ESCOLA DE GUERRA NAVAL

CC (T) MADALENA LOPES E SILVA / C-Sup 2024

GOVERNANÇA EM DEFESA:
Uma proposta de aplicação da ontologia Sec4ML-O para o Ministério da Defesa

Rio de Janeiro

2024

CC (T) MADALENA LOPES E SILVA / C-Sup 2024

GOVERNANÇA EM DEFESA:
Uma proposta de aplicação da ontologia Sec4ML-O para o Ministério da Defesa

Monografia apresentada à Escola de Guerra Naval, como requisito parcial para a conclusão do Curso Superior.

Orientador: CMG CHIARA LEÃO ARAÚJO DE FRANÇA DELGADO DE FREITAS C-Sup 2024

Rio de Janeiro

Escola de Guerra Naval

2024

DECLARAÇÃO DA NÃO EXISTÊNCIA DE APROPRIAÇÃO INTELECTUAL IRREGULAR

Declaro que este trabalho acadêmico: a) corresponde ao resultado de investigação por mim desenvolvida, enquanto discente da Escola de Guerra Naval (EGN); b) é um trabalho original, ou seja, que não foi por mim anteriormente utilizado para fins acadêmicos ou quaisquer outros; c) é inédito, isto é, não foi ainda objeto de publicação; e d) é de minha integral e exclusiva autoria.

Declaro também que tenho ciência de que a utilização de ideias ou palavras de autoria de outrem, sem a devida identificação da fonte, e o uso de recursos de inteligência artificial no processo de escrita constituem grave falta ética, moral, legal e disciplinar. Ademais, assumo o compromisso de que este trabalho possa, a qualquer tempo, ser analisado para verificação de sua originalidade e ineditismo, por meio de ferramentas de detecção de similaridades ou por profissionais qualificados.

Assinatura digital gov.br

Documento assinado digitalmente
gov.br
MADALENA LOPES E SILVA
Data: 20/11/2024 19:12:45-0300
Verifique em https://validar.iti.gov.br

DEDICATÓRIA

Dedico esta monografia a Deus por me permitir todas as minhas conquistas, à minha família por estar sempre ao meu lado me apoiando e a todos que contribuíram para o meu crescimento pessoal ao longo da minha trajetória acadêmica.

AGRADECIMENTO

Agradeço a todos que contribuíram para a conclusão desse trabalho acadêmico. Primeiramente, a Deus que me permitiu chegar até aqui e alcançar todas as minhas conquistas acadêmicas. À minha família pelos momentos de ausência suportados. À minha orientadora CMG (RM1) Chiara pelos valiosos ensinamentos e orientações recebidos tão necessários para o meu crescimento profissional e intelectual. Ao Dr. Luiz Henrique Cavalcanti da Silva, Encarregado pelo Tratamento de Dados Pessoais do Ministério da Defesa, pelo atendimento solícito e pelo fornecimento das informações relevantes e necessárias para esta monografia.

A privacidade é algo que podes vender
mas não podes comprar de volta.
Bob Dylan

RESUMO

O objeto de pesquisa deste trabalho é a propositura da aplicação da ontologia Sec4ML-O para o Ministério da Defesa (MD) como forma de alcançar a conformidade de seus sistemas e processos com a Lei Geral de Proteção de Dados (LGPD). Com o advento de legislações de proteção de dados ao redor do mundo, surgiu o momento que a sociedade brasileira vive hoje: a adequação de sistemas e processos para a LGPD. Esta legislação não foi a primeira nessa temática, ao contrário, surgiu de um processo de amadurecimento e evolução social com relação à necessidade de garantir a privacidade dos indivíduos que utilizam sistemas ou estão envolvidos em processos desenvolvidos por entidades ou empresas brasileiras. Apesar disso, nosso país ainda está enfrentando dificuldades para atingir a conformidade plena com relação àquela legislação. Este trabalho procura também, como objetivos específicos, identificar as capacidades adicionadas com a aplicação da ontologia Sec4ML-O e exemplificar possibilidades de classificação e tratamentos aplicáveis a cada tipo de dado tratado. As contribuições feitas por Cavoukian, Brito e Machado, Guarino e Herschel criam um embasamento conceitual necessário e suficiente para a compreensão da proposta apresentada neste trabalho. No caso do MD, já existe hoje um processo claro, bem estruturado e evolutivo sendo empreendido pela Assessoria Especial de Integridade (AESPI) do MD para essa adequação. A proposta de aplicabilidade da ontologia Sec4ML-O vem ao encontro do momento atual de amadurecimento dos processos e sistemas, visando trazer maior clareza, robustez e confiabilidade na conformidade para a LGPD.

Palavras-chave: Privacidade. Ontologia. LGPD. Proveniência. Programa de Gestão em Privacidade.

ABSTRACT

Governance in Defense: A proposal for the application of the Sec4ML-O ontology for the Ministry of Defense

The research object of this work is the proposal of the application of the Sec4ML-O ontology for the Ministry of Defense (MD) to achieve the compliance of its systems and processes with the General Data Protection Law (LGPD). With the advent of data protection legislation around the world, the moment that Brazilian society is experiencing today has emerged: the adaptation of systems and processes to the LGPD. Considering that this legislation was not the first on this topic, on the contrary, it arose from a process of maturation and social evolution in relation to the need to guarantee the privacy of individuals who use systems or are involved in processes developed by Brazilian entities or companies, our country is still facing difficulties to achieve full compliance with that legislation. This work looks at for, as specific objectives, to identify the capabilities added with the application of the Sec4ML-O ontology and to exemplify possibilities of classification of the Sec4ML-O ontology and exemplify possibilities of classification and processing applicable to each type of data processed. The contributions made by Cavoukian, Brito and Machado, Guarino and Herschel create a necessary and sufficient conceptual basis for understanding the proposal presented in this work. In the case of the MD, there is already a clear, well-structured and evolutionary process being undertaken by the Special Integrity Advisory (AESPI) of the MD for this adaptation. The proposal for the applicability of the Sec4ML-O ontology meets the current moment of maturation of processes and systems, aiming to bring greater clarity, robustness and reliability in compliance to the LGPD.

Keywords: Privacy. Ontology. LGPD. Provenance. Privacy Management Program.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LGPD	Lei Geral de Proteção de Dados
GDPR	General Data Protection Regulation
GT	Grupo de Trabalho
PGP	Programa de Gestão em Privacidade

SUMÁRIO

1 INTRODUÇÃO

A governança de dados vem ganhando relevância na gestão moderna de organizações. Especialmente no contexto de defesa, em que a segurança, a privacidade e a integridade dos dados são fundamentais para a realização das ações de defesa de maneira eficaz e segura. Em um cenário global cada vez mais interconectado e digitalizado, as forças armadas enfrentam desafios cada vez mais complexos relacionados à proteção de informações dos indivíduos, ao mesmo tempo em que devem garantir a interoperabilidade e a eficiência na troca de dados entre diversas entidades.

A evolução tecnológica trouxe desafios a proteção de dados pessoais. Técnicas de para o processamento de Big Data, algoritmos de aprendizado de máquina e outras abordagens de inteligência artificial permitem que grandes volumes de dados sejam processados e relacionados, aumentando o risco de reidentificação de indivíduos, mesmo em conjuntos de dados anonimizados. As abordagens de anonimização de dados é uma solução que, embora atraente, precisa ser aplicada de maneira que não comprometa a utilidade dos dados e deve ser revisitada rotineiramente de acordo com o avanço de novas ameaças.

A aplicação de técnicas de anonimização permite que organizações continuem a utilizando os dados que controlam para fins de pesquisa, marketing e desenvolvimento de produtos, sem violar a privacidade dos titulares de dados. Entretanto, a anonimização eficaz se tornou uma tarefa cada vez mais difícil. Diversos estudos (Fung *et al.*, 2010) demonstram que, mesmo após a aplicação de técnicas avançadas de anonimização, é possível reidentificar indivíduos por meio da combinação de diferentes fontes de dados.

No sentido de mitigar a possibilidade de reidenticação de indivíduos, abordagens de preservação da privacidade, também no setor de defesa não se limitam a proteger o cyberespaço contra ameaças cibernéticas, mas compreendem também a definição de políticas e processos que garantam qualidade, disponibilidade e transparência às informações ao longo de seu ciclo de vida. Com o advento dos sistemas informatizados, foi observado em larga escala o uso de dados pessoais para diversas finalidades que não necessariamente as previstas inicialmente quando os dados de indivíduos são por estes disponibilizados.

A fim de solucionar questões relativas à proteção da privacidade, elemento primordial da governança de dados, a União Europeia, como organização supraestatal,

promulgou a *General Data Protection Regulation* (União Europeia, 2018) com o intuito de estabelecer níveis de governança de dados em suas sociedades. No Brasil, a Lei nº 12.737/2012, conhecida como “Lei Carolina Dieckman”, alterou o Código Penal a fim de prever a tipificação criminal de delitos informáticos. O Marco Civil da Internet, Lei nº 12.965/2014, que estabelece direitos e deveres no uso da Internet no Brasil, também contempla a proteção da privacidade e dos dados pessoais.

Diante desse cenário de vulnerabilidade para com os seus cidadãos, no território brasileiro, seguindo a esteira desse movimento internacional de proteção de dados pessoais, foi promulgada, em 2018, a Lei Geral de Proteção de Dados – LGPD (Brasil, 2018), com sua entrada em vigor a partir de agosto de 2020. Tal legislação possibilita um ambiente mais protetivo para o indivíduo. Entretanto, gera demandas até então negligenciadas ou atendidas de maneira precária devido à inexistência de regulamentações específicas de proteção de dados. Dessa forma, as organizações ainda têm que lidar com os desafios para a mitigação de problemas e prejuízos derivados de violações de privacidade e vazamento de dados.

Apesar da iniciativa de promulgação da LGPD, o Brasil ainda enfrenta um cenário de muitos desafios (Reis, 2023) no que se refere à conformidade com essa legislação. As organizações, sejam elas estatais ou privadas estão enfrentando dificuldades das mais diversas naturezas, em especial na adequação de técnicas e tecnologias para o atendimento às exigências da citada lei. Em nosso país, somente no período de janeiro a março do presente ano, foram registrados pelo governo federal quase 1.600 casos de vazamentos de dados (O tempo, 2024)[1]. Além disso, a LGPD não define exatamente quais as técnicas que devem ser adotadas para a mitigação dos riscos de reidentificação, deixando lacunas de interpretação da lei e criando incertezas quanto à conformidade.

