UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS, CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Fábio Eder Cardoso

DADOS ABERTOS CONECTADOS NA PREVENÇÃO DE CRIMES: UMA CONTRIBUIÇÃO DA CIÊNCIA DA INFORMAÇÃO PARA A SEGURANÇA PÚBLICA

MARÍLIA - SP
2024

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS, CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Fábio Eder Cardoso

DADOS ABERTOS CONECTADOS NA PREVENÇÃO DE CRIMES:
UMA CONTRIBUIÇÃO DA CIÊNCIA DA INFORMAÇÃO PARA A SEGURANÇA PÚBLICA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP, campus de Marília, como requisito parcial para obtenção do título de Doutor em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

ORIENTADOR: PROF. DR. EDBERTO FERNEDA

MARÍLIA – SP
2024

C268d

Cardoso, Fábio Eder
Dados abertos conectados na prevenção de crimes: Uma contribuição da Ciência da Informação para a segurança pública / Fábio Eder Cardoso. -- Marília, 2024
184 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília
Orientador: Edberto Ferneda

1. Dados Abertos. 2. Segurança Pública. 3. Ciência da Informação. 4. Tecnologias Semânticas. 5. Predição Criminal. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade Estadual Paulista (UNESP), Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Impacto potencial desta pesquisa - (Dados abertos conectados na prevenção de crimes: Uma contribuição da Ciência da Informação para a Segurança Pública)

Este estudo propõe um modelo inovador de policiamento preditivo que pode transformar significativamente a área de segurança pública no Brasil. A aplicação de tecnologias semânticas combinadas com dados abertos tem o potencial de prever e prevenir atividades criminais eficientemente, contribuindo para uma sociedade mais segura. A transparência e a democratização das informações das instituições de segurança, podem reduzir os índices criminais. A integração dessas abordagens pode promover uma cultura de segurança pública mais colaborativa e baseada em evidências.

Potential Impact of the Research (Open Data in Crime Prevention: A Contribution from Information Science to Public Safety)

This study proposes an innovative model of predictive policing that could significantly transform the public safety sector in Brazil. The application of semantic technologies combined with open data has the potential to efficiently predict and prevent criminal activities, contributing to a safer society. Transparency and democratization of information from security institutions can reduce crime rates. Integrating these approaches may promote a more collaborative and evidence-based public safety culture.

Impacto Potencial de la Investigación (Datos Abiertos Conectados en la Prevención de Delitos: Una Contribución de la Ciencia de la Información para la Seguridad Pública).

Este estudio propone un modelo innovador de policía predictiva que podría transformar significativamente el área de seguridad pública en Brasil. La aplicación de tecnologías semánticas combinadas con datos abiertos tiene el potencial de predecir y prevenir actividades criminales de manera eficiente, contribuyendo a una sociedad más segura. La transparencia y la democratización de la información de las instituciones de seguridad pueden reducir los índices de criminalidad. La integración de estos enfoques puede promover una cultura de seguridad pública más colaborativa y basada en evidencias.

FÁBIO EDER CARDOSO

DADOS ABERTOS CONECTADOS NA PREVENÇÃO DE CRIMES: uma contribuição da Ciência da Informação para a Segurança Pública

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista "Júlio de Mesquita Filho" – UNESP, campus de Marília, como requisito parcial para obtenção do título de Doutor em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

BANCA EXAMINADORA

Prof. Dr. EDBERTO FERNEDA (Orientador)
Programa de Pós-Graduação em Ciência da Informação (PPGCI)
Universidade Estadual Paulista (UNESP)

Prof. Dr. CECILIO MERLOTTI RODAS
Programa de Pós-Graduação em Ciência da Informação (PPGCI)
Universidade Estadual Paulista (UNESP)

Prof. Dr. LEONARDO CASTRO BOTEGA
Programa de Pós-Graduação em Ciência da Informação (PPGCI)
Universidade Estadual Paulista (UNESP)

Prof. Dr. GUILHERME ATAÍDE DIAS
Departamento de Ciência da Informação
Universidade Federal da Paraíba (UFPB)

Prof. Dr. MARCOS LUIZ MUCHERONI
Escola de Comunicação e Artes (ECA)
Universidade de São Paulo (USP)

Marília, 22 de março de 2024

Dedico a Deus e a minha família por sempre estarem ao meu lado nos momentos mais difíceis deste processo.

A todos os meus professores, que foram de fundamental importância na construção da minha vida acadêmica e profissional.

Ao meu orientador Prof. Dr. Edberto Ferneda, pela sua paciência, conselhos e ensinamentos, que foram essenciais para o desenvolvimento deste projeto

AGRADECIMENTOS

Primeiramente quero agradecer à minha família, que me apoiou incansavelmente durante toda essa trajetória. O seu amor, compreensão e encorajamento foram a minha força motriz nos momentos mais desafiadores e cansativos. Vocês foram o meu porto seguro, me proporcionando conforto e tranquilidade para que eu pudesse focar totalmente em minha pesquisa.

Quero expressar, também, minha profunda gratidão ao meu orientador, Prof. Dr. Edberto Ferneda, cuja orientação, paciência e conhecimento profundos foram uma inspiração constante durante a elaboração desta tese. Seu compromisso com a excelência acadêmica e com a integridade intelectual me impulsionaram a superar todos os obstáculos que encontrei no decorrer desta jornada.

Também estendo meus sinceros agradecimentos ao Programa de Pós-Graduação em Ciência da Informação, que proporcionou um ambiente acadêmico estimulante e recursos essenciais para a realização desta pesquisa. A oportunidade de trabalhar dentro de um programa tão enriquecedor e colaborativo foi uma experiência verdadeiramente valiosa, que certamente moldou minha trajetória profissional e acadêmica.

Ao Grupo de Pesquisa GIHC, sob a orientação do Prof. Dr. Leonardo Botega, o meu muito obrigado por proporcionar um ambiente de pesquisa motivador e intelectualmente estimulante. As discussões enriquecedoras, os constantes desafios e a solidariedade inabalável do grupo foram fundamentais para o desenvolvimento e conclusão deste trabalho.

Um agradecimento especial à Polícia Militar, e em particular ao Tenente Anderson Garrido, pela colaboração, disposição em compartilhar conhecimentos e experiências práticas, e pelo suporte essencial à aplicação realista deste trabalho.

Este trabalho é o resultado de muitas mãos, mentes e corações. A todos que contribuíram, direta ou indiretamente, para a minha jornada acadêmica e para a realização deste projeto, minha mais sincera gratidão.

A todos vocês, o meu mais sincero muito obrigado!

Há algo maior que move a todos que fazem o caminho: o inusitado, a dimensão do sonho, o desejo de superação, a vontade de chegar ao destino almejado. A cada passo, as dificuldades vão se tornando motivos de júbilo, [e é o] que faz o caminho ter um sentido, que faz a nossa vida valer a pena: a de avançar sempre, superando-nos e às nossas inseguranças pela coragem de enfrentar o que ainda não conhecemos

Jussara Hoffman, (2001, p. 137).

RESUMO

No período atual, no Brasil, foi percebida a necessidade de abordagens tecnológicas mais eficazes na área de segurança pública. Nesse contexto, a Ciência da Informação desempenha um papel significativo, contribuindo de maneira eficaz ao sugerir métodos de cooperação e intercâmbio de informações em diversas áreas do conhecimento, alinhados com um dos seus principais objetivos: tornar a informação disponível e acessível. O objetivo deste trabalho é apresentar um modelo informacional voltado para o desenvolvimento de policiamento preditivo, resultante da combinação de dados abertos conectados com tecnologias semânticas, visando ao aprimoramento tecnológico na segurança pública da sociedade. Explora-se também como a transparência e a democratização das instituições de segurança pública podem contribuir para a redução dos índices criminais por meio de predições de ocorrências. A perspectiva de implementação prática nos impulsiona em direção a inovações tecnológicas na área de segurança. Este trabalho apresenta a interseção entre dados abertos, democracia, tecnologia da informação e segurança pública, com a visão de uma sociedade mais segura.

Palavras-chave: dados abertos; segurança pública; democracia; tecnologias semânticas; predições de crimes.

ABSTRACT

In recent years in Brazil, there has been a recognized need for more effective technological approaches in the field of public security. Within this context, Information Science plays a significant role, contributing effectively by suggesting methods of cooperation and information exchange across various knowledge domains. This aligns with one of its main goals: to make information available and accessible. The aim of this paper is to present an informational model for the development of predictive policing. This model results from the integration of open data with semantic technologies, aiming at technological enhancement in the public security sector of society. It also explores how the transparency and democratization of public security institutions can contribute to the reduction of crime rates through the prediction of occurrences. The prospect of practical implementation drives us towards technological innovations in the field of security. This paper presents the intersection of open data, democracy, information technology, and public security, envisioning a safer society.

Keywords: open data; public security; democracy; semantic technologies; crime predictions.

