com aspectos de privacidade, foi a partir de 2020 (com a publicação da LGPD no Brasil) que o número de registros de software teve crescimento acentuado. Assim, o software Privacy Finder desenvolvido no contexto deste trabalho está alinhado à tendência de produção técnica e seu registro no INPI será solicitado em breve.

Tomando como exemplo, dois dos registros de programas de computador selecionados, tem-se os softwares SysPAD e LGPD compliance. O SysPAD, com registro em 2023, busca mitigar a possibilidade de vazamento de dados em sistemas legados, permitindo empresas e instituições armazenarem e/ou divulgarem dados sem ferir as leis de privacidade (Brasil, 2023). O LGPD compliance é uma plataforma web que disponibiliza um questionário (*checklist*)

para avaliação da adequabilidade de projetos de software à LGPD, provendo dicas e ações práticas visando a conformidade desses projetos (Costa; Alves; Guimarães, 2022). Além desses dois programas, outros mais figuram dentre os analisados, com destaque para aqueles que visam a dar suporte a empresas se adequarem e operarem segundo a LGPD. Neste contexto, acredita-se que o Privacy Finder, utilizado em conjunto com a abordagem proposta neste trabalho, se coloca como uma opção possível para ser utilizada em conjunto com os programas de computador existentes, na medida em que permite melhor descrever APIs de serviços web disponibilizadas, por exemplo, por tais sistemas.

6 Considerações finais

Este trabalho apresentou uma abordagem para melhorar a descrição de serviços web REST baseada em anotações semânticas que são ancoradas em uma ontologia de referência do domínio de Privacidade de Dados, a OntoPrivacy. Os conceitos especificados na OntoPrivacy, juntamente com o mecanismo de anotação, fornecem um elo entre a descrição OpenAPI e os conceitos da ontologia.

Como validação, a abordagem foi aplicada na descrição da API Pix do Banco Central do Brasil a fim de enriquecê-la semanticamente com metadados acerca de privacidade de dados. Ademais, uma ferramenta foi implementada a fim de dar suporte à busca/localização das anotações realizadas em descrições de serviços usando a abordagem. Com isso, apoia-se o usuário consumidor do serviço, na medida em que apresenta todas as anotações já realizadas e os respectivos locais com dados sensíveis.

Quanto aos resultados esperados, espera-se, com a abordagem, apoiar organizações a mapearem o tratamento de dados pessoais em seus sistemas de informação. Mais especificamente, do ponto de vista do desenvolvimento de sistemas, a abordagem pode ser integrada às rotinas de implementação de APIs serviços web, de modo a exigir o mínimo de esforço adicional possível para a equipe de Tecnologia da Informação (TI). Do ponto de vista de áreas organizacionais responsáveis pela Gestão da Informação e Governança da Informação, tal abordagem pode ser útil na medida em que se propõe a facilitar

a incorporação de metadados sobre quais funcionalidades e quais dados trocados entre os sistemas lidam com informações sensíveis. Isso pode, por exemplo, ampliar a conformidade requerida no trato das informações organizacionais.

Como trabalhos futuros pode-se: (i) incorporar à abordagem proposta, diretrizes de descrição de serviços, a fim de orientar os desenvolvedores das APIs no uso da abordagem; (ii) aprimorar o conjunto de *tags* de extensão usados atualmente (p.ex., *x-refersTo* e *x-kindOf*) a fim de propor *tags* que sejam mais bem fundamentadas conceitualmente; (iii) avançar no mecanismo de busca de anotações rumo ao a buscas semânticas (com mecanismos de inferência), incluindo também mecanismos de validação; (iv) aplicar a abordagem junto a equipes de desenvolvimento de sistemas a fim de avaliá-la; e (v) aplicar a abordagem junto a equipes de Gestão e de Governança da Informação para verificar a eficácia no que tange a questões de conformidade e implementação de políticas organizacionais baseadas na LGPD.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR ISO/IEC 29100**: Tecnologia da informação - Técnicas de segurança - Estrutura de privacidade. Rio de Janeiro: ABNT, 2020.