No âmbito do governo federal, surgiram normativos importantes para a conformidade, como a Estratégia de Governo Digital – EGD, período 2020 a 2022. O referido documento estabeleceu, dentre outras questões, que os órgãos e as entidades da administração pública federal direta, autárquica e fundacional deveriam instituir o Comitê de Governança Digital para deliberar sobre os assuntos relativos à

1 Jornal O Tempo. GSI alertou governo sobre vazamento de dados do sistema de pagamentos. Disponível em: https://www.otempo.com.br/politica/governo/gsi-alertou-governo-sobre-vazamento-de-dados-do-sistema-de-pagamentos-1.3443705. Acesso em: 24 abr. 2024.

implementação das ações do governo e ao uso de recursos de tecnologia da informação (TI) nessa implementação. No Ministério da Defesa (MD), este Comitê teve sua instituição atualizada pela Portaria Normativa GM-MD nº 3.572/2022. Além disso, um grupo de trabalho (GT) foi instituído pela Portaria GM-MD nº 5.148/2021, com o propósito de propor a estrutura e as ações de Proteção de Dados Pessoais no âmbito do MD. Derivada dos trabalhos desse GT, foi criada a Diretriz para a Proteção de Dados Pessoais no MD, que norteia a conformidade com a LGPD nesse ministério (Brasil, 2021; 2022).

Além do cenário atual estatal, novos paradigmas de desenvolvimento de sistemas, como *Privacy by Design* (Cavoukian, 2006), têm sido estudados e novos projetos de soluções tem sido norteados por esses princípios. Nesse sentido, processos, projetos e sistemas novos devem ser empreendidos visando esta nova sistemática. Para isso, já há abordagens que possibilitam uma melhor adequação para conformidade com a LGPD, como a abordagem Sec4ML (Silva, 2022), na qual é apresentado um modelo de dados que permite a realização da coleta dos dados de proveniência necessários ao atendimento das obrigatoriedades que a referida lei exige.

O objetivo desta monografia é identificar a aplicabilidade da ontologia Sec4ML-O para a governança em defesa com foco na preservação da privacidade de dados no âmbito do MD. Serão analisados os principais desafios e oportunidades que surgem na implementação de um modelo eficaz de preservação da privacidade para o setor da defesa. No capítulo 2 será apresentado o referencial teórico que fundamenta este trabalho, inclusive aspectos da abordagem Sec4ML, especificamente os relacionados com a Ontologia Sec4ML-O e o seu modelo de dados propostos, além do detalhamento de como a abordagem realiza a Captura de Dados de Proveniência. No terceiro capítulo serão detalhados aspectos importantes do Programa de Gestão em Privacidade, especialmente o Inventário de Dados Pessoais. A Aplicabilidade da ontologia Sec4ml no contexto de conformidade com a LGPD no MD será discutida no capítulo 4. No capítulo 5 é apresentada a conclusão.

2 REFERENCIAL TEÓRICO

Em seu documento, Cavoukian (2006) apresenta os sete princípios fundamentais para o termo "*Privacy by Design*", definidos como o *framework* necessário para o desenvolvimento de sistemas de informação, processos e arquiteturas que considerem, desde a sua concepção, os aspectos necessários para a conformidade em preservação da privacidade. Naquele momento, houve uma mudança de visão relacionada à concepção, projeto e implementação de sistemas de informação e arquiteturas.

Os sete princípios abordam as seguintes questões: (i) a adoção de posicionamento mais proativo com relação a questões de privacidade; (ii) implementação de mecanismos que garantam a privacidade por *default,* ou seja como premissa do sistema e/ou processo a ser desenvolvido; (iii) a adoção de abordagens e frameworks nos quais mecanismos de preservação da privacidade estejam compreendidos; (iv) a funcionalidade deve ser total, ou seja não se pode abrir mão de nenhum requisito do sistema, processo ou arquitetura em função da adoção de medidas para a preservação da privacidade; (v) garantias de preservação da privacidade "fim a fim", em todos os pontos da arquitetura e em todo o ciclo de vida do sistema; (vi) visibilidade e transparência de maneira que permita a conformidade e a responsabilização e (vii) respeito à privacidade do usuário, levando-se em consideração o consentimento expressado.

Fung *et al.* (2010) fornece uma visão abrangente sobre os conceitos e técnicas de anonimização para preservação da privacidade. Os autores trazem conceitos de coleta e compartilhamento de dados para a publicação de dados preservando a privacidade dos titulares contra as ameaças de ataques de reidentificação. São explorados modelos de privacidade além de técnicas de anonimização, como generalização e supressão de dados. O livro destaca a importância de equilibrar a privacidade e a utilidade dos dados anonimizados, apresentando diferentes abordagens para lidar com essa questão, além de discutir as tendências futuras e técnicas emergentes. Esta obra aborda também aspectos ligados a dimensões específicas de dados, tais como redes sociais e dados de saúde. Dentro do seu conteúdo, são abordados também modelos de privacidade, explicando como cada um aborda aspectos específicos da proteção de dados, tais como o **K-anonymity** que garante que cada registro seja indistinguível de pelo menos outros k-1 registros, o **L-diversity** que complementa o K-anonymity ao assegurar diversidade dentro dos

grupos de equivalência e o **T-closeness** que garante que a distribuição de atributos sensíveis em qualquer grupo seja semelhante à distribuição global.

Um pouco mais tarde, Brito e Machado (2014) apresentaram os conceitos relacionados à preservação de privacidade de dados. Após fazerem uma breve análise do cenário do tratamento de dados no âmbito mundial naquele momento, evidenciam a classificação dos dados conforme a sua utilidade e o presumido risco à preservação da privacidade dos dados levando-se em consideração a categoria no

Figura 1 – Categorização de Técnicas de Anonimização

Fonte: Silva (2022).

qual os dados foram classificados, tais como: (i) generalização que é a técnica em que se transforma dados específicos em dados mais genéricos, (ii) supressão que ocorre quando são dados são removidos total ou parcialmente dos conjuntos de dados e (iii) permutação que ocorre quando valores do mesmo atributo são permutados em registros diferentes. As técnicas apresentadas e seus tipos podem ser visualizados na Figura 1. A fim de ilustrar como os dados são afetados pela aplicação de técnicas de anonimização, nas Tabelas 1 e 2 são apresentados registros de dados fictícios antes da aplicação de técnicas de anonimização e depois da aplicação de técnicas de anonimização, respectivamente.

Se usarmos a primeira técnica, uma data de nascimento pode ser generalizada para um intervalo de anos ou a profissão do indivíduo pode ser recolocada em uma categoria superior (“enfermeira” passa a ser “profissional de saúde”), como na Tabela

2 o atributo Ocupação. Já na segunda técnica, parte de código postal pode ser suprimido impedindo a identificação do logradouro a que se refere.

Tabela 1 – Exemplos de dados antes da aplicação de técnicas de anonimização.

ID	Nome	Ocupação	CEP	Telefone
1	José Alves Fonseca	Médico	22545-323	986770517
2	Maria Luiza Bezerra Souza	Enfermeira	21690-455	996900234
3	Leandro Nogueira dos Santos	Analista	13250-123	912330877
4	Luís Antônio Lopes	Desenvolvedor	20234-001	978990057
5	Priscila Sampaio Neves	Arquiteta	14687-300	964081200

Fonte: Autoria própria.

Tabela 2 - Exemplos de dados depois da aplicação de técnicas de anonimização.

ID[1]	Nome[2]	Ocupação[3]	CEP[4]	Telefone[5]
4	José Alves Oliveira	Profissional de Saúde	22545	996900234
5	Maria Luiza Batista Souza	Profissional de Saúde	21690	986770517
6	Leandro Silveira dos Santos	Profissional de TI	13250	912330877
7	Luís Antônio Carvalho	Profissional de TI	20234	912330877
8[6]				

Fonte: Autoria própria. Legenda: 1- Adição de Ruído 2- Substituição 3- Generalização 4- Truncagem 5- Permutação 6- Supressão

No caso da terceira técnica, valores numéricos podem ser permutados aleatoriamente no mesmo atributo, dificultando a identificação. Todas as técnicas citadas devem ser consideradas para os dados a serem armazenados já anonimizados.

Entretanto, há a situação em que os dados serão armazenados sem nenhum tratamento de preservação de privacidade, mas, no momento da publicação e/ou consulta, teremos a necessidade de adição de ruído para garantir a privacidade dos titulares dos dados. Os autores citam modelos matemáticos como o de Privacidade

Diferencial em que o ruído é calculado e aplicado sobre o resultado de uma consulta realizada sobre um conjunto de dados. É importante ressaltar que modelos matemáticos como este mantêm as características estatísticas dos dados apresentados, o que torna esse tipo de técnica bastante empregado no mercado.

A definição de ontologia também é um conceito importante para a compreensão da proposta desta monografia. Guarino (2009), nesse trabalho precursor, nos traz esse conceito fundamental para o entendimento de como podemos criar conceitualizações formais para os diversos conceitos existentes no mundo. O autor define que, derivado do conceito aristotélico, uma ontologia é a especificação explícita de uma conceitualização compartilhada. Então para cada conceito existente no mundo, a ontologia que o define deverá especificar esse conceito de maneira explícita e dar-lhe plena divulgação que este seja plenamente conhecido e assim compartilhado. Logo, uma ontologia define formalmente determinados conceitos e torna possível, da maneira menos custosa ao interessado, encontrar e utilizar essa conceitualização.

No contexto de ciência da computação, as ontologias são representadas por modelos em determinadas linguagem de representação. A correta representação de “coisas” do mundo real, quer sejam concretas ou abstratas, é fundamental para a eficácia de uma ontologia. Uma ontologia é tão boa quanto melhor se aproximar e definir o mundo real. A definição ontológica de conceitos e suas relações através de uma linguagem de representação constitui um artefato de *software* e pode ser processado por mecanismos automatizados.

Alguns artefatos ontológicos basearam a construção da ontologia Sec4ML-O, cerne da proposta apresentada nessa dissertação. Podemos citar alguns tais como *Friend of a Friend*[2] (FOAF) com os conceitos relacionados a pessoas, *Common Warehouse Metamodel*[3] (CWM) com os conceitos relacionados a metadados de grandes armazenamentos de dados, *Dublin Core*[4] (DC) com os conceitos de metadados de repositórios de dados de pesquisa, *Provenance Ontology* (PROV-O) (Lebo, 2013) com os conceitos relacionados a metadados de proveniência, GDPR

[2] http://xmlns.com/foaf/spec/
[3] https://www.omg.org/spec/CWM/1.1/About-CWM
[4] https://www.dublincore.org/

Provenance Ontology[5] (GDPROV) com os conceitos definidos pela GDPR, entre outras.

Para o cumprimento das exigências legais exigidas pela LGPD, é mandatório a coleta de dados de proveniência. Proveniência (Herschel *et al.*, 2017) é toda informação que pode ser coletada sobre entidades, indivíduos e processos para ser utilizada com finalidades de garantir privacidade, qualidade, responsabilização, reprodutibilidade, análise de fluxo de processos, entre outras aplicações. Para modelar esses conceitos, há um padrão estabelecido pela W3C[6], o PROV-O (Lebo, 2013), no qual os conceitos envolvidos com a coleta de dados de proveniência estão definidos. Esta ontologia fundamentou a ontologia Sec4ML-O sob o ponto de vista da coleta de proveniência. Além dos reusos apontados, toda a ontologia é estereotipada com base na ontologia de proveniência PROV-O.