LISTA DE FIGURAS

Figura 1 – Formação dos dados governamentais abertos 36
Figura 2 – Modelo de e-democracia 40
Figura 3 – Interconexão dos dados por meio do *Linked Data* 74
Figura 4 – Distribuição 5 estrelas para Dados Abertos 83
Figura 5 – Dados disponibilizados sob licença aberta 84
Figura 6 – Dados processados por softwares proprietários 84
Figura 7 – Dados em formato aberto 85
Figura 8 – Dados utilizando URI (*Uniform Resource Identifie*r) 85
Figura 9 – Dados conectados permitindo a descoberta de informações relacionadas 86
Figura 10 – Alinhamento de dados via conexões RDF 95
Figura 11 – Estrutura do método 119
Figura 12 – Etapas do fluxo do processo KDD 123
Figura 13 – Tipos de classificação de ontologias e suas relações. 128
Figura 14 – Exemplo das triplas RDF 131
Figura 15 – Proposta de Ontologia 135
Figura 16 – Fluxo do desenvolvimento do sistema 140
Figura 17 – Modelo de Taxonomia genérica de crimes 142
Figura 18 – Taxonomia em formato de mapa mental 144
Figura 19 – Chamada da Biblioteca RDFLib 147
Figura 20 – Resultado obtido com a correlação de Pearson 148
Figura 21 – Mapa de calor de ocorrências e suas predições 149
Figura 22 – Mapa em grade com localização dos crimes ocorridos 150
Figura 23 – Mapa em grade com a predição criminal 151

LISTA DE TABELAS

Tabela 1 – Modelo de Taxonomia utilizado na aplicação 143

LISTA DE ABREVIATURAS E SIGLAS

ABNT..... Associação Brasileira de Normas Técnicas
API ..... Application Programming Interface
BD ..... Banco de Dados
BDTD..... Biblioteca Digital Brasileira de Teses e Dissertações
BIG..... Banco de Informação de Geração
CGU ..... Controladoria-Geral da União
CI..... Ciência da Informação
CP..... Community Policing
CRC..... Conselho Regional de Contabilidade
CSV..... Comma-separated values
DA..... Despesas Administrativas
DAC ..... Dados Abertos Conectados
DAG..... Dados Abertos Governamentais
DATA ..... Um tipo de dado referente as datas numéricas
DDACTS .. Data-Driven Approaches to Crime and Traffic Safety
DE ..... Diagnóstico de Enfermagem
DGA..... Dados Governamentais Abertos
DM ..... Data Mining
DO..... Despesas Operacionais
DOI ..... Digital Object Identifier
DOS..... Disk Operating System
DSR..... Design Science Research
DTD ..... Document Type Definition
EM..... Electron Multipliers
EPD ..... Encarregado de Proteção de Dados
ETL ..... Extração, Transformação e Carga
EUA ..... Estados Unidos da América
FBSP ..... Fórum Brasileiro de Segurança Pública
FOIA ..... Freedom of Information Act
G1..... Grupo Globo
GDPR..... General Data Protection Regulation
GIHC..... Grupo de Interação Humano-Computador
GIS ..... Geographic Information System
HTTP..... HyperText Transport Protocol
IA ..... Inteligência Artificial
IBICT ..... Instituto Brasileiro de Informação em Ciência e Tecnologia
ICT ..... Instituto de Ciência e Tecnologia
IEEE..... Institute of Electrical and Electronics Engineers
IETF ..... Internet Engineering Task Force
ILP..... Interledger Protocol
JSON ..... JavaScript Object Notation
KDD..... Knowledge Discovery in Data Base
LAI..... Lei de Acesso à Informação

LD ..... Linked Data
LDP ..... Linked Data Platform
LGD ..... Linked Government Data
LGPD ..... Lei Geral de Proteção de Dados
LOD ..... Linking Open Data
LODC..... Linked Open Data Cloud
LODD ..... Linked Open Drug Data
MAPS..... Mapping and Analysis for Public Safety
MD ..... Mineração de dados
ML..... Machine learning
NIJ..... National Institute of Justice
NIST..... National Institute of Standards and Technology
ODCAL .... Open Data Commons Attribution License
ODCODL . Open Data Commons Open Database License
OECD..... Organisation for Economic Co-operation and Development
OF ..... Open Format
OGP..... Open Government Partnership
OKF..... Open Knowledge Foundation
ONU..... Organização das Nações Unidas
OWL ..... Ontology Web Language
PC..... Personal Computer
pdf..... Portable Document Format
PIB ..... Produto Interno Bruto
PM..... Polícia Militar
PMESP..... Polícia Militar do Estado de São Paulo
POP ..... Problem-Oriented Policing
RDF..... Resource Description Framework
RE ..... Readable Machine
SBC..... Sociedade Brasileira de Computação
SKOS ..... Simple Knowledge Organization System
SPARQL ... SPARQL Protocol and RDF Query Language
TI..... Tecnologia da Informação
TIC ..... Tecnologias da Informação e Comunicação
TWG ..... Geospatial Technology Working Group
UCLA..... University of California
UE ..... União Européia
UN..... United Nations
UNESCO... Organização das Nações Unidas para a Educação, a Ciência e a Cultura
URI..... Uniform Resource Identifier
W3C ..... World Wide Web Consortioum
WWW ..... World Wide Web
XML..... EXtensible Markup Language

SUMÁRIO

1 INTRODUÇÃO 17
1.1 Problema de Pesquisa 20
1.2 Hipótese 21
1.3 Tese 22
1.4 Justificativa e Motivação 23
1.5 Objetivos 24
1.5.1 Objetivo Geral 24
1.5.2 Objetivos Específicos 25
1.6 Delimitação do Universo da Pesquisa 26
1.7 Procedimentos Metodológicos 27
1.7.1 Natureza, tipo e método de pesquisa 28
1.7.2 Metodologia de Pesquisa em Design Science (DSR) 30
1.8 Estrutura da Tese 31

2 DEMOCRACIA, TRANSPARÊNCIA E TECNOLOGIA 34
2.1 Governo Aberto 35
2.2 E-Democracia 38
2.3 A democratização das instituições públicas: Polícia Militar 42
2.3.1 História da Polícia Militar do Estado de São Paulo 44
2.4 Policiamento Preditivo 45
2.4.1 Policiamento proativo 47
2.4.2 Tecnologias 48
2.4.3 Rentabilização de recursos 48
2.5 Inteligência Policial Baseada em Dados 49

3 DADOS ABERTOS 52
3.1 O que são Dados Abertos 54
3.1.1 Relevância dos Dados Abertos 55
3.1.2 Disponibilidade e acesso 56
3.1.3 Benefícios 57
3.1.4 Questões de governança relacionadas aos Dados Abertos 59
3.1.5 Desafios 60
3.2 Legislação relacionada aos Dados Abertos 61
3.2.1 Direitos autorais e direitos de propriedade intelectual e industrial relacionados aos dados abertos 65
3.2.2 Direitos de privacidade e proteção de dados 67
3.2.3 Usos abusivos de dados abertos 68
3.2.4 Segurança de dados 70
3.3 Gerenciamento e compartilhamento de dados abertos 71
3.3.1 Armazenamento de dados abertos 72
3.3.2 Redes de dados 74
3.3.3 Tecnologias relacionadas a dados abertos 76
3.4 Considerações relevantes sobre esta seção 77

4 DADOS ABERTOS CONECTADOS 79
4.1 Definição e Contextualização 79
4.2 Padrões de representação e tecnologias utilizadas em Dados Abertos Conectados 82

4.3 Publicação em Dados Abertos Conectados 88
4.3.1 Fontes de Dados Abertos Conectados 88
4.3.2 Processos de Publicação de Dados Abertos Conectados 90
4.3.3 Ferramentas e Plataformas de Publicação em DAC 90
4.4 Conexão de Dados Abertos 91
4.4.1 Mapeamento de dados abertos 92
4.4.2 Alinhamento e reconciliação de dados abertos 94
4.4.3 *Linking Open Data* (LOD) 97
4.5 Uso e aplicação de dados abertos conectados 99
4.5.1 Análise de dados abertos conectados 100
4.5.2 Visualização de dados abertos conectados 103
4.5.3 Recomendação de dados abertos conectados 105
4.5.4 Outras aplicações de dados abertos conectados 107
4.5.5 Aplicações de Dados Abertos Conectados em Segurança Pública 109

5 DADOS ABERTOS CONECTADOS NA SEGURANÇA PÚBLICA 111
5.1 Levantamento de fontes de dados relevantes 114
5.1.1 Extração dos Dados Pluviométricos e de Vazão 117
5.1.2 Integração com Dados de Segurança Pública 118
5.2 Compreender e Identificar a estrutura dos dados 120
5.3 Padronização de dados 121
5.3.1 Knowledge Discovery in Data Base 122
5.4 Mapeamento da estrutura semântica dos dados com uso de tecnologias semânticas e dados conectados 126
5.4.1 Enriquecimento semântico dos dados por meio de ontologia 126
5.4.2 Classificação de Ontologias 127
5.5 Resource Description Framework (RDF) 130
5.6 SPARQL 132
5.7 *Extensible Markup Language* (XML) 133
5.8 Policiamento preditivo: sugestão de modelo informacional 134
5.9 Considerações sobre esta seção 139

6 Desenvolvimento da Aplicação para a predição de Crimes 140
6.1 Taxonomia de dados abertos em segurança pública 141
6.2 Benefícios do uso de taxonomia na segurança pública 144
6.3 Geração do Modelo RDF 145
6.4 Desenvolvimento do Sistemas Preditivo 146
6.5 Interfaces da aplicação 149

7 CONCLUSÃO 153

1 INTRODUÇÃO

O crime, como um fenômeno persistente na sociedade, não apenas desafia a ordem e a segurança pública, mas também instiga a contínua inovação e adaptação das estratégias de prevenção e controle (Sento-sé, 2011). É impossível evitar ocorrências criminais, entretanto, predizer alguns tipos de crimes, com o uso de informações e tecnologia, deixa de ser uma possibilidade remota e se torna uma realidade.

Em nosso país, o contexto sobre policiamento preditivo ainda é incipiente, uma vez que a segurança pública opera em contextos altamente complexos, nos quais cada ocorrência, mesmo sendo da mesma natureza, difere das demais. Essa imprevisibilidade ocorre em momentos cruciais de tomada de decisão, gerando uma grande volatilidade nas informações disponíveis.

Ultimamente, há um grande interesse no uso de aplicações que utilizam dados abertos conectados. Entretanto, na área da segurança pública, no Brasil, essa tecnologia é subutilizada, uma vez que não há um ente público que organize processos de integração e combinação de conjunto de dados heterogêneos para que o seu uso seja pleno (Lima, 2008; Lima Neto; Vieira, 2014). Os dados abertos conectados estão relacionados à disponibilização de informações de maneira acessível, interoperável e conectada por meio do uso de tecnologias semânticas (Isotani; Bittencourt, 2015). A aplicabilidade desses conceitos na segurança pública permite o aprimoramento nas tomadas de decisões, aumento na eficácia e na transparência da prestação de serviço, bem como, promover a cooperação entre diferentes órgãos e setores de segurança.