BANCO CENTRAL. **API Pix, 2024**. Brasília: Banco Central do Brasil, 2024.

BOOTH, D.; HAAS, H.; MCCABE, F.; NEWCOMER, E.; CHAMPION, M.; FERRIS, C.; ORCHARD, D. **Web Services Architecture**. Wakefield: W3C Working Group Note, 2004.

BRASIL. Lei n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**: seção 1, Brasília, ano 155, n. 157, p. 59, 15 de ago. 2018.

BRASIL. Ministério da Educação. **Vitrine MEC de tecnologias**: SysPAD: proteção e privacidade de dados em sistemas legados. Brasília: MEC, 7 dez. 2023.

COSTA, L. M. T.; ALVES, M. B.; GUIMARÃES, R. F. **LGPD compliance**: aplicando *checklist* para avaliação de sistemas à luz da Lei Geral de Proteção de Dados. 2022. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) - Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Campina Grande, 2022.

CREMASCHI, M.; PAOLI, F. Toward automatic semantic API descriptions to support services composition. *In: De Paoli, F., Schulte, S., Broch Johnsen, E. (ed.) Service-Oriented and Cloud Computing - ESOC 2017*. Cham: Springer, 2017. p. 159-167.

DUTRA, F. G. de C.; BARBOSA, R. R. Modelos e etapas para a gestão da informação: uma revisão sistemática de literatura. **Em Questão**, Porto Alegre, v. 26, n. 2, p. 106-131, 2020. Disponível em: <https://doi.org/10.19132/1808-5245262.106-131> . Acesso em: 19 dez. 2024.

EL GHOSH, M.; ABDULRAB, H. Capturing the basics of the GDPR in a well-founded legal domain modular ontology. *In: NEUHAUS, F.; BRODARIC, B. (ed.). Formal ontology in information systems: proceedings of the twelfth International Conference (FOIS 2021)*. Amsterdam: IOS Press Ebooks, 2021. p. 144-158.

FONSECA, F. The double role of ontologies in information science research. **Journal of the American Society for Information Science and Technology**, New Jersey, v. 58, n. 6, p. 786-793, 2007. Disponível em: <https://doi.org/10.1002/asi.20565> . Acesso em: 19 dez. 2024.

INFORMATION GOVERNANCE. *In: GARTNER Glossary*. Stamford: Gartner Inc., c2024.

GHARIB, M.; GIORGINI, P.; MYLOPOULOS, J. COPri v.2 - A core ontology for privacy requirements. **Data & Knowledge Engineering**, Amsterdam, v. 133, p. 1-20, 2021. Disponível em: <https://doi.org/10.1016/j.datak.2021.101888> . Acesso em: 19 dez. 2024.

GOMADAM, K.; RANABAHU, A.; SHETH, A. **SA-REST**: Semantic annotation of web resources. Wakefield: W3C Member Submission, 2010.

GRÜNEWALD, E. *et al.* TIRA: an openapi extension and toolbox for gdpr transparency in restful architectures. *In: IEEE EUROPEAN SYMPOSIUM ON SECURITY AND PRIVACY WORKSHOPS*, 2021, Vienna. **Proceedings [...]**. New York: IEEE, 2021. p. 312-319.

GUARINO, N. (ed.). **Formal ontology and information systems**. Amsterdam: IOS Press, 1998.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. 2005. Thesis (Master's) - Centre for Telematics and Information Technology, Enschede, 2005.

GUIZZARDI, G. On ontology, ontologies, conceptualizations, modeling languages, and (meta) models. *In: VASILECAS, J. E.; CAPLINSKAS, A. (ed.). Databases and information systems IV: selected papers from the seventh*

International Baltic Conference DBandIS'200. Amsterdam: IOS Press, 2007. p. 18-39.