Os autores classificam a proveniência em quatro categorias: (i) metadados de procedência, ou seja, dados de procedência mais geral, independente de modelos; (ii) procedência de sistemas de informação, em que metadados sobre os sistemas de informação utilizados são capturados; (iii) procedência de fluxo de trabalho, em que são capturados metadados de fluxos de trabalho e (v) procedência de dados, em que são capturados metadados sobre o processamento de dados de maneira individualizada, permitindo o rastreamento dessas ações.

Curado Jr. (2021) apresentou, em seu trabalho de conclusão de curso, uma análise da situação, à época, em que havia uma lacuna normativa para a implantação da adequabilidade da LGPD no âmbito do MD. Este autor apontou que a aplicação dessa lei no setor de defesa nacional apresenta complexidades significativas. Aspectos como a possibilidade de incompatibilidade legal foram ressaltados. A Lei de Acesso à Informação (LAI) pode ser analisada, em alguns aspectos, como incompatível com a LGPD, gerando possíveis conflitos interpretativos, ressaltou também o autor. Um outro ponto ressaltado foi o possível impacto em instituição de estado, como a área de defesa nacional. Destacou a necessidade de criação de regulamentações específicas para essa área. São ressaltadas ainda as seguintes questões: a garantia de tráfego de dados pessoais entre os diversos órgãos militares preservando a privacidade dos titulares, a evitação da divulgação imprópria de dados

[5] https://openscience.adaptcentre.ie/projects/CDMM/GDPRov/
[6] https://www.w3.org/

pessoais e o fortalecimento da segurança da informação como pilar para a garantia da proteção de dados pessoais.

Reis (2023), apresentou, na sua dissertação de mestrado, uma Revisão Sistemática da Literatura (RSL), em que procurou investigar os impactos das leis de proteção de dados nas atividades de engenharia de *software*. Investigou também o quanto os processos de desenvolvimento de *software* precisaram ser modificados ou ter algo acrescentado para garantir o cumprimento de tais legislações. Dentre as questões de pesquisa que procurou responder, identificou que a maioria dos estudos citam que os maiores impactos oriundos da adequação da lei ocorrem na fase de especificação ou coleta de requisitos. Os trabalhos pesquisados em sua RSL também ressaltam dificuldades específicas pelo fato de os requisitos serem especificados em linguagem natural, o que pode gerar ambiguidade. No caso de requisitos baseados em legislações, a dificuldade pode ser maior ainda pois pode haver mais de uma interpretação da referida norma. Outro aspecto destacado é o fato de que, no passado, esse tipo de requisito tenha sido considerado não funcional, para o quais as empresas geralmente adaptavam os processos de *software*, situação que hoje não se mostra mais possível.

2.1 A ABORDAGEM Sec4ML

A abordagem Sec4ML foi proposta para apoiar o problema de escassez de conjuntos de dados disponíveis para pesquisa e desenvolvimento de novas soluções de TI, especialmente no âmbito de segurança da informação. Para tal, propôs, entre outras funcionalidades: (i) a aplicação de técnicas de preservação da privacidade sobre os dados processados e (ii) uma ontologia que fundamenta a captura de dados de proveniência sobre os processos aplicados aos dados. Apresentou também um experimento que demonstra a efetividade dos processos de preservação de privacidade e a capacidade de tornar um determinado processo ou sistema em conformidade com a LGPD ou GDPR.

2.1.1 ONTOLOGIA E MODELO DE DADOS

A dissertação de Silva (2022) propôs a ontologia Sec4ML-O. Esta ontologia define os novos conceitos que, conjuntamente com os conceitos reusados necessários, permitem a captura de metadados de proveniência de sistemas de informação e/ou processos que sejam relacionados à conformidade com legislações de proteção de dados, dentre outras possibilidades.

A ontologia proposta baseou-se no reuso de conceitos previamente estabelecidos para a conformidade com a GDPR ou LGPD. Foi considerado o fato de essa legislação ser anterior à LGPD, e esta norma ter sido criada com base no normativo europeu. Além daqueles conceitos, são reutilizados conceitos de ontologias bem estabelecidas, como PROV-O, FOAF, CWM e Dublin Core para estabelecer uma base semântica rica. Os conceitos foram agrupados em 4 grandes grupos a saber: *Dataset Metadata, Operator Metadata, Law Compliance Metadata* e *Worflow Metadata*. O primeiro grupo trata de dados de proveniência que se referem ao conjunto de dados de processo. Já o segundo grupo trata metadados relacionados com os operadores aplicados nesses processos aos que os dados foram submetidos. O terceiro grupo contempla os conceitos que estão relacionados à conformidade com legislações de proteção de dados, como a LGPD.

A ontologia fornece uma estrutura que possibilita a resposta a questionamentos sobre os dados utilizados em determinado fluxo de processamento, incluindo as suas caraterísticas, a forma como foram processados e os algoritmos utilizados. Dentre esses questionamentos, é possível responder, por meio dos dados coletados de acordo com a estrutura proposta pela ontologia, (i) características sobre o conjunto de dados tratado, (ii) composição e execução do *workflow* e seus steps, (iii) os operadores e os parâmetros utilizados nos algoritmos aplicados e (iv) dados relativos à conformidade com as legislações de proteção de dados. O desenho da ontologia completa pode ser visualizado na Figura contida no Anexo A.

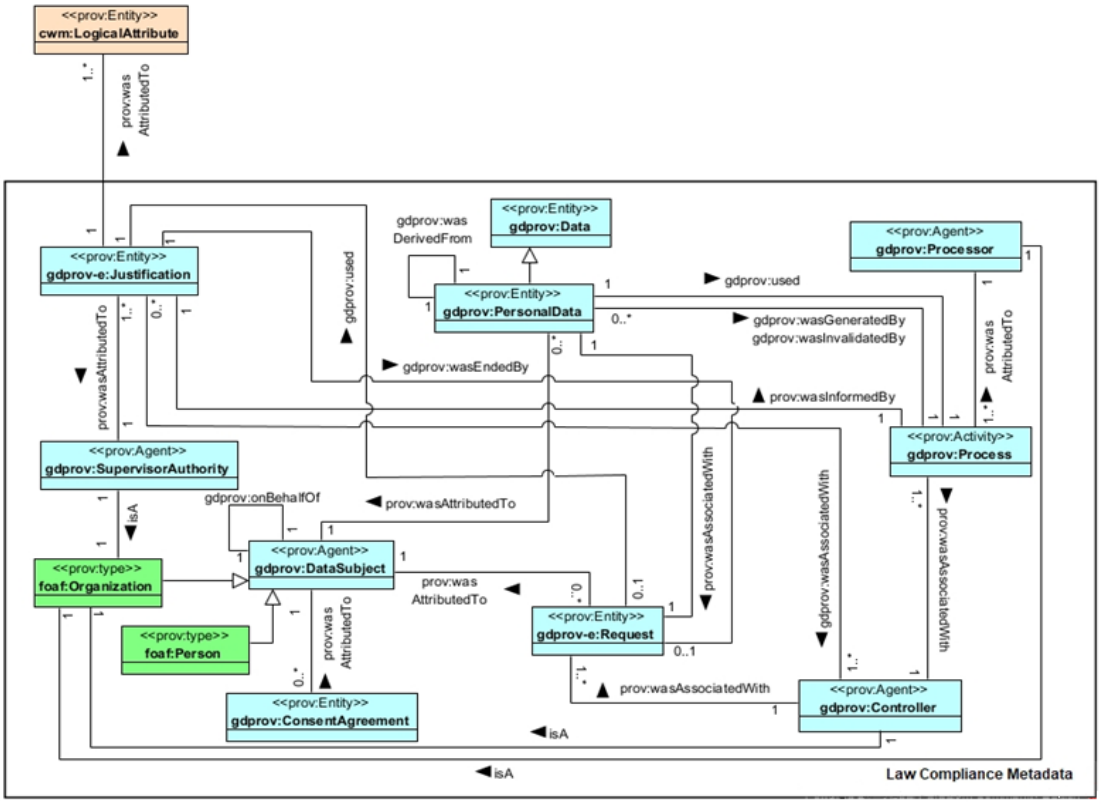
Os metadados coletados para se obter a conformidade legal sobre os processos aplicados aos dados tratados são orientados de acordo com a visão de indivíduos. As classes do grupo *Law Compliance Metadata* permitem a coleta de metadados voltados diretamente para a conformidade com a LGPD. Os titulares são representados *gdprov:DataSubject* especializado pela classe *foaf:Person*. A coleta do consentimento é representada pela classe *gd-prov:ConsentAgreement*, em que o indivíduo declara o consentimento relacionado as ações que seus dados podem sofrer. Quando um indivíduo responde por outro, esta relação é representada pela

associação *gdprov:onBehalfOf*. Os indivíduos podem ter instâncias de dados atribuídas pela classe *gdprov:PersonalData*, especialização da classe *gdprov:Data* para dados pessoais. A classe *gdprov:Process* permite a captura de metadados sobre o processo no qual os dados estão sendo tratados. Este grupo de classes e suas associações está ilustrado na Figura 2.

O grupo de classes *Operator Metadata* prevê classes que podem capturar informações sobre os operadores aplicados aos dados para prover alguma técnica de anonimização.

A classe *dmop:Algorithm* provê qual algoritmo foi aplicado com o software especificado pela classe *dc:Software*. A classe que especifica a categoria do operador a ser empregado é a *OperatorCategory.* Por meio da classe *LogicalAttributeCategory,* podemos realizar as relações necessárias para alcançar o atributo que de fato foi alterado com a técnica de anonimização empregada.

Figura 2 - Representação gráfica das conceitualizações da ontologia Sec4ML-O agrupadas como *Law Compliance Metadata.*



Fonte: Silva (2022).

2.1.2 CAPTURA DE DADOS DE PROVENIÊNCIA

No trabalho de Silva (2022) a captura de proveniência das tarefas empreendidas é realizada ao longo da execução delas. Como já citado, são coletadas informações sobre 4 grandes grupos de metadados relativos aos: conjunto de dado, os operadores utilizados sobre os dados, conformidade legal e metadados acerca da execução de tarefas sobre os dados. Para que a conformidade com a LGPD seja assegurada, qualquer sistema e/ou processo deve incorporar a captura dos metadados definidos nas classes e suas associações, principalmente, do terceiro e do quarto grupos. Nestes grupos podemos encontrar metadados que possibilitam responder aos principais questionamentos previstos como direitos do titular na LGPD (Brasil, 2018), conforme descrito a seguir.

> *Art. 18. O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição:*
>
> *I - confirmação da existência de tratamento;*

A coleta e armazenamento dos dados modelados nas classes *gdprov:Process* e *gdprov:PersonalData* poderão prover as informações necessárias para o atendimento do Art. 18, inciso I. A primeira classe possibilita informar o tipo de processo a que o dado foi submetido, quem é o responsável por esse processamento, quem é o controlador e a justificativa para o tratamento realizado. Por meio da segunda classe é possível identificar, além do dado em si, o tipo do dado, a declaração de consentimento e o indivíduo ao qual esse dado está relacionado.