A segurança pública é um dos pilares fundamentais para a convivência pacífica e o desenvolvimento de uma sociedade. É dever do Estado e responsabilidade de todos prover um ambiente seguro para seus cidadãos, por meio da prevenção e repressão de crimes, além da promoção da ordem pública e do bem-estar geral. Nesse contexto, a disponibilidade de informações precisas e atualizadas é essencial para o planejamento, execução e avaliação das políticas de segurança pública (Lima Neto: Vieira, 2014).

A Ciência da Informação desempenha um papel crucial nesse contexto, uma vez que é responsável pelo estudo dos processos de produção, organização, recuperação e uso da informação (Souza; Alvarenga, 2004). A disponibilidade de dados abertos conectados na segurança pública representa uma nova fronteira para a Ciência da Informação, que se

molda e se desenvolve com o uso de novas metodologias e abordagens para lidar com essas informações conectadas.

Existem diversas fontes de dados que podem ser utilizadas na segurança pública, como estatísticas de criminalidade, cadastros de criminosos, informações sobre ações policiais, mapas criminais, dentre outros. No entanto, muitas dessas informações estão dispersas e fragmentadas, dificultando sua utilização e integração. A abertura e a conexão desses dados podem superar essas limitações, permitindo a análise integrada e em tempo real de múltiplas fontes de informação (Sauerwein *et al*., 2019).

Além disso, a aplicação de tecnologias semânticas pode enriquecer o significado dos dados, possibilitando a fusão e a análise de informações de diferentes fontes. A utilização de padrões e protocolos para a representação e o intercâmbio de dados, como o RDF (*Resource Description Framework*) e o SPARQL (*SPARQL Protocol and RDF Query Language*), permite criar uma rede de dados abertos que pode ser acessada e consultada por diferentes órgãos e instituições.

A utilização de dados abertos conectados na segurança pública pode trazer benefícios significativos para a sociedade. A análise integrada de informações sobre ocorrências criminais pode auxiliar na identificação de padrões de criminalidade e na adoção de ações preventivas mais eficazes. Além disso, a disponibilidade de informações em tempo real pode permitir uma resposta mais rápida a incidentes e emergências.

No entanto, é importante ressaltar que a utilização de dados abertos conectados na segurança pública também apresenta desafios. Dentre esses desafios, destacam-se a garantia da privacidade e proteção dos dados pessoais, a padronização e a qualidade dos dados, além da necessidade de capacitação e treinamento dos profissionais envolvidos na coleta, organização e análise dessas informações.

A disponibilização de dados abertos conectados na segurança pública representa uma nova abordagem para o uso da informação nesse contexto. A Ciência da Informação desempenha um papel fundamental para aproveitar todo o potencial desses dados, desenvolvendo metodologias e abordagens que permitem a utilização e integração dessas informações de forma eficiente e segura. A aplicação de dados abertos conectados na segurança pública tem o potencial de melhorar a tomada de decisões e promover um ambiente mais seguro e transparente para a sociedade.

Neste sentido, o esforço dedicado ao desenvolvimento da aplicação apresentada nesta tese não apenas reflete uma inovação tecnológica na área de segurança pública, mas

também simboliza uma parceria estratégica e pioneira entre o Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Unesp Marília e a Polícia Militar, por meio do décimo oitavo Batalhão do Interior, sediado no município de Presidente Prudente. Essa colaboração interinstitucional demonstra a importância da união de competências e recursos na busca por soluções eficazes para problemas complexos, como o combate e prevenção ao crime.

Esta parceria é muito significativa, pois combina o rigor acadêmico e a capacidade de pesquisa do programa com a experiência prática e o conhecimento específico da polícia na área de segurança pública. Essa sinergia possibilitou o desenvolvimento de uma aplicação inovadora, baseada em princípios de dados abertos conectados, que visa transformar a maneira como informações são utilizadas para a prevenção e o combate ao crime.

O modelo informacional adotado, fundamentado em Ontologia e taxonomias cuidadosamente elaboradas, é um dos aspectos mais valorosos deste trabalho. A criação de uma Ontologia específica para a área de segurança pública, acompanhada de taxonomias que organizam e categorizam os dados de maneira lógica e acessível, constitui uma base sólida para a análise e interpretação de informações complexas. Essa estruturação informacional não apenas facilita a compreensão dos dados, mas também potencializa a capacidade de previsão de eventos criminosos, permitindo intervenções mais assertivas e fundamentadas.

A adoção de Ontologias e taxonomias representa um avanço significativo no tratamento e na gestão de informações na área de segurança pública, possibilitando uma visão holística e integrada dos dados. Essa abordagem não somente melhora a qualidade das análises realizadas, mas também fomenta a interoperabilidade entre diferentes sistemas e instituições, promovendo uma cooperação mais efetiva no combate à criminalidade.

O desenvolvimento desta aplicação ressalta a relevância da Ciência da Informação enquanto campo disciplinar capaz de contribuir significativamente para a resolução de desafios sociais críticos. Por meio da aplicação de seus métodos e técnicas, a Ciência da Informação demonstra sua capacidade de inovar e de fornecer soluções práticas para o aprimoramento das políticas de segurança pública, enfatizando o papel crucial da gestão informacional na sociedade contemporânea.

Portanto, este trabalho não apenas contribui para o avanço científico e tecnológico na área de segurança pública, mas também destaca a importância das parcerias interdisciplinares e interinstitucionais na busca por soluções inovadoras e eficientes. A colaboração entre o PPGCI e a Polícia Militar de Presidente Prudente serve de modelo para futuras iniciativas, demonstrando como a união de esforços e conhecimentos diversificados pode levar ao desenvolvimento de ferramentas poderosas para o benefício da sociedade.

1.1 Problema de Pesquisa

O problema de pesquisa proposto aborda desafios multifacetados e complexos na utilização de dados abertos não conectados para melhorar a eficiência dos serviços prestados pela polícia militar. Primeiramente, enfrenta-se o desafio técnico e metodológico de coletar, integrar e organizar dados atualmente dispersos e não conectados de várias fontes, incluindo registros policiais, relatórios de ocorrências, bases de dados públicas, entre outros (Santos, 2021, p. 294).

Silva e Sena (2015, p. 45) corroboram para este problema:

> [...] a aposta na produção e na avaliação de diagnósticos como ferramentas para a tomada de decisão, gradualmente, vêm sendo incorporada à realidade institucional. Contudo, a adoção dessas medidas esbarra em um dos principais desafios na gestão policial: a produção, a organização, o processamento e a difusão das informações de maneira sistêmica e com foco no alcance de resultados.

Este aspecto envolve não apenas a conexão física de dados, mas também sua padronização e parametrização para assegurar consistência, interoperabilidade e relevância analítica. Segundo Ribeiro (2010),

> [...] A principal dificuldade para a operacionalização de estudos sobre o sistema de justiça criminal no Brasil diz respeito à inexistência de um sistema oficial de estatística que congregue informações sobre todas as fases, desde a policial até a do sistema prisional. Isso ocorre porque, nessa realidade, as bases de dados são fragmentadas, produzidas por cada organização que compõe o sistema, segundo a sua própria lógica e conforme os documentos que interessam a essa fase de processamento, sem a preocupação com o desdobramento desse dado nas fases posteriores (Ribeiro, 2010, p. 164).

Adicionalmente, o problema se estende para a compreensão e modelagem da influência da sazonalidade nos padrões criminais. Isso requer a análise de como variáveis temporais, como estações do ano, feriados, e até eventos esporádicos, impactam as taxas de criminalidade (Bruederle; Peters; Roberts, 2017; Vu Thuy Huong Le *et al.*, 2022). Esta

análise demanda uma abordagem multidisciplinar, combinando criminologia, estatística, e ciência de dados, para desenvolver modelos preditivos que considerem essas variações.

Além disso, busca-se desenvolver um sistema de policiamento preditivo que utilize dados para prever e identificar áreas e períodos de maior risco de ocorrências criminais. O objetivo é criar um modelo que permita às forças policiais alocar seus recursos de maneira mais eficaz, direcionando esforços preventivos e operacionais para locais e momentos específicos, com maior probabilidade de ocorrência de delitos.

Neste sentido, deve considerar os aspectos éticos na coleta e uso de dados, assegurando a proteção de informações sensíveis e respeitando os direitos individuais. Ademais, é imperativo avaliar a eficácia do modelo em termos de impacto real na prevenção e combate à criminalidade, além de sua capacidade de adaptação a mudanças nos padrões de criminalidade e na disponibilidade de dados.

De modo geral, o problema de pesquisa proposto não é apenas técnico, mas também estratégico e operacional, visando aprimorar a capacidade de resposta e prevenção da polícia militar em relação à criminalidade, por meio de uma abordagem inovadora e baseada em dados.

1.2 Hipótese

O presente trabalho apresenta a seguinte hipótese: O uso de dados abertos conectados conduz a uma melhoria na identificação do *modus operandi* criminal e contribui com a tomada de decisões estratégicas em segurança pública, onde:

a) A conexão de dados abertos permitiu a identificação e a predição de tendências criminais, bem como identificou os fatores que influenciam o comportamento dos criminosos.

b) Com o uso dos princípios da Ciência da Informação como, gestão, arquitetura e recuperação da informação, o processo de análise de dados pode ser mais simples.

c) Organizar os dados de forma adequada contribui na identificação das relações casuais entre o cometimento de crimes e, com isso, provê uma gestão estratégica em segurança pública. Isso inclui alocação de recursos mais eficazes, definição de prioridades para áreas de ação preventiva e implementação de estratégias de policiamento personalizadas.