HJERPPE, K.; RUOHONEN, J.; LEPPÄNEN, V. Annotation-based static analysis for personal data protection. *In: FRIEDEWALD, M. et al. (ed.). Privacy and identity management data for better living: AI and privacy: IFIP International Summer School on Privacy and Identity Management. Cham: Springer, 2019. p. 343-358.*

KARAVISILEIOU, A. *et al.* Automated ontology instantiation of OpenAPI REST service descriptions. *In: ARAI, K. (ed.) Advances in Information and Communication - FICC 2021. Cham: Springer, 2021. p. 945-962.*

KOPECHÝ, J. *et al.* Sawsdl: Semantic annotations for wsdl and xml schema. **IEEE Internet Computing**, New York, v. 11, n. 6, p. 60-67. 2007. Disponível em: <https://doi.org/10.1109/MIC.2007.134> . Acesso em: 19 dez. 2024.

LANTHALER, M.; GÜTL, C. A semantic description language for restful data services to combat semaphobia. *In: IEEE International conference on digital ecosystems and technologies, 5., 2011, Daejeon. Proceedings [...]. New York: IEEE 2011. p. 47-53.*

LANTHALER, M.; GÜTL, C. Seamless integration of RESTful services into the web of data. **Advances in Multimedia**, New Jersey, p. 1-14, 2012. Disponível em: <https://doi.org/10.1155/2012/586542> . Acesso em: 19 dez. 2024.

LANTHALER, M.; GÜTL, C. Hydra: a vocabulary for hypermedia-driven Web APIs. *In: WWW2013 WORKSHOP ON LINKED DATA ON THE WEB, Rio de Janeiro, 2013. Proceedings [...]. Aachen: CEUR Workshop Proceedings, 2013. v. 996, p. 35-38.*

MARTIN, D. *et al.* **OWL-S**: semantic markup for web services. Wakefield: W3C Member Submission, 2004. v. 22, n. 4.

MELO, R.; BATISTA, G. V.; LIMA, P. R. S. A Lei Geral de Proteção de Dados Pessoais à luz da Ciência da Informação: uma análise bibliométrica. **P2P & Inovação**, Rio de Janeiro, v. 10, n. 1, p. 356-370, 2023. Disponível em: <https://doi.org/10.21728/p2p.2023v10n1.p356-370> . Acesso em: 19 dez. 2024.

MONTEIRO, S. A.; DUARTE, E. N. Bases teóricas da gestão da informação: da gênese às relações interdisciplinares. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, Brasil, v. 9, n. 2, p. 89-106, 2018. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v9i2p89-106> . Acesso em: 18 nov. 2024.

NARDI, J. C.; FALBO, R. de A.; ALMEIDA, J. P. A. Foundational ontologies for semantic integration in EAI: a systematic literature review. *In: Douligieris, C.*

et al. (ed.). **Collaborative, trusted and privacy-aware e/m-services - I3E 2013**. Heidelberg: Springer, 2013.

NETEDU, A. *et al.* A web service composition method based on openapi semantic annotations. *In: Chao, K. M. et al.* (ed.). **Advances in E-Business Engineering for Ubiquitous Computing - ICEBE 2019**. Cham: Springer, 2019. p. 342-357.

SEMINÁRIO DE PESQUISA EM ONTOLOGIAS NO BRASIL - ONTOBRAS, 17., 2024, Vitória. **Anais [...]**. Porto Alegre: Instituto de Informática/UFRGS, 2024.

OLIVEIRA, A. S. S.; GALLOTTI, M. M. C.; BARBOSA NETO, P. A. Relação da gestão da informação e Lei Geral de Proteção de Dados Pessoais: um estudo bibliográfico e bibliométrico. **Ciência da Informação em Revista**, Maceió, v. 11, e15444, 2024. Disponível em: <https://doi.org/10.28998/cirev.2024v11e15444> . Acesso em: 19 dez. 2024.

OASIS. **Reference Model for Service Oriented Architecture 1.0**: Committee Specification 1. Woburn: OASIS Open Group, 2006.

OPENAPI INITIATIVE. **OpenAPI Specification, 2024**. San Francisco: GitHub Inc., 2024.

OREN, E. *et al.* **What are semantic annotations**. Galway: Digital Enterprise Research Institute, 2006. p. 1-14.