> *IV - anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei;*

Caso seja aplicada alguma técnica de anonimização, as classes *DataAnonymizationOperator, OperatorCategory e dmop:Algorithm* podem fornecer dados para possibilitam responder ao questionamento previsto no Art.18, inciso IV.

Por meio dos metadados armazenados nestas classes podemos identificar as técnicas de anonimização empregadas nesses dados, as categorias dos operadores envolvidos e até mesmo qual o algoritmo utilizado na tarefa de anonimização.

> *V - portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa, de acordo com a regulamentação da autoridade nacional, observados os segredos comercial e industrial;*
>
> *VI - eliminação dos dados pessoais tratados com o consentimento do titular, exceto nas hipóteses previstas no art. 16 desta Lei;*

Os direitos expressos no Art. 18, incisos V e VI, podem ser atendidos utilizando-se os dados capturados pelas classes *foaf:Person, gdprov:DataSubject, gdprov:PersonalData, gdprov:Data*. Quanto ao direito de portabilidade, é necessário que seja possível relacionar todos os dados armazenados pelo titular dos dados para que este conjunto de dados seja transferido para um novo controlador. Os metadados garantidos pelas estruturas das classes já mencionadas possibilitam essa transferência. Utilizando-se das relações entre *gdprov:Process* e *gdprov-e:Justification*, podemos encontrar para qual ou quais atributos foram solicitados a portabilidade ou eliminação. Este tipo de solicitação é representado pela classe *gdprov-e:Request*. No caso de eliminação dos dados a tarefa de levantamento de todos os dados existentes se faz necessária da mesma forma. Entretanto, neste caso, não há a transferência dos dados, mas sim a sua eliminação ou expurgo da base de dados atual. Aqui é necessária uma ressalva quanto à pertinência desse direito. Em casos como os previstos no Art. 7, incisos I e II, a possibilidade de exclusão dos dados do titular fica impossibilitada, sob pena de impedimento de obrigações legais, regulatórias e/ou execução de políticas públicas.

> *VII - informação das entidades públicas e privadas com as quais o controlador realizou uso compartilhado de dados;*

As entidades que realizam uso compartilhado dos dados podem ser expressas pela classe *gdprov:Processor*. Além dos demais dados já evidenciados, essa classe agrega informações de qual entidade é responsável pelo tratamento dos dados de determinado titular.

> *IX - revogação do consentimento, nos termos do § 5º do art. 8º desta Lei.*
>
> *§ 3º Os direitos previstos neste artigo serão exercidos mediante requerimento expresso do titular ou de representante legalmente constituído, a agente de tratamento.*

Esse tipo de solicitação e outros podem ser representados pela classe *gdprov-e:Request*. Nesta classe são armazenadas informações sobre todos os tipos de requisição que o titular dos dados pode realizar, inclusive a revogação do consentimento.

> *§ 6º O responsável deverá informar, de maneira imediata, aos agentes de tratamento com os quais tenha realizado uso compartilhado de dados a correção, a eliminação, a anonimização ou o bloqueio dos dados, [...]*

A classe *gdprov:Processor* pode fornecer a relação de entes envolvidos em determinado tratamento dos dados, quer sejam controladores ou operadores, tornando possível que esses entes recebam a comunicação prevista no Art. 18, inciso IX, § 6º.

> *Art. 19. A confirmação de existência ou o acesso a dados pessoais serão providenciados, mediante requisição do titular:*
>
> *I - em formato simplificado, imediatamente; ou*
>
> *II - por meio de declaração clara e completa, que indique a origem dos dados [...],*

> *§ 1º Os dados pessoais serão armazenados em formato que favoreça o exercício do direito de acesso. [...]*
>
> *§ 3º Quando o tratamento tiver origem no consentimento do titular ou em contrato, o titular poderá solicitar cópia eletrônica integral de seus dados pessoais [...]*

A captura dos dados, utilizando-se das classes propostas e do armazenamento em um banco de dados que permita a rápida consulta e exportação dos dados requeridos, possibilita o atendimento do direito previsto no Art. 19, § 1º e § 3º.

3 PROGRAMA DE GESTÃO EM PRIVACIDADE

Na sequência das ações empreendidas pela MD para a adequação das ações de tratamento de dados pessoais no âmbito de seus sistemas e processos, foi desenvolvido o Programa de Gestão em Privacidade - PGP, com a finalidade de basear a adoção de estratégias que garantam o cumprimento dos normativos e boas práticas existentes relacionados com a proteção de dados.

Este documento tem por objetivo melhorar as atividades de tratamento de dados pessoais e promover melhorias com foco no cumprimento da legislação de proteção de dados. Surgiu da necessidade de especificar, mais detalhadamente, orientações sobre ações que visam a conformidade, visto que a LGPD entrou em vigor em 2020 com previsão de sanções administrativas a partir de agosto de 2021.

O programa ressalta a necessidade de definição, conforme contida na LGPD, para os papéis de controlador e operador. O controlador (Brasil, 2018), conforme definido no Art. 5º, inciso VI, é a natural ou jurídica, de direito público ou privado, a quem competem as decisões referentes ao tratamento de dados pessoais. Nos casos em que as pessoas jurídicas sejam de direito público, o controlador é a União, podendo ser descentralizadas as funções de controlador para os órgãos da administração pública. Já o papel de operador (Brasil, 2018), definido no Art. 5ª, inciso VII, é toda pessoa natural ou jurídica, de direito público ou privado, que realiza o tratamento de dados pessoais em nome do controlador. O operador é aquele que realiza de fato o tratamento dos dados pessoais definidos e delimitados pelo

controlador. Será sempre um indivíduo distinto do órgão, não atuando como profissional subordinado ou membro do Controlador.

O referido programa ressalta a natureza multidisciplinar e multisetorial dessa legislação, fato que exige a adoção de medidas em todas as fases do ciclo de vida do dado, assim definido naquele documento: (i) coleta - obtenção, recepção ou produção de dados pessoais, (ii) retenção - arquivamento ou armazenamento dos dados pessoais, (iii) processamento - classificação, utilização, reprodução, extração ou modificação destes dados, (iv) compartilhamento – transmissão, distribuição, comunicação, transferência ou compartilhamento dos dados pessoais e (v) eliminação - apagar, expurgar ou eliminar os dados pessoais. Os ativos de pessoas, bases de dados, documentos e sistemas são diretamente afetados pelas mudanças necessárias à conformidade com a LGPD.

O GT instituído em 2021 propôs então ações a empreender de curto, médio e longo prazos a fim de atingir a plena adequação do setor à LGPD. Dentre elas, algumas possuem potencial de poderem receber contribuições no sentido de maximizar a efetividade das atividades realizadas estão ressaltadas. Foram elencadas seis ações que podem receber contribuição com a aplicação da ontologia Sec4ML e que estão relacionadas na Tabela 3. O detalhamento das ações a empreender e as contribuições desse trabalho para cada uma delas será visto no Capítulo 4.

3.1 INVENTÁRIO DE DADOS PESSOAIS

No PGP há a definição da ação a empreender de curto prazo nº **28 - Realizar inventário de serviços e sistemas que tratam dados pessoais,** com a finalidade de realização da investigação no ecossistema de determinado sistema ou processo, apresentando os dados por eles tratados, as finalidades dos tratamentos recebidos por esses dados e tempo de retenção, além de possibilitar que sejam estabelecidos os métodos para atingir a conformidade a LGPD. Para tal atividade, o PGP prevê um modelo de inventário[7] que foi, originalmente, proposto no âmbito do Programa de Privacidade e Segurança da Informação (PPSI) como ferramenta para a implementação das melhores práticas nessas temáticas[8].

[7] Disponível em: https://www.gov.br/governodigital/pt-br/seguranca-e-protecao-de-dados/ppsi/template_inventario_dados_pessoais.xlsx
[8] https://www.gov.br/governodigital/pt-br/privacidade-e-seguranca/guias-e-modelos

Tabela 3 - Ações a empreender que podem receber contribuição da aplicabilidade da ontologia Sec4-ML-O.

Nº	Ação a Empreender	Prazo	Contribuição
10	Estabelecer medidas para assegurar que processos, serviços ou sistemas sejam projetados, desde a concepção e por padrão, em conformidade com a LGPD (Privacy by Design e Privacy by default)	médio	Adotar o modelo de dados de acordo com a ontologia Sec4ML
12	Implementar sistema de gestão de consentimentos e exercício dos direitos dos titulares	médio	Registrar o consentimento utilizando a estrutura de dados proposta de acordo com a ontologia
24	Implementar mecanismos para atender os direitos dos titulares de dados pessoais	curto	Uso dos dados de proveniência coletados
31	Gerar evidências para comprovar que tomou medidas de segurança técnicas e administrativas para proteger os dados pessoais	curto	Uso dos logs gerados pela aplicação em conjunto com os dados coletados de proveniência
36	Manter rastreabilidade dos dados pessoais em meio físico e digital	médio	Uso dos dados de proveniência coletados
42	Promover a manutenção de sistemas e serviços que tratam dados pessoais, a fim de adequá-los às normas atinentes ao tema	médio	Modificação do modelo de dados existente com base na ontologia e incremento dos processos de captura de proveniência

Fonte: Adaptado de Brasil (2023)

Como apontado por Reis (2023), de uma maneira geral, o pessoal envolvido em projetos de engenharia de *software* possui dificuldade em compreender o que deve ser feito para a adequação de seus sistemas com leis de privacidade, dificultando

a adoção de boas práticas da engenharia de *software*. Assim, o preenchimento e arquivamento do formulário é uma forma de documentar e controlar os registros dos tratamentos previstos que os dados podem receber, minimizando as dificuldades de implementação apontadas por esse autor.

O formulário possui uma estrutura que possibilita a descrição de informações tais como:

— os agentes de tratamento envolvidos e o encarregado;

— qual ou quais as finalidades para que os dados tenham sido coletados;

— as hipóteses de tratamento previstas nos arts. 7º e 11 da LGPD;

— a previsão legal que prevê a finalidade do tratamento que os dados receberão;

— a relação de dados pessoais tratados pela instituição;

— a categoria dos titulares dos dados pessoais;

— o tempo de retenção dos dados pessoais;

— instituições com as quais os dados pessoais serão compartilhados;

— transferência internacional de dados se aplicável conforme o art. 33 LGPD; e

— as medidas de segurança atualmente adotadas.