Ao longo da investigação, foi realizado um estudo empírico que envolveu a coleta, organização e análise de dados reais. Técnicas e metodologias foram empregadas para testar a hipótese proposta. A eficácia do uso de princípios da Ciência da Informação e dados abertos e conectados na identificação de padrões criminais e na tomada de decisões estratégicas sobre segurança pública, foi avaliada.

Os resultados que foram obtidos com a implementação deste modelo perpassaram por análises estatísticas e foram comparados com outros métodos ou abordagens existentes. Isso permitiu confirmar ou rejeitar a hipótese proposta. Assim, confirmou-se que a integração de dados abertos em segurança pública, em conjunto com a Ciência da Informação, representou uma contribuição significativa para melhorar a identificação de padrões criminais, aprimorando a tomada de decisões estratégicas no campo da segurança pública.

1.3 Tese

O uso de estatísticas criminais pode não ser suficientemente eficaz para subsidiar tomadas de decisões em segurança pública. Assim, Lima (2008) afirma que:

> [...] Um dos temas que mais chamam a atenção na discussão sobre segurança pública no Brasil é, sem dúvida, a (in)existência de estatísticas criminais que permitam mensurar e subsidiar a tomada de decisões e o planejamento de políticas públicas eficientes e democráticas na área (Lima, 2008, p. 65).

Neste sentido, durante a década de 1970, a demanda por transparência nas decisões do governo e o uso de estatísticas criminais existentes para reestruturar o sistema de justiça criminal e reivindicar direitos levaram a tensões nos padrões e regulamentos de produção de dados.

> [...] só no final dos anos 1990 é que a adoção de ferramentas de georreferenciamento muda esse cenário e começa a indicar a utilidade e atualidade dos dados para planejamento operacional e tático, portanto, circunscritos às polícias (Lima, 2008, p. 3).

Portanto, a tese central deste estudo vincula-se a uma nova abordagem tecnológica e informacional baseada em como os dados abertos conectados em segurança pública, sob a ótica da Ciência da Informação, podem contribuir significativamente para o combate à criminalidade, diminuindo os índices e os indicadores criminais, como uso de predições de ocorrências.

Deste ponto de vista, pode-se especular que a distribuição e utilização efetivas de dados abertos conectados podem resultar em benefícios significativos para a segurança

pública. A organização, integração e análise desses dados são influenciadas de forma crucial pela Ciência da Informação, permitindo a identificação de padrões criminais, a compreensão da fenomenologia e a tomada de decisões apoiadas em evidências.

O uso de dados abertos conectados permite uma compreensão mais abrangente e integrada das questões relacionadas ao crime. Ao combinar várias fontes de dados, incluindo *logs* de eventos, dados demográficos, informações geoespaciais e outras fontes pertinentes, é possível obter predições mais precisas sobre padrões criminais, suas causas e variações sazonais.

Com base nesta tese, investigou-se como criar estratégias eficazes de prevenção ao crime. Por meio da análise integrada desses dados, foram identificadas áreas de alto risco, previu-se tendências criminais e direcionou-se recursos de forma mais assertiva, ajudando a reduzir a criminalidade e aumentar a sensação de segurança da sociedade.

1.4 Justificativa e Motivação

Com a disponibilidade de dados abertos conectados, principalmente dados governamentais, há a possibilidade de aprimorar estratégias de prevenção de crimes com foco no fortalecimento de políticas públicas de segurança. A Ciência da Informação, como uma área abrangente, desempenha um papel crucial neste contexto, oferecendo meios para contribuir significativamente com este cenário (Rautenberg; Burda; Souza, 2018).

Nos últimos anos, houve um aumento significativo na quantidade de dados disponíveis para acesso público que fornecem diversas informações úteis sobre vários aspectos da sociedade, incluindo índices criminais. No entanto, é crucial ir além da disponibilidade básica dos dados e investigar como eles podem ser efetivamente usados para melhorar a prevenção do crime e a segurança pública, em geral.

Rautenberg, Burda e Souza (2018) corroboram para o uso de dados públicos no suporte a pesquisa ao afirmar:

> [...] Web de Dados, se obtém os Dados Abertos Conectados, os quais são convertidos em dados para subsidiar as pesquisas. Desta forma, minimizam-se os esforços nos processos de aquisição, triagem, tratamento e utilização de dados primários. Por isso, é pertinente estabelecer medidas quanto à manutenção de Dados Abertos Conectados, fomentando a base informacional da Web de Dados (Rautenberg; Burda; Souza, 2018, p. 111).

O uso de dados abertos conectados na segurança pública possibilita uma visão integrada dos problemas relacionados à criminalidade. Ao conectar diferentes fontes de dados é possível obter informações mais detalhadas e robustas sobre os padrões criminais,

suas causas e suas variações sazonais. Dessa forma, é possível identificar áreas de maior risco, prever tendências criminais e direcionar o policiamento de forma mais efetiva.

Assim, pode-se considerar que a Ciência da Informação desempenha um papel fundamental no que se refere a seus princípios e abordagens, como a organização, a integração e a análise de dados, é possível explorar o potencial dos dados abertos conectados na prevenção de crimes. Ela oferece metodologias e técnicas que permitem a parametrização e a estruturação desses dados, facilitando sua interpretação e uso pelas instituições responsáveis pela segurança pública.

A relevância desta pesquisa se evidencia com base no desenvolvimento de estratégias eficazes de prevenção de crimes, sendo esta, uma das necessidades constantes na sociedade (Lima Neto; Vieira, 2014). A utilização de dados abertos conectados, aliada aos princípios da Ciência da Informação, oferece a possibilidade de avançar nesse campo, proporcionando melhorias na identificação de padrões criminais e na tomada de decisões estratégicas. Isso contribui diretamente para fortalecer a segurança pública, reduzir a criminalidade e promover uma sociedade mais segura, democrática e harmoniosa.

Assim, este trabalho se justifica pela importância de aproveitar o potencial dos dados abertos conectados, pela necessidade de aprimorar as estratégias de prevenção de crimes e pela contribuição que a Ciência da Informação pode oferecer nesse contexto. Ao explorar e compreender melhor essa temática, é possível proporcionar avanços significativos na segurança pública e contribuir para o bem-estar da sociedade como um todo.

1.5 Objetivos

1.5.1 Objetivo Geral

Desenvolver um modelo de policiamento preditivo que utilize dados abertos conectados e princípios da Ciência da Informação. Esse modelo buscará integrar e analisar os dados não conectados, considerando a sazonalidade e outros fatores relevantes, a fim de prever e identificar áreas com maior probabilidade de ocorrência de crimes.

A aplicação desse modelo de policiamento preditivo permitiu que a polícia militar otimize suas ações e recursos, direcionando-os de maneira mais efetiva para prevenção e combate ao crime. Além disso, ao disponibilizar os dados de forma conectada, organizada e parametrizada, será possível facilitar a tomada de decisões baseadas em evidências,

melhorar a compreensão dos padrões criminais e fortalecer a segurança pública, na totalidade.

Ao abordar o problema de pesquisa relacionado à utilização de dados abertos conectados na prevenção de crimes, considerando a influência da sazonalidade, espera-se contribuir para o avanço da Ciência da Informação e para a área de segurança pública. Através do desenvolvimento desse modelo de policiamento preditivo, será possível aprimorar a eficiência e a eficácia das ações policiais, proporcionando uma resposta mais rápida e efetiva aos desafios da criminalidade, e, consequentemente, contribuir para a construção de uma sociedade mais segura.

1.5.2 Objetivos Específicos

- **Identificar as fontes de dados abertos conectados relevantes para a segurança pública** - Neste objetivo específico, foi realizada uma busca e seleção das fontes de dados abertos que possuem informações relevantes para a análise e prevenção de crimes. Foram consideradas bases de dados governamentais, registros de ocorrências, informações demográficas, dados geoespaciais e outras fontes que pudessem contribuir para uma visão abrangente do cenário da criminalidade.
- **Analisar as técnicas de análise de dados aplicáveis nesse contexto -** Neste objetivo específico, foram estudadas e avaliadas as técnicas e metodologias de análise de dados mais adequadas para explorar as informações disponíveis, tendo sido consideradas as técnicas de: mineração de dados, aprendizado de máquina, análise estatística, modelagem preditiva, entre outras, considerando suas aplicações específicas para a prevenção de crimes e o fortalecimento da segurança pública.
- **Avaliar os impactos da aplicação desses dados na prevenção de crimes -** Este objetivo específico visa medir e avaliar os impactos da utilização dos dados abertos conectados na prevenção de crimes. Foram analisados os resultados obtidos a partir das análises realizadas, verificando-se a eficácia das estratégias e intervenções baseadas nesses dados. Foi considerado o grau de redução da criminalidade, a efetividade das ações preventivas, a otimização na alocação de recursos e outros indicadores que permitiram mensurar os benefícios da aplicação desses dados na segurança pública.

- **Propor um modelo de policiamento preditivo -** Neste objetivo específico, foi proposto um modelo de policiamento preditivo baseado nos dados abertos conectados e nas técnicas de análise estudadas. O modelo teve como objetivo antecipar a ocorrência de crimes, identificar áreas de maior risco e orientar ações preventivas. Foram considerados fatores como a sazonalidade, a localização geográfica, as características socioeconômicas e outros elementos relevantes para a previsão e prevenção de crimes.

Ao alcançar esses objetivos específicos, pretende-se contribuir para a melhoria da segurança pública por meio da aplicação de dados abertos conectados. A identificação de fontes de dados relevantes, a análise adequada dessas informações, a avaliação dos impactos e a proposição de um modelo de policiamento preditivo serão passos importantes para aprimorar as estratégias de prevenção de crimes e fortalecer a segurança da sociedade.