PALMIRANI, M. *et al.* Pronto: privacy ontology for legal reasoning. *In: Kő, A.; Francesconi, E.* (ed.). **Electronic Government and the Information Systems Perspective - EGOVIS 2018**. Cham: Springer, 2018. p. 139-152.

PANDIT, H. J. *et al.* GConsent-a consent ontology based on the GDPR. *In: Hitzler, P., et al* (ed.). **The Semantic Web - ESWC 2019**. Cham: Springer, 2019. p. 270-282.

PENG, C.; BAI, G. Using tag based semantic annotation to empower client and rest service interaction. *In: THE INTERNATIONAL CONFERENCE ON COMPLEXITY, FUTURE INFORMATION SYSTEMS AND RISK*, 3., 2018, Funchal. **Proceedings [...]**. Lisboa: INSTICC, 2018. p. 64-71.

POPOV, B. *et al.* Towards semantic web information extraction. *In: HUMAN LANGUAGE TECHNOLOGIES WORKSHOP AT THE INTERNATIONAL SEMANTIC WEB CONFERENCE*, 2., 2003. **Proceedings [...]**. [S.l.: s.n.], 2003.

PRIVACY FINDER. 1.0. **Colatina, 2021**. [S.l.]: Snowflake Inc, 2024.

- SÁNCHEZ-FERNANDEZ, Luis; FERNANDEZ-GARCÍA, Norberto. **The Semantic Web: fundamentals and a brief state-of-the-art**. **UPGRADE**, [s.l.], v. 6, n. 6, p. 5-11. 2005.
- SANTANA, C. J. L. **SWS-Editor: uma ferramenta baseada em serviços para anotação semântica de serviços web RESTful**. 2015. Dissertação (Mestrado em Sistemas e Computação) - Universidade Salvador, Salvador, 2015.
- SILVA, A. M.; ESTRELA, S. C. L.; ARAÚJO, E. P. de O. **Gestão da informação nas PME industriais de Portugal (Norte e Centro): diagnóstico e bases para um modelo orientador**. Porto: Centro de Investigação Transdisciplinar Cultura, Espaço e Memória, 2024.
- SMITH, B. The relevance of philosophical ontology to information and computer science. *In*: Hagenbruger, R.; Riss, U. V. (ed.), **Philosophy, computing and information science**. London: Pickering & Chattoo, 2014. p. 75-83.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: principles and methods. **Data & Knowledge Engineering**, Amsterdam, v. 25, n. 1-2, p. 161-197, 1998. Disponível em: [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6). Acesso em: 19 dez. 2024.
- SWAGGER. **Swagger Petstore - OpenAPI 3.0, 2024**. Pet store sample. Somerville: Swagger, 2024.
- TAPSOBA, L. S. K.; TRAORE, Y.; MALO, S. Interoperability approach for hospital information systems based on the composition of web services. **Procedia Computer Science**, Amsterdam, v. 219, p. 1161-1168, 2023. Disponível em: <https://doi.org/10.1016/j.procs.2023.01.397>. Acesso em: 19 dez. 2024.
- THOMAS, C (ed.). **Ontology in information science**. Rijeka: IntechOpen, 2018.
- TUMER, A.; DOGAC, A.; TOROSLU, I. H. A semantic-based user privacy protection framework for web services. *In*: Mobasher, B.; Anand, S. S. (ed.). **Intelligent Techniques for Web Personalization - ITWP 2003**. Heidelberg: Springer, 2003. p. 289-305.
- USCHOLD, M.; GRUNINGER, M. Ontologies: principles, methods and applications. **The Knowledge Engineering Review**, Cambridge, v. 11, n. 2, p. 93-136, 1996. Disponível em: <https://doi.org/10.1017/S0269888900007797>. Acesso em: 19 dez. 2024.
- VERBORGH, R. *et al.* **Description and interaction of restful services for automatic discovery and execution**. *In*: FTRA INTERNATIONAL WORKSHOP ON ADVANCED FUTURE MULTIMEDIA SERVICES, 2011.

Proceedings [...]. [S.l.]: Future Technology Research Association International, 2011.