O formulário é estruturado nas seguintes seções: (i) **Identificação dos serviços / processos,** parte do inventário no qual possível informar o serviço ou processo que está sendo inventariado, qual seja que utilize ou não sistema informatizado, além de atribuir um identificador e data de criação; (ii) **Identificação dos agentes de tratamento envolvidos,** em que é possível informar os agentes de tratamento envolvidos, ou seja, o controlador, o operador e o encarregado; (iii) **Identificação da ação do operador ao longo do ciclo de vida de tratamento de dados pessoais,** em que podemos registrar em qual ou quais fases do ciclo de vida do tratamento de dados pessoais o operador atua. Esse levantamento auxilia na identificação de quais operações de tratamento são realizadas sobre os dados. Nos casos em que não há operador distinto do controlador, este assumirá o papel de operador; (iv) **Identificação da ação do operador ao longo do ciclo de vida de tratamento de dados pessoais,** na qual podemos coletar a descrição dos tratamentos aplicados aos dados. Os dados coletados nesta parte do inventário são fundamentais para os futuros atendimentos às demandas previstas para os titulares de dados no Capítulo III da LGPD; (v) **Identificação do escopo e da natureza de**

dados pessoais, em que coletados dados que definam a natureza e escopo dos dados pessoais tratados; (vi) **Finalidade do tratamento dos dados pessoais,** em que são coletados dados que definam a finalidade, hipótese de tratamento e previsão legal dos dados pessoais tratados. É importante ressaltar que a hipótese de tratamento e a previsão legal são informações imprescindíveis para o planejamento da conformidade dos dados tratados, condicionando até mesmo se há obrigatoriedade de coleta do consentimento; (vii) **Categoria dos Dados Pessoais**, na qual se inicia de fato o levantamento dos dados de maneira agrupada em 14 categorias de acordo com a temática dos atributos e (viii) **Categoria dos Dados Pessoais Sensíveis**, a última parte, em que são relacionados especificamente os dados sensíveis, conforme LGPD Art. 5, inciso II.

4 APLICABILIDADE DA ONTOLOGIA Sec4ML NO MINISTÉRIO DA DEFESA

Pesquisas apontam que a anonimização anterior à publicação dos dados é a abordagem mais promissora para a proteção da privacidade (Fung *et al.*, 2010). Entretanto, a busca por esta proteção deve ser balanceada no sentido de aplicar uma abordagem de anonimização que preserve também a utilidade dos dados para fins estatísticos, de pesquisa e apoio à decisão.

Mas há situações que os dados pessoais devem ser tratados individualmente, como as informações importantes de cada indivíduo. Nesse cenário, torna-se mais difícil a anonimização em virtude da possibilidade de não identificação e não caracterização do titular dos dados. Para esses casos, temos a possibilidade de aplicar modelos de privacidade diferencial (Brito e Machado, 2014). A utilização desses modelos busca adicionar ruídos nas saídas de consultas de modo que esses dados sejam perturbados garantindo a privacidade dos indivíduos. A utilização de técnicas desse tipo permite a inserção de ruídos na saída dos dados agrupados, de maneira a preservar suas capacidades probabilísticas. Entretanto, não é escopo deste trabalho analisar a aplicabilidade com relação às técnicas de privacidade diferencial.

Como apontado por Reis (2023), os principais desafios enfrentados no processo de adequação às legislações de proteção de dados se encontram na especificação ou coleta de requisitos. Ademais, o referido autor também ressaltou dificuldades específicas com requisitos baseados em legislações, pois pode haver mais de uma interpretação da referida norma.

Com o objetivo de mitigar os desafios apontados por Reis (2023), o programa em curso no MD, o PGP, como já visto no capítulo 3, aborda todas os aspectos necessários para a implantação de um processo de conformidade adequado, gradativo e que não interrompa os processos e sistemas em uso atualmente. Dentre as diversas ações a empreender apresentadas pelo programa e que foram destacadas na Tabela 1, em seis destas ações foi observado maior potencial de aplicabilidade da ontologia Sec4ML. São detalhados neste capítulo os aspectos principais da aplicabilidade da ontologia para o sucesso do empreendimento das ações citadas e uma análise detalhada dos dados que podem ser inventariados pelo setor, utilizando-se o inventário de dados, já detalhado na subseção 3.1, com a possível categorização quanto à preservação da privacidade e técnicas que podem ser aplicadas. Veremos o potencial de aplicabilidade com relação às ações do PGP a seguir.

A ação 10, "**Estabelecer medidas para assegurar que processos, serviços ou sistemas sejam projetados, desde a concepção e por padrão, em conformidade com a LGPD (*Privacy by Design e Privacy by default*)**", direciona as ações voltadas para todas as atividades necessárias para que novos sistemas sejam projetados em conformidade com a LGPD. Esta ação deve determinar quais atividades devem ocorrer na fase de desenho da arquitetura do sistema e de sua base de dados envolvida, se houver. A ontologia Sec4ML compreende todas as classes de metadados necessários para a captura da proveniência que permitirá responder aos questionamentos previstos na legislação no futuro. Assim, a adoção total da ontologia citada, no momento do design da solução proposta, asseguraria o atendimento dessa ação e empreender.

Analisando a ação 12, "**Implementar sistema de gestão de consentimentos e exercício dos direitos dos titulares**" e a ação 24, "**Implementar mecanismos para atender os direitos dos titulares de dados pessoais**", podemos observar dois pontos importantes: a coleta do consentimento e a garantia do exercício dos direitos dos titulares, expressos nos Art. 18 e 19 da LGPD. Há diversas classes relacionadas com esses dois pontos e elas foram evidenciadas na subseção 2.1.2.

Com relação à ação 31, "**Gerar evidências para comprovar que tomou medidas de segurança técnicas e administrativas para proteger os dados pessoais**", podem ser apontadas ações de uso de registros de atividades passadas conhecidos como "logs", em conjunto com os dados de proveniência coletados com o

uso da ontologia estudada. Nos registros de logs geralmente são coletados dados como dia e hora de determinado evento, assim como o usuário que realizou determinada tarefa e dados sobre identificação do equipamento utilizado para tal atividade, como endereço IP. O modelo de dados da ontologia Sec4ML, constante no Apêndice A do trabalho de Silva (2022), apresenta vários atributos que podem ser acrescentados a outros dados oriundos de registros de logs enriquecendo as evidências de que se trata a ação 31.

A ação 36, "**Manter rastreabilidade dos dados pessoais em meio físico e digital**", pressupõe que os sistemas considerem criar e manter uma estrutura de rastreio das atividades realizadas. As classes da ontologia Sec4ML agrupadas nos grupos *Operator Metadata* e *Worflow Metadata* permitem a captura de dados de rastreio que são muito importantes e que podem ser persistidos nas próprias bases de dados dos seus sistemas.

Com relação à ação 42, "**Promover a manutenção de sistemas e serviços que tratam dados pessoais, a fim de adequá-los às normas atinentes ao tema**", é a ação mais difícil de empreender pois implica na adaptação dos modelos de dados existentes, para que estes possam contemplar as estruturas de dados necessárias para a conformidade com a LGPD. Alterar sistemas em produção é sempre um desafio, considerando que o planejamento deve prever a coexistência de duas versões do sistema em paralelo. Este cenário requer uma gestão do projeto de desenvolvimento mais apurada e flexível, a fim de não inviabilizar o uso da versão anterior e, ao mesmo tempo, acrescentar ao sistema existente as capacidades que a conformidade requer. Da mesma forma que em ações anteriores, a adoção completa da ontologia pode ser um fator facilitador nesse processo.

No que concerne a uma análise detalhada dos dados com uma possível categorização quanto à preservação da privacidade e técnicas que podem ser aplicadas, foi realizada com base no Inventário de Dados Pessoais detalhado na subseção 3.1. A planilha eletrônica resultado dessa análise encontra-se no Apêndice A. Tal planilha foi adaptada do Inventário de Dados Pessoais e, como neste documento, os dados estão divididos em 15 agrupamentos detalhados a seguir, assim como as estratégias de anonimização sugeridas.

O grupo 7.1 **Identificação Pessoal** aborda dados de identificação, abrangendo também os gerados por instituições governamentais além de dados de identificação e localização eletrônica. Grupo formado somente por dados identificadores e semi-

identificadores, apresenta os dados com maior potencialidade de identificação do titular, pois são dados tais como nome, endereço, telefone e e-mail. Para os atributos identificadores, sugere-se que sejam aplicadas técnicas de substituição, permutação e adição de ruído. No caso de nomes a técnica de substituição de parte no nome é a mais aplicável por preservar a maior parte do nome do titular ao mesmo tempo que impede a sua identificação. No caso de permutações, os valores de registros diferentes são trocados entre si, impedindo também a identificação por meio do atributo ao qual será submetido o tratamento. Já na adição de ruído, mais aplicável a dados numéricos, são definidos valores a serem adicionados aos dados originais gerando as suas novas versões com a perturbação desejada.

No grupo 7.2 **Dados Financeiros,** temos os dados financeiros a serem armazenados. Dentro desse grande grupo somente os dados de identificação com as instituições bancárias, como números de contas bancárias e de cartões de crédito, necessitam de tratamento específico. Novamente, como são dados numéricos, sugere-se a aplicação de ruído como forma de evitar a identificação do titular dos dados. Entretanto, há dados que, mesmo não sendo identificadores, nem semi-identificadores e tampouco sensíveis, podem receber tratamento de mascaramento com a finalidade de preservar a segurança do titular. Dados dessa natureza podem revelar informações importantes sobre patrimônio e riqueza, informações que podem gerar mais risco à segurança do indivíduo e de sua família.

O grupo 7.3 **Características Pessoais** apresenta alguns dados semi-identificadores como idade e sexo e dados sensíveis como altura, peso e cor da pele e dos olhos. Dos dados semi-identificadores que fazem parte desse grupo, idade e sexo, como dados numéricos, podem ser perturbados utilizando-se adição de ruído, dificultando que este dado seja usado para, em combinação com outros dados, identificar um titular de dados. O atributo sexo é geralmente apresentado como uma combinação de representações seja F ou M, 1 ou 2 ou qualquer outra combinação de valores. Além disso, é um tipo de dado com pouca variabilidade dentro do conjunto de dados, o que dificulta o tratamento para anonimização. Apesar desses fatores, ainda é possível aplicarmos a técnica de substituição em que, por exemplo valores de F e M podem ser substituídos por outros valores como 10 e 20 ou letras como A e B sem semântica relacionada ao dado informado, dificultando o seu uso para facilitar a identificação de seu titular. O local de nascimento, por ser um dado informado em texto e que possui hierarquia de valores, pode sofrer generalização. Por exemplo, se

esse dado for informado pela cidade em que a pessoa nasceu, pode ser generalizado pelo estado de nascimento. Os dados que podem ser sensíveis se relacionados à dados de saúde são peso e altura. Sugere-se que ambos recebam adição de ruído. Atributos de cor dos olhos, de pele e outras características físicas podem ser relacionados com questões raciais e étnicas sendo, portanto, considerados sensíveis. Para eles, sugere-se a supressão de valores, em que determinados valores, dentro do espectro existente, são escolhidos e eliminados do conjunto de dados.