1.6 Delimitação do Universo da Pesquisa

A presente pesquisa teve um escopo específico, sendo importante delimitar o universo de estudo para um melhor direcionamento das análises e resultados. Dessa forma, as seguintes delimitações foram consideradas:

- **Crimes que envolvem sazonalidades de clima -** A pesquisa teve foco nos crimes que apresentam influência das sazonalidades climáticas. Foram analisados os padrões criminais que pudessem estar associados a variações climáticas, como o aumento ou diminuição da criminalidade em determinadas estações do ano. Foi investigado como essas variações podem impactar os índices de criminalidade e como a prevenção de crimes pode ser aprimorada considerando essas sazonalidades.
- **Furto:** A pesquisa foi direcionada ao crime de furto, enfocando esse delito em particular. Foram examinados os casos de furto, levando em consideração suas características específicas, como os fatores de risco, os perfis dos infratores e das vítimas, e as possíveis flutuações sazonais desse tipo de crime. Investigou-se como os dados abertos e a Ciência da Informação podem auxiliar na prevenção e redução do furto.
- **Microrregião da área de Presidente Prudente -** A pesquisa teve como âmbito geográfico uma microrregião específica, localizada na área do município de Presidente Prudente. Segundo o Censo de 2022, sua população é de 225.668 habitantes. A área territorial do município é de 560,637 km², com uma densidade

demográfica de 402,52 habitantes por km². O município conta com distritos como Ameliópolis, Eneida, Floresta do Sul e Montalvão, subdivididos em 255 bairros. O Índice de Desenvolvimento Humano Municipal (IDHM) de Presidente Prudente, conforme dados de 2010, é de 0,806, o que é considerado muito alto. A cidade possui uma importante tradição cultural e é um dos principais polos industriais, culturais e de serviços do oeste de São Paulo (IBGE, 2023). Essa delimitação permitiu uma análise mais detalhada e contextualizada dos dados, considerando as características socioeconômicas, demográficas e geográficas da região. Foram consideradas as informações disponíveis nessa microrregião para a identificação de padrões criminais e para o desenvolvimento do modelo de policiamento preditivo.

Com essas delimitações, buscou-se concentrar o estudo em uma área específica e em aspectos relevantes, como a influência das sazonalidades climáticas e a violência doméstica. Essa abordagem permitiu uma análise mais aprofundada e direcionada, visando contribuir de forma mais efetiva para a prevenção de crimes e o fortalecimento da segurança pública, na referida microrregião.

1.7 Procedimentos Metodológicos

A ciência e a tecnologia se complementam por meio de um processo construtivo do conhecimento e esse processo caminha por vias da comunicação e da informação.

De acordo com Garvey (1979),

> [...] as atividades associadas com a produção, disseminação e uso da informação, perpassando desde a hora em que o cientista teve a ideia da pesquisa até o momento em que os resultados de seu trabalho são aceitos como parte integrante do conhecimento científico.

Um pesquisador precisa, além do conhecimento do assunto, possuir a curiosidade, criatividade, integridade intelectual e sensibilidade social. São igualmente importantes a humildade, disciplina, perseverança e paciência (Gil, 2008).

A presente pesquisa pretendeu trazer uma compreensão da realidade no contexto da segurança pública e Ciência da Informação, contribuindo para o avanço do conhecimento científico nos campos da gestão, arquitetura e recuperação da informação, do fluxo informacional e do uso da informação.

1.7.1 Natureza, tipo e método de pesquisa

A presente pesquisa utiliza-se de uma abordagem qualitativa, que, sendo esta uma abordagem metodológica enfatiza a compreensão profunda dos fenômenos sociais dentro do contexto em que ocorrem. Esta abordagem valoriza o processo de pesquisa tanto quanto, ou mais do que, os resultados obtidos. Nessa perspectiva, o pesquisador se imerge no ambiente natural onde o fenômeno é observado, coletando dados diretamente da fonte. Esta imersão possibilita uma compreensão mais rica e detalhada dos aspectos sociais, culturais e humanos envolvidos. A natureza descritiva da pesquisa qualitativa permite que os dados sejam analisados de forma indutiva, construindo teorias e compreensões a partir das observações feitas no campo.

“Em segundo lugar, apesar de haver afirmado que a dimensão teórica da pesquisa qualitativa seria dada pelo pesquisador, devemos afirmar, sem que isto se constitua numa proposição essencial, que o tipo de pesquisa qualitativa denominada "pesquisa participante" (ou "participativa") pode prestar-se melhor a um enfoque dialético, histórico-estrutural que tenha por objetivo principal transformar a realidade que se estuda” (Triviños, 1987, p. 125).

A metodologia *Design Science Research* (DSR) compartilha várias características com a abordagem qualitativa descrita por Triviños (1987), especialmente no que se refere à ênfase na compreensão profunda dos fenômenos dentro de seus contextos naturais e na valorização do processo de pesquisa. Assim como na pesquisa qualitativa, onde o pesquisador se imerge no contexto natural para coletar dados e construir teorias a partir da observação direta, o DSR foca no desenvolvimento e na análise de artefatos (como modelos, métodos ou sistemas) para resolver problemas práticos e complexos. Esta abordagem de pesquisa é iterativa e reflexiva, envolvendo a criação e o teste de inovações no mundo real. Similarmente à pesquisa participante destacada por Triviños, o DSR não se limita a uma compreensão teórica, mas busca transformar a realidade ao interagir diretamente com ela. Os artefatos criados no DSR são avaliados em contextos reais, e as lições aprendidas são utilizadas para refinar tanto os artefatos quanto as teorias subjacentes.

Na aplicação da metodologia *Design Science Research* (DSR) a este projeto, a natureza exploratória e descritiva da pesquisa se alinha perfeitamente com o objetivo de criar um modelo informacional inovador para a conexão de dados abertos. Este modelo visa facilitar a tomada de decisões eficazes na redução dos índices criminais. Nesse sentido, a utilização de metodologias que permitam a compreensão, normalização e ligação eficientes de dados abertos é necessária devido à crescente disponibilidade deles. Os dados

conectados demonstraram ser uma abordagem promissora que permite a interoperabilidade e o enriquecimento de conjuntos de dados através da utilização de tecnologias semânticas.

Nesta tese, a metodologia pretendida abrangeu várias etapas cruciais. Em primeiro lugar, foi necessário identificar as fontes de dados pertinentes para o projeto, procurando em portais de dados abertos públicos e outras fontes especializadas. Esta fase de identificação permitiu a combinação de vários conjuntos de dados díspares. Assim, foi necessário:

- **Realizar um levantamento abrangente das possíveis fontes de dados relevantes para o projeto** - Explorar plataformas de dados abertos como o "data.gov", *sites* de agências governamentais, institutos de pesquisa e órgãos reguladores, *sites* especializados e outras fontes relevantes para encontrar conjuntos de dados disponíveis. Avaliar a qualidade, confiabilidade e relevância das fontes de dados identificadas, bem como sua persistência, no sentido de que os dados estarão sempre disponíveis.
- **Compreender e Identificar a Estrutura dos Dados** - Analisar as estruturas de dados disponíveis nas fontes identificadas, que podem ser em formatos do tipo CSV, JSON, XML ou outros. Utilizar técnicas de análise exploratória de dados para compreender a estrutura dos conjuntos de dados. Identificar as variáveis, atributos e relacionamentos presentes nos conjuntos de dados.

Uma vez determinadas as fontes de dados, o passo seguinte consistiu em compreender e determinar a estrutura destas coleções. Isto envolveu a análise dos vários formatos em que os dados podem ser disponibilizados e a exploração das variáveis e relações inerentes aos conjuntos de dados. Esta compreensão foi crucial para a etapa seguinte, o arquivo de dados.

- **Padronização dos dados -** O objetivo da padronização dos dados é garantir a interoperabilidade entre vários conjuntos de dados. Isto significa que deve ser definido um modelo de dados normalizado ou devem ser adotadas normas existentes, como as da *Open Knowledge Foundation* (OKF). Além disso, foi necessário desenvolver transformações ou *scripts* que permitam converter os formatos de dados encontrados para o formato preferido. Este passo garante que os dados são consistentes e podem ser facilmente integrados e utilizados em muitos contextos.

- **Mapeamento da estrutura semântica dos dados utilizando tecnologias semânticas e de dados conectados -** Isto envolve a utilização de recursos como RDF (*Resource Description Framework*), OWL (*Ontology Web Language*) e SKOS (*Simple Knowledge Organization System*), para expressar os conceitos e as relações existentes nos dados. O desenvolvimento de ontologias ou vocabulários controlados permite o estabelecimento de relações claras entre diversos conglomerados de dados, possibilitando a conexão e ampliação das informações.

1.7.2 Metodologia de Pesquisa em Design Science (DSR)

A metodologia de *Design Science Research* (DSR) foi particularmente adequada para este projeto, dado o seu foco na criação de artefatos práticos que solucionam problemas específicos. No contexto da segurança pública e Ciência da Informação, a DSR permite o desenvolvimento de modelos informacionais que são não apenas teoricamente sólidos, mas também praticamente aplicáveis.

O primeiro passo na aplicação da DSR foi uma análise detalhada do problema de pesquisa. Isso envolveu a identificação das lacunas no conhecimento existente e a compreensão das necessidades específicas no campo da segurança pública. A análise abrangeu tanto a teoria quanto as práticas atuais, identificando áreas onde a gestão e análise de dados abertos podem ser otimizadas.

Após a identificação do problema, o próximo passo foi o desenvolvimento de artefatos. Neste caso, os artefatos podem incluir modelos de dados, algoritmos, ou sistemas de informação que facilitam a gestão eficiente e a análise de dados abertos na segurança pública. Este processo foi iterativo, envolvendo prototipagem e refinamento contínuo.

Uma vez desenvolvidos, os artefatos foram testados em cenários do mundo real para demonstrar sua utilidade e eficácia. Este processo de demonstração ajudou a validar os modelos e sistemas propostos. A avaliação incluiu tanto critérios qualitativos quanto quantitativos, assegurando que os artefatos pudessem atender aos objetivos definidos.