VITVAR, T. *et al.* WSMO-lite annotations for web services. *In: Bechhofer, S. et al. (ed.). The semantic web: research and applications - ESWC 2008.* Heidelberg: Springer, 2008. p. 674-689.

WESTIN, A. F. Privacy and freedom. **Washington and Lee Law Review**, Washington, v. 25, n. 1, p. 166, 1968.

WU, J. *et al.* Privacy information classification: a hybrid approach. **arXiv**, [s.l.], 2021. Disponível em: <https://doi.org/10.48550/arXiv.2101.11574> . Acesso em: 19 dez. 2024.

XAVIER, O. C. **Serviços web semânticos baseados em RESTful: um estudo de caso em redes sociais online.** 2011. Dissertação (Mestrado em Ciência da Computação) - Programa de Pós-Graduação, Instituto de Informática, Universidade Federal de Goiás, Goiânia. 2011.

Support in the adoption of the General Data Protection Law through semantic annotations in description of web services

Abstract: With the General Data Protection Law in effect in Brazil since 2020, many organizations are being compelled to implement data processing practices. This has been addressed by different knowledge areas such as Information Science, Information Management, and Information Technology. These areas aim to understand and incorporate the legal requirements of General Data Protection Law in organization processes and to implement them in information systems. In this context, a challenge faced is the execution of data processing through web services (e.g., collection, storage, and deletion) made available and accessed by various information systems, including those across different organizations. Identifying which services perform personal data processing and, based on that, conducting practices in compliance with the General Data Protection Law is a crucial aspect. In this work, an approach is proposed for semantic enrichment of REST web service descriptions using annotations based on OpenAPI and a domain reference ontology for Data Privacy (OntoPrivacy). This approach aims to provide additional information, through metadata, about the processing operations and the types of sensitive data being handled. Alongside this approach, a software tool (Privacy Finder) was developed to support the automatic location of the annotations in described service APIs. For validation, the proposed solution was applied in the context of the Pix *Application Programming Interface* of the Central Bank of Brazil. Additionally, the solution was analyzed in light of related works in the literature, as well as technical productions whose computer program fillings were found in the Brazilian National Institute of Industrial Property database. The analysis also

revealed a trend in the growth of such productions, particularly since the publication year of the mentioned law.

Keywords: data privacy; GDPL; ontology; semantic annotation; web services; metadata; information management; information governance.

Declaração de autoria

Concepção e elaboração do estudo: Dornélio Mori Junior, Julio Cesar Nardi, Fabiano Borges Ruy.

Coleta de dados: Dornélio Mori Junior, Julio Cesar Nardi, Fabiano Borges Ruy

Análise e interpretação de dados: Dornélio Mori Junior, Julio Cesar Nardi, Fabiano Borges Ruy, Giovany Frossard Teixeira.

Redação: Dornélio Mori Junior, Julio Cesar Nardi, Fabiano Borges Ruy, Giovany Frossard Teixeira.

Revisão crítica do manuscrito: Dornélio Mori Junior, Julio Cesar Nardi, Fabiano Borges Ruy, Giovany Frossard Teixeira.

Autoria para correspondência

Julio Cesar Nardi
julionardi@ifes.edu.br

Como citar

MORI JUNIOR, Dornélio; NARDI, Julio Cesar; RUY, Fabiano Borges; TEIXEIRA, Giovany Frossard. Apoio na adoção da Lei Geral de Proteção de Dados Pessoais por meio de anotações semânticas em descrições de serviços Web. **Em Questão**, Porto Alegre, v. 31, e-139608, 2025. <https://doi.org/10.1590/1808-5245.31.139608>

Recebido: 10/04/2024

Aceito: 02/12/2024



¹ WILSON, T. D. Information management. In: FEATHER, J.; STURGES, P. (ed.). International Encyclopedia of Information and Library Science. London: Routledge, 2002. *Apud* Monteiro e Duarte (2018).

² Linguagem de serialização de dados legível por humanos bastante utilizada em arquivos de configuração.

³ Relutância na aplicação de abordagens que requerem amplo conhecimento sobre tecnologias semânticas (Lanthaler; Gütl, 2011).