No conjunto 7.4 **Hábitos Pessoais** há os atributos que designam o consumo de álcool e tabaco, além de outros hábitos alimentares. Estes dados denotam características referentes à saúde dos titulares, senso assim classificados como sensíveis. Além disso aparecem também atributos que evidenciam questões de estilo de vida como consumo de bens e comportamentos de consumo. Se estes últimos estão relacionados com questões de saúde, como por exemplo, compras em redes de farmácia, também são considerados atributos ligados a informação de saúde. Pra todos eles sugere-se a aplicação de supressão de valores. No subgrupo de contatos sociais há atributos que podem identificar outras pessoas e empresas. Considerando-se que sejam semi-identificadores, sugere-se a aplicação de substituição de parte dos dados visando a não identificação das partes envolvidas. Assim, como dados de acidentes e incidentes podem estar relacionados a questões de saúde, também são classificados como sensíveis. Para estes, sugere-se a aplicação de supressão de células visando a garantia da privacidade.

No grupo 7.5 **Características Psicológicas**, se temos dados que definem personalidade ou caráter, podemos classificá-los como sensíveis. Sugere-se a aplicação de supressão de células. No grupo 7.6 **Composição Familiar** temos novamente dados de nomes que podem ser identificadores (nome do cônjuge) e semi-identificadores (nome de solteiro do cônjuge). Para ambos se sugere a aplicação da técnica de substituição para a troca de parte do nome. A data de casamento, por ser um atributo numérico semi-identificador, pode receber o tratamento de adição de ruído. De maneira análoga, o nome de cônjuge ou parceiro anterior pode receber tratamento de substituição.

No grupo 7.7 **Interesses de Lazer** não há dados a serem anonimizados. No grupo 7.8 **Associações**, se os dados referentes às associações se referem a organizações ou grupos políticos, temos mais um atributo sensível, para o qual sugere-se o tratamento de mascaramento.

O grupo 7.9 **Processos Judiciais e Administrativos**, apesar de não apresentar dados sensíveis, pode apresentar dados relacionados a questões de justiça. Sugere-se também, caso necessário para preservar a segurança do titular, a aplicação do tratamento de mascaramento dos dados.

Os grupos 7.10 **Hábitos de Consumo**, 7.11 **Dados Residenciais** e 7.12 **Educação e Treinamento** não apresentam dados que necessitem de tratamento.

O grupo 7.13 **Profissão e Emprego** apresentam dado descritivo do cargo e função do titular que pode ser semi-identificador. Para esses casos sugere-se a aplicação da técnica de generalização. Além desse atributo, há dados que informam o local de trabalho e tipo de empresa. Da mesma forma que outros atributos anteriores, por questões de segurança do titular, pode ser aplicável o tratamento de mascaramento.

O grupo 7.14 **Registros de vídeo, imagem e voz** refere-se a estes tipos de conteúdo que necessitam de tratamentos mais específicos considerando-se caso a caso. Entretanto, caso esses dados sejam utilizados com a intenção de identificar uma pessoa, deve ser considerado como semi-identificador e pode necessitar receber algum tipo de tratamento que garanta a privacidade das pessoas retratadas.

No último grupo relacionado na planilha contida no Apêndice A, 8. **Categorias de Dados Pessoais Sensíveis**, surgem todos as categorias definidas na LGPD Art. 5, inciso II. Considerando-se que todos estes atributos são sensíveis, sugere-se a aplicação da técnica de supressão de valores na medida do possível, garantindo a privacidade dos titulares.

O trabalho de Silva (2022) trouxe diversas outras contribuições para a questão da preservação da privacidade, incluindo: a implementação da abordagem em ferramenta de controle de fluxo de tratamento dos dados que possibilitou a realização de experimentos com dados de *benchmark,* a divulgação dos dados anonimizados nesses experimentos e os resultados dos experimentos de anonimização (utilizando os dados anonimizados e os dados não anonimizados) que demonstraram que não houve perda significativa na aplicação dos dados com algoritmos de classificação de aprendizado de máquina supervisionado. Os dados tratados nesse trabalho são dados semissintéticos de incidentes de segurança da informação. No referido trabalho, foram implementados somente três técnicas de anonimização aplicadas sobre os dados. Ademais, caso fosse do interesse do MD, como a ferramenta possibilita o tratamento de quaisquer conjuntos de dados, este ministério poderia utilizar todo ou parte da

implementação para a adaptação de seus sistemas e/ou processos no caso de utilizarem somente as técnicas implementadas por Silva. No caso de adotarem outras técnicas de anonimização, como sugerido nesta monografia, seria necessário a adaptação e/ou complemento da ferramenta utilizada no trabalho de Silva.

A Tabela de Categorização de Dados Pessoais apresentada no Apêndice A foi construída com base no Inventário de Dados Pessoais, já discutido na subseção 3.1. A estrutura dessa tabela traz duas grandes contribuições nas colunas Categorização e Sugestão de Tratamento. Na primeira, os dados apresentados são categorizados de acordo com a categorização apresentada por Brito e Machado (2014). Esta categorização é necessária pois baseia todas as demais ações a serem empreendidas sobre determinada categoria de dados, principalmente a adoção de determinada técnica de anonimização. Essa sugestão de qual técnica adotar aparece na segunda coluna da referida tabela, orientando o gestor de cada setor/sistema no MD sobre uma possibilidade de tratamento a ser dado que preserve a privacidade dos indivíduos.

Importante ressaltar que os esforços de adequação à conformidade com a LGPD serão sempre multidisciplinares, envolvendo o pessoal de governança, os setores de cada sistema/processo (negócio) e o pessoal de TI. A adoção total ou parcial das classes de dados e técnicas sugeridas nesta monografia possibilitará aos setores envolvidos em cada processo e/ou sistema atender, além da conformidade com a LGPD, a questões de transparência e garantir, para os sistemas novos, a aplicação do conceito de *Privacy by Design*. O MD está empreendendo a adequação com a LGPD iniciando pelo levantamento realizado por meio do Inventário e Mapeamento de Dados. Na sequência serão definidos os Termos de Uso, Avisos de Privacidade e a adequação dos contratos. Após, serão produzidos os Relatórios de Impacto e, por fim, realizada a fase de adequação dos sistemas a serem desenvolvidos. No presente momento, o setor ainda está trabalhando no inventário de dados pessoais e não há ainda previsão de uso de outras ferramentas no processo. Todas as sugestões apresentadas neste trabalho são subsídios para a implementação por equipe de TI e/ou desenvolvimento de sistemas conjuntamente com o pessoal de negócio de cada setor.

5 – CONCLUSÃO

Ao apresentar a aplicabilidade da ontologia Sec4ML-O no ambiente de governança em defesa, foi possível evidenciar como o uso da ontologia poderia trazer ganhos ao processo de adequação do MD à conformidade com a LGPD. A proposta apresentada neste trabalho reforça a importância da captura e gestão adequada dos dados de proveniência como uma forma de atender às exigências legais impostas pela LGPD. Um dos principais aspectos abordados neste estudo é a capacidade da Sec4ML-O de mapear e organizar os dados de forma que cada operação realizada, desde a coleta até o processamento e compartilhamento, seja captura e armazenada. Isso não só facilita a auditoria e o cumprimento das obrigações legais, como também proporciona maior transparência e rastreabilidade dos dados, elementos essenciais para o fortalecimento da governança em defesa.

Ademais, o uso da ontologia permitiria que o MD implementasse práticas de *Privacy by Design*, incorporando a proteção da privacidade desde a concepção dos sistemas. Além da adoção de abordagens como o *Privacy by Design*, a inclusão de técnicas de anonimização, conforme detalhadas neste trabalho, pode se tornar um fator fundamental para mitigar riscos e garantir que os sistemas do MD estejam alinhados com as melhores práticas em proteção de dados, assegurando essa proteção mesmo durante o compartilhamento entre diferentes entidades ou a divulgação pública para fins de transparência.

Também foi analisado o Programa de Preservação da Privacidade do MD e de que forma o uso da ontologia pode ser encaixar no programa hoje adotado. Abrangendo todos os aspectos necessários a essa adequação, esse programa fornece diversas ferramentas para auxiliar os envolvidos com cada sistema e/ou processo no MD a realizar as atividades previstas. O inventário de dados pessoais, parte integrante do referido programa, é o ponto de partida que possibilita o mapeamento dos dados tratados por cada sistema/processo, facilitando o apontamento de quais abordagens poderiam ser adotadas para cada categoria de dados mapeada.

A ontologia Sec4ML-O apresenta-se como uma solução abrangente e altamente aplicável, não apenas uma adequação às normas de proteção de dados, mas também um compromisso com a transparência, a segurança e o respeito aos direitos dos titulares de dados, assegurando a robustez dos sistemas e processos em um cenário cada vez mais exigente e interconectado. Assim, a sua adoção pode contribuir para o fortalecimento da governança em defesa, promovendo a segurança

e estabelecendo um padrão elevado de respeito à privacidade, tornando-se um elemento essencial para a modernização e a eficiência dos sistemas de defesa no Brasil.

REFERÊNCIAS

ALMEIDA, M. ***An unified approach to ontologies: Information Science, Computer Science and Philosophy***. In: Perspectivas em Ciência da Informacao. 2014. P. 242-258. Disponível em: https://www.researchgate.net/publication/287044064_An_unified_approach_to_ontologies_Information_Science_Computer_Science_and_ Philosophy. Acesso em: 23 agosto 2024.

BRASIL. **Lei Geral de Proteção de Dados**. Brasília, DF. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 25 abril 2024.

BRASIL. **Portaria GM-MD nº 5.148/2021**. Institui Grupo de Trabalho para propor a estrutura e as ações de Proteção de Dados Pessoais no âmbito da administração central do Ministério da Defesa. Disponível em: https://www.editoraroncarati.com.br/v2/Diario-Oficial/Diario-Oficial/PORTARIA-GM-MD-N%C2%BA-5-148-DE-14-12-2021.html. Acesso em: 22 julho 2024.

BRASIL. **Portaria Normativa GM-MD nº 3.572/2022.** Institui o Comitê de Governança Digital do Ministério da Defesa - CGD-MD e aprova a Implantação da Estratégia de Governo Digital - EGD no âmbito da administração central do Ministério da Defesa. Brasília, DF. Disponível em https://www.gov.br/defesa/pt-br/acesso-a-informacao/governanca-e-gestao/colegiados/reunioes/arquivos/reunioes - 2022/1o-reuniao-extraordinaria/portaria_gm_md_n__3-572__de_29_de_junho_de_2022___ portaria_gm_md_n__3-572__de_29_de_junho_de_2022___dou___imprensa_ nacional.pdf. Acesso em: 02 julho 2024.

BRASIL. **Programa de Gestão em Privacidade**. Brasília, DF. Disponível em: https://www.gov.br/defesa/pt-br/acesso-a-informacao/lei-geral-de-protecao-de-dados-pessoais-lgpd/arquivos/ProgramadeGestoemPrivacidade.pdf. Acesso em: 04 agosto 2024.