A metodologia DSR é inerentemente iterativa. Com base no *feedback* e nos resultados das fases de demonstração e avaliação, os artefatos serão continuamente aprimorados. Este ciclo de aprimoramento garante que os artefatos permaneçam relevantes e eficazes diante das mudanças nas necessidades e nos contextos da segurança pública.

Uma parte crucial da DSR é a comunicação eficaz dos resultados da pesquisa. Isso inclui não apenas a publicação em periódicos acadêmicos, mas também a disseminação dos resultados para partes interessadas no campo da segurança pública. Esta comunicação ajudará a garantir que os artefatos desenvolvidos sejam adotados e aplicados de maneira eficaz.

A pesquisa em DSR busca contribuir tanto para a teoria quanto para a prática. No contexto deste projeto, espera-se que o trabalho contribua para o corpo teórico da Ciência da Informação, especialmente no que diz respeito à gestão de dados abertos. Praticamente, espera-se que os artefatos desenvolvidos tenham um impacto significativo na melhoria das estratégias de segurança pública.

Por fim, é essencial considerar aspectos éticos e de sustentabilidade na aplicação da DSR. Isso inclui garantir a proteção de dados pessoais, considerando o impacto ambiental das soluções tecnológicas propostas e garantindo que os artefatos sejam acessíveis e utilizáveis para uma ampla gama de usuários dentro do campo da segurança pública.

Assim, a aplicação da metodologia DSR nesta pesquisa oferece uma abordagem sistemática e rigorosa para o desenvolvimento de soluções inovadoras e eficazes que atendam as necessidades complexas e dinâmicas da segurança pública e da Ciência da Informação.

Ao utilizar esta metodologia para trabalhar com dados abertos e conectados, é possível obter uma compreensão mais abrangente e integrada das informações disponíveis, possibilitando análises mais profundas. Além disso, esta estratégia auxilia no desenvolvimento de uma infraestrutura de dados mais resiliente e adaptável, capaz de atender às necessidades das demandas cada vez mais complexas e interconectadas.

Desta forma, a pesquisa justifica-se pela contribuição com o órgão público de segurança, bem como pela importância de aproveitar ao máximo todo o potencial dos dados abertos conectados, pela necessidade de apresentar modelos preditivos para aprimorar e criar estratégias de prevenção criminal e, por fim, pelo potencial contributo que a Ciência da Informação pode oferecer nesta situação.

1.8 Estrutura da Tese

A proposta para a estrutura da presente tese está organizada em quatro seções, a fim de abordar de forma coerente e abrangente o tema da aplicação de dados abertos conectados

na prevenção de crimes, sob a perspectiva da Ciência da Informação. A estrutura da Tese é a seguinte:

- Seção 1: Introdução e Metodologia

Esta seção compreende a introdução ao tema da pesquisa, apresentando o contexto, a problemática, a justificativa, o objetivo geral e específicos, além da delimitação do universo de estudo. Também está sendo apresentada a metodologia adotada na pesquisa, descrevendo os procedimentos utilizados, a coleta e análise de dados, e as técnicas empregadas na obtenção dos resultados. A seção 1 fornece uma visão geral da Tese e estabelece as bases para as seções subsequentes.

- Seção 2: Democracia, Transparência e Tecnologia

Nesta seção, foi realizada uma revisão teórica e conceitual sobre o tema da democracia. Foram explorados os principais fundamentos, princípios e características da democracia, bem como sua relação com a segurança pública e a prevenção de crimes.

- Seção 3: Dados Abertos

Nesta seção foi abordado o conceito e a importância dos dados abertos. Foram explorados os fundamentos e princípios dos dados abertos, destacando sua relevância para a transparência, a participação cidadã e o fortalecimento da governança democrática. Foram apresentados exemplos de iniciativas de dados abertos em diferentes contextos e setores, demonstrando seu potencial para aprimorar a tomada de decisão e promover a inovação social. Foram discutidos também os desafios e limitações relacionados à disponibilização e uso de dados abertos.

- Seção 4: Dados Abertos Conectados

Na seção 4 foi aprofundado o tema dos dados abertos conectados. Foram exploradas as estratégias, tecnologias e padrões utilizados para conectar e integrar os dados abertos, visando sua maior efetividade e utilização. Foram discutidos conceitos como *linked data*, ontologias, Web Semântica e interoperabilidade, apresentando sua relevância para a conexão e a organização dos dados abertos. Foram apresentados casos de uso e exemplos de aplicação de dados abertos conectados na área da segurança pública, evidenciando seu potencial para a prevenção de crimes e o aprimoramento dos serviços à sociedade.

- Seção 5: Dados Abertos Conectados na Segurança Pública

Nesta seção está sendo descrita toda a estrutura de desenvolvimento prático da tese. São descritos os procedimentos utilizados para o desenvolvimento do trabalho, sendo eles: levantamento de fontes de dados relevantes; compreensão e identificação da estrutura dos

dados; padronização dos dados e mapeamento da estrutura semântica dos dados. São abordadas as técnicas, bases de dados utilizadas e a metodologia de enriquecimento semântico dos dados por meio de ontologia. São apresentadas as tecnologias RDF, SPARQL e XML.

- Seção 6: Desenvolvimento da aplicação para a predição de crimes

Esta seção apresenta todas as fases do desenvolvimento da aplicação contendo seus fluxos e processos. Nesta seção é descrita a estrutura do desenvolvimento de uma aplicação dedicada à predição de crimes, detalhando cada componente essencial para a criação e funcionamento da ferramenta bem como o resultado por meio de mapas.

- Seção 7: Conclusões

2 DEMOCRACIA, TRANSPARÊNCIA E TECNOLOGIA

O Sistema Democrático é, atualmente, o sistema político mais aderido em todo o mundo (Huntington, 1993). Esta popularidade, contudo, resulta de uma longa evolução histórica, que se iniciou na Grécia Antiga, com a democracia em sua forma direta. Roma também a praticou durante a Idade Antiga, como resultado das reivindicações dos plebeus para participarem das decisões políticas da sociedade. Durante a Idade Média, porém, a democracia foi pouco praticada, refletindo na sua escassa evolução, principalmente devido à presença de sistemas como o feudalismo, a monarquia e a influência da Igreja Católica. Essa ideia foi posteriormente adotada por outros países europeus ao longo da história, como a Inglaterra, que desenvolveu um sistema parlamentar em que os representantes eleitos governam em nome do povo (Ventura, 2016).

Ao falar de democracia, consideram-se as deliberações universais que beneficiam toda a coletividade sendo tomadas por pessoas eleitas para essa finalidade. Como expressou Norberto Bobbio (2005, p.18),

> [...] essas deliberações não são diretamente tomadas por todos os membros da coletividade. No entanto, esse entendimento é pré-estabelecido com o conceito comum de democracia moderna. Quando se discutem teoremas de democracia direta, é necessária uma explicação mais aprofundada, pois nesse caso se sai do contexto usual [...].

Segundo Lijphart (1997, p. 21), "a democracia tem sido defendida como uma forma de governo que respeita os direitos individuais e promove a liberdade de expressão, a tolerância e a igualdade". Esses valores se tornaram cada vez mais importantes ao longo do tempo, especialmente após as duas guerras mundiais e a ascensão do totalitarismo, em várias partes do mundo.

Outro fator que contribuiu para a disseminação da democracia foi a descolonização, que ocorreu no final do século XX. Os países que conquistaram a independência, após séculos de dominação colonial, frequentemente adotaram a democracia para se libertar da opressão política e social. A Organização das Nações Unidas promove os direitos humanos e a democracia como valores universais (Lijphart, 1997).

É importante destacar que a democracia tem suas limitações e desafios, e que não é um sistema perfeito. No entanto, a sua popularidade em todo o mundo reflete a crença de que é uma forma de governo que, apesar de suas falhas, oferece a melhor oportunidade para a promoção do bem-estar humano e da justiça social.

Nesse sentido, esta tese caracteriza o modelo democrático aberto como um importante instrumento de transformação social, especialmente quando relacionado ao uso de dados abertos para a tomada de decisões coletivas.

2.1 Governo Aberto

O abandono do modelo hierárquico de organização, que se baseia no cumprimento estereotipado de regras, no controle e na unidade de comando, é um tema recorrente na literatura sobre a reforma da Administração Pública, desde meados da década de mil novecentos e setenta (Peters, 1996). Em vez disso, as organizações públicas estão se movendo em direção a um modelo complexo baseado em parcerias e em uma verdadeira rede de governança. Essa transformação de um modelo de "Administração Profissional", próprio do "Estado de Welfare" (Gomes, 2018, p. 3), que se alicerçava no profissionalismo e experiência, para um modelo que depende de colaborações e parcerias tem profundas implicações para a gestão das organizações públicas e, mais especificamente, para os administradores públicos.

A mudança de um modelo de "Administração Profissional" para um modelo baseado em parcerias e em uma verdadeira rede de governança tem profundas implicações para a gestão das organizações públicas, pois abre a possibilidade de aumentar a transparência das informações governamentais e permitir a participação da sociedade civil nas decisões políticas. Esta mudança também exige o desenvolvimento de habilidades de liderança em colaboração, a capacidade de trabalhar em equipes interdisciplinares, a capacidade de gerenciar relações de parceria complexas e o desenvolvimento de uma compreensão aprofundada das complexidades do ambiente em que as organizações públicas operam. Nessa perspectiva, Mello (2009, p. 31) apresenta tendências teóricas da governança eletrônica na gestão pública e, define como meta, minimizar os problemas de departamentos no governo aberto, reduzindo a **assimetria informacional** entre o ator (gestor público) e o cidadão, assim como os problemas relacionados ao comportamento do gestor público, de modo que ele tome decisões que maximizem o bem comum e não seus próprios interesses. O debate sobre a concepção de governo aberto abrange temas essenciais, como a ligação entre a internet e a democracia e a governança da própria internet (Mello, 2009). Estas reflexões visam alcançar formas de aumentar a transparência das informações governamentais, por meio de Dados Governamentais Abertos (DGA), e proporcionar a participação da sociedade civil nas decisões políticas. Por meio de uma análise aprofundada das vantagens e desafios deste novo paradigma, há a possibilidade de

se compreender melhor os reais impactos das transformações digitais na construção de uma sociedade mais igualitária.

A concepção de Dados Governamentais Abertos (DGA) se refere ao princípio de disponibilização de informações governamentais em formatos que possibilitam sua reutilização e interligação com outras fontes, o que leva a um aumento na eficiência, eficácia e transparência da Administração Pública. A disponibilização de DGA é vista como um instrumento para aproximar o Estado de seus cidadãos (Lathrop; Ruma, 2010; Rodrigues; Sant'ana; Ferneda, 2015).

Em 2011, líderes governamentais e defensores da sociedade civil se uniram para criar uma parceria única, que combina essas forças poderosas para promover a governança aberta, participativa, inclusiva e responsável. A Parceria para Governo Aberto (OGP)[1] inclui 76 países, representando mais de dois bilhões de pessoas e milhares de organizações da sociedade civil (Yu; Robinson, 2012; Peixoto, 2013).

De acordo com Gonzalez-Zapata e Heeks (2015, p. 442 *apud* Aleixo 2020, p. 129),

> [...] toma-se como ponto de partida para o aparecimento do conceito de Dados Governamentais Abertos três elementos principais: a abertura do governo para o acesso à informação, o governo como fornecedor de informação e os dados como fonte de conhecimento [...]"

conforme ilustra a Figura 1.

Figura 1 – Formação dos dados governamentais abertos

Fonte: Gonzalez-Zapata; Heeks (2015, p. 442).

A abertura do governo, também conhecida como governo aberto, é uma evolução significativa na administração pública, visto que promove o acesso inclusivo às

[1] Disponível em: https://www.opengovpartnership.org/. Acesso em: 20 fev. 2023.

informações governamentais (Bonsón *et al*., 2012). Com essa abertura, é possível ampliar a participação da sociedade na tomada de decisões governamentais, contribuindo para o aumento da transparência na gestão pública (Martinez-Prieto *et al.,* 2015). Esta nova prática permite que os cidadãos acessem informações essenciais para decisões coletivas, bem como para tomadas de decisões individuais. A abertura dos dados governamentais também fornece ao governo as ferramentas para compreender melhor os problemas enfrentados por seus cidadãos, bem como as tendências de desenvolvimento e mudança na sociedade. A ideia de governo aberto tem raízes antigas nos Estados Unidos, tendo surgido na década de 1950 como uma forma de promover a transparência na administração pública (Santos, 2014).

A noção de governo aberto representou um salto inovador na gestão pública, possibilitando uma maior transparência e acessibilidade. Embora possa parecer um desenvolvimento recente, os fundamentos da teoria de governo aberto possuem raízes historicamente datadas nos Estados Unidos, na década de 1950, como uma forma de responder à ocultação da ação governamental, que se manteve após o término da Segunda Guerra Mundial (Santos, 2014).

Em 1955, o Congresso Americano criou um subcomitê especial para informações governamentais, que serviu de base para a Lei de Liberdade de Informação (*FOIA – Freedom of Information Act*), nos anos 1960[2]. Foi nesse documento do subcomitê que se citou pela primeira vez o termo "governo aberto", conectado com o "direito de saber" ou "direito à informação", promovendo as primeiras discussões relacionadas à transparência, tornando possível contribuir para o aumento da compreensão e conscientização sobre o contexto (Yu; Robinson, 2012; Peled, 2013).

De acordo com Rodrigues (2020), no âmbito de Governo Aberto, a transparência tem sido frequentemente abordada como uma solução para problemas relacionados à ineficiência, corrupção e má gestão, por meio da disponibilização de informações relevantes, confiáveis, oportunas e compreensíveis sobre as ações e decisões políticas.

A ideia de governança transparente remonta à chamada "Batalha das Luzes"[3] contra o absolutismo no século XVIII, quando os principais proponentes Jean-Jacques Rousseau

[2] Disponível em: https://nsarchive2.gwu.edu/nsa/foia/FOIARelease66.pdf. Acesso em: 20 fev. 2023

[3] A "Batalha das Luzes" foi um movimento contra o absolutismo no século XVIII que defendia a transparência e a prestação de contas do governo para a sociedade. Os principais proponentes dessa ideia foram Jean- Jacques Rousseau e Jeremy Bentham, que debateram sobre a importância da governança transparente na época. A ideia era que o governo deveria ser mais responsável e aberto sobre suas ações e decisões, a fim de evitar abusos de

(1712-1778) e Jeremy Bentham (1748-1832) debateram a questão. Bentham foi o maior defensor da transparência, sendo provavelmente o primeiro a usar o termo como base para monitorar o governo. Para o filósofo, a transparência seria uma estratégia para inibir o abuso de poder. De acordo com Rousseau, os agentes públicos deveriam atuar sob vigilância pública, de modo a prevenir intrigas desestabilizadoras (Rodrigues, 2020).

Estas ideias foram contestadas pelos predecessores de Bentham e Rousseau, como Jean Bodin (1530-1596) e Arnold Clapmar (1574-1604), que defendiam o sigilo da política imperial. A discussão sobre transparência foi intensificada nas últimas décadas, sendo importante para a promoção da qualidade dos serviços públicos (Meijer, 2014).

A transparência no governo aberto é um tema de grande relevância atualmente, principalmente quando se trata da prestação de contas por parte dos gestores públicos. Diversas iniciativas têm sido implementadas para fomentar a transparência governamental, como a regulamentação da Lei de Acesso à Informação (LAI), a implementação de sistemas eletrônicos de gestão de documentos, a realização de audiências públicas e a publicação de relatórios de gestão (Silva *et al*., 2019).

A avaliação de desempenho é uma ferramenta importante para promover a transparência no governo aberto. Além de mensurar a eficiência e a eficácia da gestão pública, a avaliação de desempenho pode também identificar desvios ou irregularidades. É crucial que tais avaliações sejam conduzidas de forma regular e transparente, permitindo a participação da sociedade civil e garantindo a publicidade dos resultados. Maia, Correia e Costa (2022) destacam a importância da avaliação de desempenho na promoção da transparência e da responsabilidade no governo aberto.

A responsabilidade, no governo aberto, está relacionada à prestação de contas, ou seja, à obrigação do gestor público de prestar esclarecimentos sobre suas atividades e os recursos públicos sob seu exercício. Nesse sentido, é fundamental que os relatórios de gestão sejam claros, objetivos e acessíveis, permitindo que qualquer cidadão possa compreender as informações contidas neles.

2.2 E-Democracia

Os governos em todo o mundo estão trabalhando para a adoção do modelo democrático e, nesse sentido, um grande desafio é implementar instituições e processos

poder e garantir a proteção dos direitos dos cidadãos. Esse movimento é considerado um marco histórico na luta pela transparência e pelo controle democrático do governo (Fukuyama, 2011).

mais sensíveis às necessidades dos cidadãos comuns. Ronchi; Todaro; Serra, (2023, p. 6) asseveram que a democracia representativa permeia um grande número de Estados e que são estruturados de maneiras diferentes e em diversas camadas de órgãos representativos: municipal, estadual e federal.

> Mais países do que nunca trabalham para construir uma governança democrática. O seu desafio consiste em desenvolver instituições e processos que respondam melhor às necessidades dos cidadãos comuns, incluindo os pobres, e que promovam o desenvolvimento. Atualmente, um grande número de Estados é governado por uma democracia representativa, estruturada de diferentes formas, mas sempre escalados em diferentes níveis de órgãos representativos eleitos direta ou indiretamente pelos cidadãos: governos municipais, estaduais ou federais (Ronchi; Todaro; Serra, 2023, p. 6, tradução nossa).

Uma das tentativas de adotar modelos democráticos e participação popular é a e-democracia. Existe consenso de que a e-democracia se relaciona com o "[...] emprego de tecnologias da informação e comunicação (TIC) [...] como meio para aprimorar a governança" (OECD, 2003, p. 11), a fim de aumentar a eficiência dos processos administrativos, aprimorar os serviços públicos destinados aos cidadãos/consumidores, contribuir para a consecução de resultados específicos em políticas públicas como, na área de saúde, meio ambiente e educação, ou resultados econômicos, reduzindo a corrupção e promovendo a modernização da gestão pública (Kneuer, 2016).

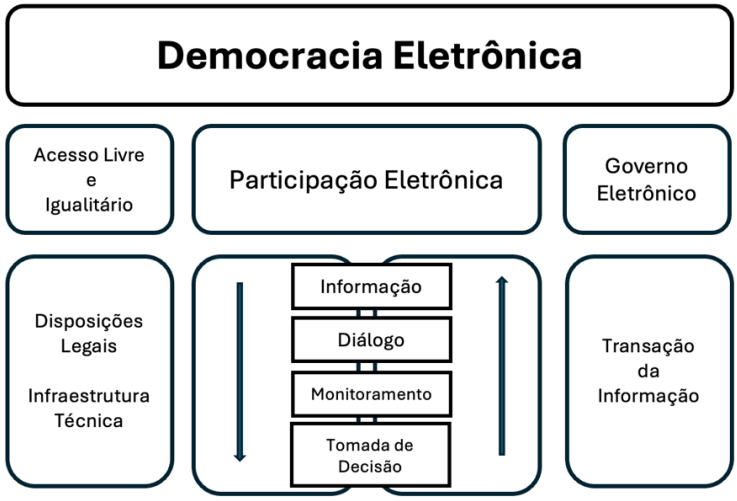
Kneuer (2016, p. 671) sugere três dimensões, ou conceito tridimensional, para caracterizar os processos de "e-democracia". A **primeira dimensão** refere-se às condições básicas para o uso de TIC em democracias e baseia-se em dois pré-requisitos centrais para a existência e durabilidade da e-democracia: a infraestrutura técnica (acesso a mídias digitais); e a liberdade da internet. Estas últimas são disposições legais relativas ao uso livre da internet ou regulação de conteúdo. Tão importante quanto a liberdade da internet legalmente garantida é o acesso livre à internet com base na infraestrutura técnica e a questão de se há ou não filtragem, ou bloqueio por órgãos reguladores. Criar uma dimensão específica para a liberdade da internet e o acesso igualitário fornece uma imagem mais clara da situação legal e técnica nos países do que incluir esses fatores em escores agregados.