BRITO, F.; MACHADO, J. **Preservação de Privacidade de Dados: Fundamentos, Técnicas e Aplicações**. *In*: BETHLEM, N. Jornadas de Atualização em Informática. [*S. l.*]: Sociedade Brasileira de Computação, 2017. cap. 3, p. 70-87. Disponível em: https://www.researchgate.net/publication/318726149_Preservacao_de_Privacidade_de_Dados_ Fundamentos_Tecnicas _e_Aplicacoes. Acesso em: 22 julho 2024.

CAVOUKIAN, A. ***Privacy by Design The 7 Foundational Principles Implementation and Mapping of Fair Information Practices***. Information & Privacy Commissioner. Ontario, Canada. p. 12, 2006. Disponível em: https://privacy.ucsc.edu/resources/privacy-by-design---foundational-principles.pdf. Acesso em: 24 abril 2024.

CURADO JR., C. F.; CALEGARI, R. P.; DOWSLEY, R. B. **A lacuna normativa no tratamento de dados pessoais na defesa nacional e segurança do estado.** (Trabalho de Conclusão de Curso) - ESG. Escola Superior de Defesa. Rio de Janeiro: ESG, 2021. Disponível em: https://repositorio.esg.br/handle/123456789/

1538. Acesso em: 06 agosto 2024.

FUNG, B. C., WANG, K., FU, A. W.-C., and YU, P. S. ***Introduction to Privacy Preserving Data Publishing: Concepts and Techniques***. Chapman & Hall/CRC. Florida, United States of America. 2010. Disponível em: https://dl.acm.org/doi/10.5555/1855031. Acesso em: 06 agosto 2024.

GUARINO, N.; OBERLE, D.; STAAB, S. ***What is an ontology*****?** In: Handbook on Ontologies. Springer Berlin Heidelberg, 2009. p. 1–17. ISBN 978-3-540-70999-2. Disponível em: http://link.springer.com/10.1007/978-3-540-92673-3_0. Acesso em: 04 agosto 2024.

LEBO, T.; SAHOO, S.; MCGUINNESS D. **Prov-O:** ***The prov ontology***. W3C Recommendation, 2013. Disponível em: https://www.w3.org/TR/2013/REC-prov-o-20130430/. Acesso em: 02 agosto 2024.

HERSCHEL, M.; DIESTELKäMPER, R.; LAHMAR, H. B. ***A survey on provenance: What for? what form? what from?*** The VLDB Journal 26, 2017. p. 881–906. Disponível em: https://link.springer.com/article/10.1007/s00778-017-0486-1. Acesso em: 02 agosto 2024.

SILVA, M. L. **Sec4ML: anonimização de dados de incidentes de segurança da informação para tarefas de aprendizado de máquina**. (Dissertação de Mestrado) - IME. Programa de Pós-Graduação em Sistemas e Computação do Instituto Militar de Engenharia. Rio de Janeiro: IME, 2022. Disponível em: http://www.comp.ime.eb.br/pos/modules/files/dissertacoes/2022/2022-MadalenaLopes.pdf. Acesso em: 02 julho 2024.

REIS, D. B. de C. **Como as leis de privacidade e de proteção de dados impactam as atividades de engenharia de software: uma revisão sistemática da literatura. In:** (Dissertação de Mestrado) – UFRGS. Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande do Sul. Porto Alegre: UFRGS, 2023. Disponível em: https://repositorio.ufu.br/handle/123456789/36489. Acesso em: 01 julho 2024.

O TEMPO. Jornal. **GSI alertou governo sobre vazamento de dados do sistema de pagamentos**. Disponível em: https://www.otempo.com.br/politica/governo/gsi-alertou-governo-sobre-vazamento-de-dados-do-sistema-de-pagamentos-1.3443705. 2024. Acesso em: 24 abril 2024.

UNIÃO EUROPÉIA. **General Data Protection Regulation**. Disponível em: https://gdprinfo.eu/. Londres, Reino Unido. Acesso em: 25 abril 2024.

APÊNDICE A – Tabela de Categorização de Dados Pessoais

7 – Categorização de Dados Pessoais			
7.1 -Dados de Identificação Pessoal			
Grupo de Dados	Detalhamento de Atributos	Categorização	Sugestão de Tratamento
7.1.1 - Informações de identificação pessoal	Nome	Identificador	Substituição
	Endereço residencial	Semi-identificador	Mascaramento
	Histórico de endereços anteriores	Semi-identificador	Mascaramento
	Número de telefone fixo residencial	Semi-identificador	Adição de Ruído
	Número celular pessoal	Semi-identificador	Adição de Ruído
	e-mail pessoal	Identificador	Mascaramento
7.1.2 - Informações de identificação atribuídas por instituições governamentais	CPF	Identificador	Permutação
	RG	Identificador	Permutação
	Número do passaporte	Identificador	Permutação
	Número da carteira de motorista	Identificador	Permutação
	Número da placa de veículo	Semi-identificador	Truncagem
	Número de registro em conselho profissional	Identificador	Truncagem
7.1.3 - Dados de identificação eletrônica	Endereços IP	Identificador	Mascaramento ou criptografia
	Cookies	Semi-identificador	Mascaramento
	Data e hora de conexões	Semi-identificador	Mascaramento
7.1.4 - Dados de localização eletrônica	Dados de comunicação de Torres de celulares	Identificador	Adição de Ruído
	Dados de GPS	Semi-identificador	Adição de Ruído

7.2 -Dados Financeiros			
Grupo de Dados	Detalhamento de Atributos	Categorização	Sugestão de Tratamento
7.2.1 - Dados de identificação financeira	Números de identificação	Identificador	Adição de Ruído
	Números de contas bancárias	Identificador	Adição de Ruído
	Números de cartões de crédito ou débito	Identificador	Adição de Ruído
	Códigos secretos	Semi-identificador	Permutação
7.2.2 - Recursos financeiros	Dados sobre Renda	Não sensível	Mascaramento por segurança do titular
	Dados sobre Bens	Não sensível	Mascaramento por segurança do titular
	Investimentos	Não sensível	Mascaramento por segurança do titular
	Renda total e profissional	Não sensível	Mascaramento por segurança do titular
	Poupança	Não sensível	Mascaramento por segurança do titular
	Datas de início e término dos investimentos	Não sensível	Mascaramento por segurança do titular
	Receita de investimento	Não sensível	Mascaramento por segurança do titular
	Dívidas sobre ativos	Não sensível	Mascaramento por segurança do titular
7.2.3 - Dívidas e despesas	Aluguel	Não sensível	Mascaramento por segurança do titular
	Empréstimos	Não sensível	Mascaramento por segurança do titular
	Hipotecas e outras formas de crédito.	Não sensível	Mascaramento por segurança do titular
7.2.4 - Situação financeira (Solvência)	Avaliação do rendimento e	Não sensível	Mascaramento por segurança do titular
	Avaliação de capacidade de pagamento	Não sensível	Mascaramento por segurança do titular
7.2.5 - Empréstimos, hipotecas, linhas de crédito	Natureza e valor do empréstimo, saldo remanescente, data de início, período do empréstimo, taxa de juros, visão geral do pagamento, detalhes sobre as garantias	Não sensível	Mascaramento por segurança do titular
7.2.6 - Assistência financeira	Benefícios, assistência, bonificações, subsídios	Não sensível	Mascaramento por segurança do titular
7.2.7 - Detalhes da apólice de seguro	Natureza da apólice de seguro, detalhes sobre os riscos cobertos, valores segurados, período segurado, data de rescisão, pagamentos feitos, recebidos ou perdidos, situação do contrato	Não sensível	Mascaramento por segurança do titular
7.2.8 - Detalhes do plano de pensão	Data efetiva do plano de pensão, natureza do plano, data de término do plano, pagamentos recebidos e efetuados, opções, beneficiários	Não sensível	Mascaramento por segurança do titular
7.2.9 - Transações financeiras	Valores pagos e a pagar pelo titular dos dados, linhas de crédito concedidas, avais, forma de pagamento, visão geral do pagamento, depósitos e outras garantias	Não sensível	Mascaramento por segurança do titular
7.2.10 - Compensação	Compensações reivindicadas, valores pagos ou outros tipos de compensação	Não sensível	Mascaramento por segurança do titular
7.2.11 - Atividades profissionais	Natureza da atividade, natureza dos bens ou serviços utilizados ou entregues pela pessoa no registro, relações comerciais	Não sensível	---
7.2.12 - Acordos e ajustes	Acordos ou ajustes comerciais, acordos sobre representação ou acordos legais	Não sensível	---

7.3 -Características Pessoais			
Grupo de Dados	Detalhamento de Atributos	Categorização	Sugestão de Tratamento
7.3.1 - Detalhes pessoais	Idade	Semi-identificador	Adição de Ruído
	Sexo	Semi-identificador	Substituição
	data de nascimento	Semi-identificador	Adição de Ruído
	local de nascimento	Semi-identificador	Generalização
	estado civil	Não sensível	---
	nacionalidade	Não sensível	---
7.3.2 - Detalhes militares	situação militar	Não sensível	---
	patente	Não sensível	---
	distinções militares	Não sensível	---
7.3.3 - Situação de Imigração	Dados sobre visto	Não sensível	---
	dados sobre autorização de trabalho	Não sensível	---
	limitações de residência ou movimentação	Não sensível	---
	autorização de residência	Não sensível	---
7.3.4 - Descrição Física	Altura	Sensível se relacionado a questões de saúde	Adição de Ruído
	Peso	Sensível se relacionado a questões de saúde	Adição de Ruído
	Cor dos olhos	Sensível	Supressão de valores
	Cor da pele	Sensível	Supressão de valores
	Outras características	Sensível	Supressão de valores

7.4 – Hábitos Pessoais			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.4.1 - Hábitos	Uso de tabaco	Sensível	Supressão de valores
	Uso de álcool	Sensível	Supressão de valores
	Hábitos alimentares	Sensível	Supressão de valores
7.4.2 - Estilo de vida	Uso de bens ou serviços	Sensível se relacionado a questões de saúde	Supressão de valores
	Comportamento de consumo	Sensível se relacionado a questões de saúde	Supressão de valores
7.4.3 - Viagens e deslocamentos	Dados sobre residências anteriores	Não sensível	---
	Deslocamentos	Não sensível	---
	Visto de viagem	Não sensível	---
7.4.4 - Contatos sociais	Amigos e familiares	Semi-identificador	Substituição
	Parceiros de negócios	Semi-identificador	Substituição
	Relacionamentos com pessoas que não sejam familiares próximos	Semi-identificador	Substituição
7.4.6 - Denúncias, incidentes ou acidentes	Informações sobre um acidente ou incidente	Sensível	Supressão de células
	denúncia na qual o titular dos dados está envolvido,	Sensível	Supressão de células
	a natureza dos danos ou ferimentos	Sensível	Supressão de células
	Pessoas envolvidas	Identificador	Substituição
7.4.7 - Distinções	Distinções civis, administrativas ou militares	Não sensível	---
7.4.8 - Uso de mídia	Comportamento de uso de mídias e meios de comunicação	Sensível se revela aspectos de dados sensíveis	Supressão de células

7.5 – Características Psicológicas			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.5.1 - Descrição Psicológica	Dados sobre personalidade ou caráter	Sensível	Supressão de células

7.6 – Composição Familiar			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.6.1 - Casamento ou forma atual de coabitação	Nome do cônjuge ou companheiro(a)	Identificador	Substituição
	Nome de solteiro(a) do cônjuge ou companheiro(a)	Semi-identificador	Substituição
	Data do casamento	Semi-identificador	Adição de Ruído
	Data do contrato de coabitação	Não sensível	---
	Número de filhos	Não sensível	---
7.6.2 - Histórico conjugal	Casamentos ou parcerias anteriores, divórcios e/ou separações	Não sensível	---
	Nomes de cônjuges ou Parceiros anteriores	Identificador	Substituição
7.6.3 - Familiares ou membros da família	Outros familiares ou Membros da família	Não sensível	---

7.7 – Interesses de Lazer			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.7.1 - Atividades e interesses de lazer	*Hobbies* e/ou esportes	Não sensível	---

7.8 – Associações			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.8.1 Associações (exceto profissionais, políticas, em sindicatos ou qualquer outra associação que se enquadre em dados pessoais sensíveis)	Participação em organizações beneficentes, clubes, parcerias, organizações ou grupos	Sensível, se referente a organizações ou grupos políticos	Mascaramento

7.9 – Processo Judicial/Administrativo/Criminal			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.9.1 - Suspeitas	Suspeitas de violações, conexões conspiratórias com criminosos conhecidos.	Não sensível	Mascaramento por segurança do titular
	Inquéritos ou ações judiciais (civis ou criminais) empreendidas por ou contra o titular dos dados	Não sensível	Mascaramento por segurança do titular
7.9.2 - Condenações e sentenças	Dados sobre condenações e sentenças	Não sensível	Mascaramento por segurança do titular
7.9.3 - Ações judiciais	Dados sobre tutela, guarda temporária ou definitiva, interdição e/ou adoção	Não sensível	Mascaramento por segurança do titular
7.9.4 - Penalidades Administrativas	Multas	Não sensível	Mascaramento por segurança do titular
	Processo disciplinar	Não sensível	Mascaramento por segurança do titular
	Advertências, bem como qualquer outro tipo de penalidade ou sanção administrativa prevista em leis, normas e regulamentos.	Não sensível	Mascaramento por segurança do titular

7.10 – Hábitos de consumo			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.10.1 - Dados de bens e serviços	Dados sobre bens e serviços vendidos, alugados ou emprestados ao titular dos dados	Não sensível	---

7.11 – Dados residenciais			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.11.1 - Residência	Natureza da residência	Não sensível	---
	Imóvel próprio ou alugado	Não sensível	---
	Duração da residência nesse endereço	Não sensível	---
	Aluguel e outros custos	Não sensível	---
	Classificação da residência	Não sensível	---
	Detalhes sobre a avaliação do imóvel	Não sensível	---
	Nomes das pessoas que possuem as chaves	Identificador	---

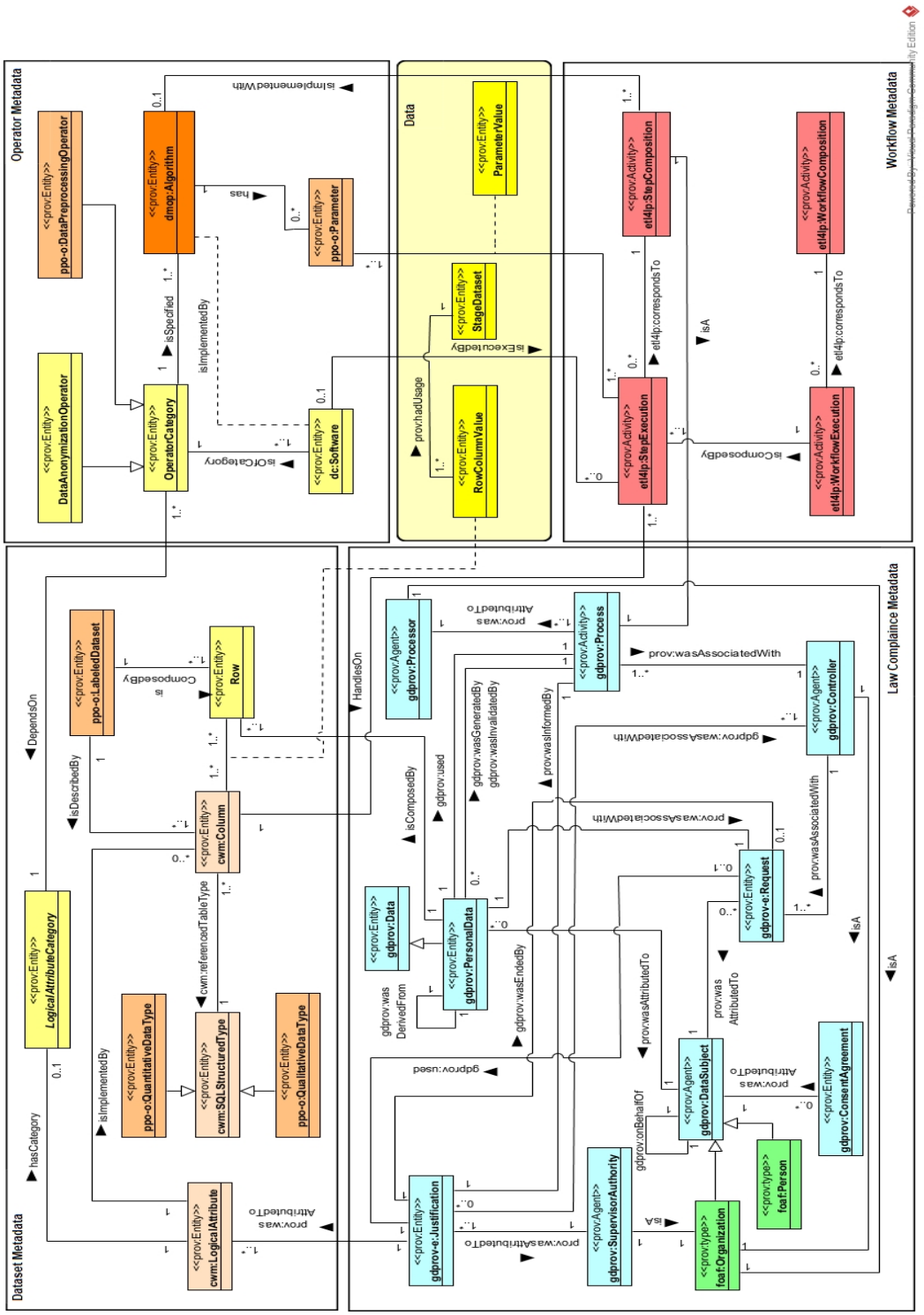
7.12 – Educação e treinamento			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.12.1 - Dados acadêmicos/escolares	Dados sobre diplomas	Não sensível	---
	Certificados obtidos	Não sensível	---
	Resultados de exames	Não sensível	---
	Avaliação do progresso dos estudos	Não sensível	---
	Histórico escolar	Não sensível	---
	Grau de escolaridade	Não sensível	---
7.12.2 Registros financeiros do curso/treinamento	Certificações profissionais, interesses profissionais, interesses acadêmicos, interesses de pesquisa e experiência de ensino	Não sensível	---

7.13 – Profissão e emprego			
Grupo de Dados	**Detalhamento de Atributos**	**Categorização**	**Sugestão de Tratamento**
7.13.1 - Emprego atual	Empregador	Não sensível	---
	Descrição do cargo e função	Semi-identificador	Generalização
	Data da contratação	Não sensível	---
	Local de trabalho	Não sensível	Mascaramento por segurança do titular
	Especialização ou tipo de empresa	Não sensível	Mascaramento por segurança do titular
	Modos e condições de trabalho	Não sensível	---
	Cargos anteriores	Não sensível	---
	Experiência anterior de trabalho no mesmo empregador	Não sensível	---
7.13.2 - Recrutamento	Data de recrutamento, método de recrutamento, fonte de recrutamento, referências e detalhes relacionados com o período de estágio	Não sensível	---
7.13.3 - Rescisão de trabalho	Data de rescisão, motivo, período de notificação, condições de rescisão	Não sensível	---
7.13.4 - Carreira	Emprego anterior e empregadores, períodos desempregado e serviço militar	Não sensível	---
7.13.5 - Absentismo e disciplina	Registros de absentismo, motivos de ausência e medidas disciplinares	Não sensível	---
7.13.6 -Avaliação de Desempenho	Avaliação de desempenho ou qualquer outro tipo de análise de qualificação ou habilidades profissionais	Não sensível	---

7.14 -Registros de vídeo, imagem e voz			
Grupo de Dados	Detalhamento de Atributos	Categorização	Sugestão de Tratamento
7.14.1 - Vídeo e imagem	Tratamento dado aos arquivos de vídeos, fotos digitais e fitas de vídeo	Semi-identificador caso a mídia possa identificar o titular	---
7.14.2 - Imagem de Vigilância	Tratamento de imagens e/ou vídeos de câmeras de segurança/vigilância	Semi-identificador caso a mídia possa identificar o titular	---
7.14.3 - Voz	Tratamento dado a arquivos digitais de voz e/ou outros registros de gravações de voz	Semi-identificador caso a mídia possa identificar o titular	---

8 - Categorias de Dados Pessoais Sensíveis		
Grupo de Dados	Categorização	Sugestão de Tratamento
8.1 - Dados que revelam origem racial ou ética	Sensível	Supressão de valores
8.2 - Dados que revelam convicção religiosa	Sensível	Supressão de valores
8.3 - Dados que revelam opinião política	Sensível	Supressão de valores
8.4 - Dados que revelam filiação a sindicato	Sensível	Supressão de valores
8.5 - Dados que revelam filiação a organização de caráter religioso	Sensível	Supressão de valores
8.6 - Dados que revelam filiação ou crença filosófica	Sensível	Supressão de valores
8.7 - Dados que revelam filiação ou preferências política	Sensível	Supressão de valores
8.8 - Dados referentes à saúde ou à vida sexual	Sensível	Supressão de valores
8.9 - Dados genéticos	Sensível	Supressão de valores
8.10 - Dados biométricos	Sensível	Supressão de valores

ANEXO A – Ontologia Sec4ML-O



Fonte: SILVA, M. L. **Sec4ML: anonimização de dados de incidentes de segurança da informação para tarefas de aprendizado de máquina**. p. 52. Disponível em: http://www.comp.ime.eb.br/pos/modules/files/dissertacoes/2022/2022-MadalenaLopes.pdf. Acesso em: 02 julho 2024.