A **segunda dimensão** é a "e-participação", que, nesse contexto, três elementos são identificados. Como enfatizado por vários estudiosos, há uma distinção básica entre duas dimensões na "e-participação": (1) uma direção no sentido "*top-down*" (Coleman; Blumler, 2009, p. 90) e (2) uma direção *"botton-up"* (OECD, 2003, p. 30). Em terceiro lugar, a

votação eletrônica nos processos de tomada de decisão, já que - assim como no mundo offline - estas são ações funcionalmente diferentes (OECD, 2003, p. 32).

A **terceira dimensão** relaciona-se ao "e-monitoramento", com base na suposição de que por meio de mídias digitais os cidadãos não só podem obter um melhor conhecimento de problemas ou desenvolvimentos sociais, mas também possuir direitos para expressar suas opiniões, caracterizado pelo autor como "alarmes" (Hindman, 2009, p. 136-138). Já existem inúmeras organizações de monitoramento parlamentar *online*, como o modelo *Parliament Watch* na Alemanha, que desempenham uma função de monitoramento, responsabilizando os membros do parlamento[4]. A Figura 2 retrata um modelo de democracia eletrônica.

Figura 2 – Modelo de Democracia eletrônica



Fonte: Modelo de democracia eletrônica sugerido por Kneuer (2016, p. 672, tradução nossa).

Conforme a Figura 2, são distinguidas duas direções diferentes de e-participação: "*top-down"* e *"botton-up"*, e quatro níveis: e-informação (informação eletrônica); e-consulta; e-monitoramento (monitoramento eletrônico); e, e-tomada de decisão (tomada de decisão eletrônica). Estes níveis refletem diferentes graus de engajamento em termos de tempo, entrada e compromisso. Os cidadãos escolhem se usam os meios digitais apenas para informação, se querem contactar políticos (através de e-mail ou redes sociais como *Twitter* ou *Facebook*) e, assim, entrar em diálogo com eles, se escolhem se envolver no monitoramento de políticos em plataformas específicas ou se vão mais longe e tomam um passo mais exigente de se envolver na tomada de decisão (por exemplo, por "e-petições", assinando campanhas e governança colaborativa). Ele se aplica às ferramentas "*top-*

[4] Essa organização tem a função de responsabilizar os membros do parlamento por meio de monitoramento. Disponível em: https://www.abgeordnetenwatch.de/ueber-uns/mehr/international. Acesso em: 23 fev. 2023.

down", bem como para os canais *online* específicos oferecidos pelos governos. Segundo o modelo utilizado pelas Nações Unidas (2014, p. 195), os serviços *online* do governo são classificados em quatro etapas cada vez mais exigentes: emergentes; melhorados; transacionais; e conectados.

A terceira dimensão corresponde ao "e-governo", que, ao contrário da "e-participação", se limita a um mecanismo "top-down", que oferece ferramentas *online* aos cidadãos como serviço do governo com foco na prestação de serviços públicos, eficiência e resultados da política. Nesta dimensão, o papel do cidadão tende a ser concebido como aquele de um consumidor ou cliente. Os objetivos das ferramentas do e-governo focam na redução de custos e eficiência nas transações administrativas. Ao mesmo tempo, a transparência aumentada pode levar à redução da corrupção. Além disso, a transparência nos procedimentos de formulação de políticas e a satisfação aumentada dos cidadãos/consumidores podem aumentar a confiança para com o governo, o que também deve ter um impacto na legitimidade (OECD, 2003, p. 45; Clift, 2004, p. 8-14).

Embora o e-governo pareça ser puramente uma questão de serviço ou eficiência, tem implicações para um aumento possível na qualidade democrática, resultando no aumento da dimensão de saída, bem como apoiar uma atitude positiva em relação ao governo, ao regime e à democracia em geral (Kneuer, 2016).

Essas abordagens permitem que os cidadãos se envolvam nos processos de tomada de decisão em níveis mais profundos, fazendo sugestões, participando de discussões e dando sua opinião sobre os assuntos mais importantes. Esse processo resulta em decisões que refletem as opiniões da população. Além disso, o Governo Aberto e a e-democracia podem ajudar a melhorar a eficiência do governo, gerenciando os recursos públicos de maneira mais eficaz e economizando tempo e dinheiro (Kneuer, 2016).

No entanto, existem desafios associados à implementação do Governo Aberto e da e-democracia. Os governos precisam garantir que a participação cidadã seja representativa da população na totalidade e que as informações disponibilizadas aos cidadãos sejam precisas e verdadeiras. Se esses desafios forem superados, o Governo Aberto e a e-democracia podem trazer benefícios significativos para governos, cidadãos e organizações, como aumento da transparência e da eficiência, redução do desperdício e da corrupção, melhora da participação cidadã e fortalecimento da confiança dos cidadãos no governo e nas instituições públicas (Hindman, 2009).

O Governo Aberto e a e-democracia são abordagens inovadoras para a gestão dos recursos públicos que visam promover a participação cidadã e a transparência na tomada

de decisão. Essas abordagens estão se tornando cada vez mais comuns em todo o mundo com o potencial de transformar a maneira como as instituições públicas operam. Nessa perspectiva, umas das organizações públicas que pode se beneficiar significativamente na implementação dessas abordagens é a Polícia Militar.

A democratização das instituições públicas, incluindo a Polícia Militar, é um processo essencial para a promoção de uma cultura de transparência e participação cidadã na tomada de decisões governamentais. A implementação do Governo Aberto e da e-democracia pode ajudar a Polícia Militar a se tornar mais eficiente, transparente e responsável, o que é fundamental para fortalecer a confiança e a legitimidade das instituições públicas. É importante destacar que a Polícia Militar também enfrenta desafios e limitações na utilização de dados abertos, mas é necessário que a instituição esteja ciente dessas limitações e busque formas de minimizar seus efeitos negativos. A democratização das instituições públicas é um processo contínuo e complexo que requer um compromisso constante com a transparência, a responsabilidade e a participação cidadã (Lima, 2017).

2.3 A democratização das instituições públicas: Polícia Militar

A democratização das instituições públicas é um tema central para o fortalecimento da democracia e para a garantia dos direitos dos cidadãos. Deve ser possível a participação mais ativa da sociedade civil nas políticas públicas e na tomada de decisões governamentais (Mendes, 2010). Para isso, é importante que haja canais de diálogo e de participação da sociedade, como audiências públicas, conselhos regionais e locais de segurança e outras formas de consulta popular (Lima, 2017).

A necessidade de mudanças culturais e de práticas tradicionais de políticas pode ser um grande desafio, visto que, é preciso superar as barreiras da cultura burocrática e da falta de transparência, que muitas vezes impedem a participação da sociedade e a efetividade das políticas públicas (Brasil, 2009). Nesse sentido, é fundamental investir em educação e em formação política, para preparar a sociedade para o exercício de sua cidadania (Mendes, 2010).

Outro aspecto importante é a garantia da diversidade e da representatividade na tomada de decisões. Isso implica na inclusão de grupos historicamente marginalizados e na valorização da diversidade cultural e social (Lima, 2017). A participação de mulheres, negros, indígenas e outros grupos que são deixados à margem na gestão pública é fundamental para a promoção da igualdade e para o fortalecimento da democracia (Jacobi, 2006).

Além disso, a democratização das instituições públicas passa pela descentralização do poder e pela valorização dos espaços locais de participação. A gestão democrática deve ser fortalecida nos níveis municipais e regionais, para garantir a participação da população na gestão pública e na tomada de decisões que afetam diretamente suas vidas (Mendes, 2010).

É importante destacar que não se trata apenas de um processo de abertura e de participação, mas também de um processo de responsabilidade e de controle social. É preciso garantir que as instituições públicas estejam sujeitas ao escrutínio da sociedade, para prevenir a corrupção e garantir a efetividade das políticas públicas (Lima, 2017).

É preciso reconhecer ser um processo contínuo e que exige o engajamento constante da sociedade e dos gestores públicos. A busca pela promoção da igualdade, da transparência e da participação deve ser uma constante na agenda pública, para garantir uma gestão mais eficiente e democrática (Brasil, 2009).

Uma das mais importantes instituições públicas em nossa sociedade é a Polícia Militar. Como uma instituição estatal responsável pela segurança pública, deve estar incluída no processo de democratização, envolvendo a criação de mecanismos de controle e participação social, de forma que a instituição possa estar mais próxima da sociedade e ser mais responsiva às demandas da população (Brasil, 2009).

Além disso, a democratização da Polícia Militar consiste igualmente na promoção de uma cultura que priorize os direitos humanos e o respeito aos direitos civis, implicando na capacitação e formação de seus policiais, bem como na criação de dispositivos transparentes que garantam o acesso responsável às informações de segurança pública (Silva, 2020). Essa evolução envolve a garantia da diversidade e pluralidade no seu corpo funcional, significando a promoção da igualdade de oportunidades de acesso e ascensão profissional, bem como o respeito à diversidade étnica, de gênero e orientação sexual (Moreira, 2019).

Por fim, a democratização da polícia militar é fundamental para o fortalecimento da democracia e para a garantia dos direitos dos cidadãos. É preciso que a instituição esteja aberta ao diálogo e à participação da sociedade, de forma que possa ser mais eficiente e eficaz na promoção da segurança pública e no respeito aos direitos humanos (Fernandes, 2018).

2.3.1 História da Polícia Militar do Estado de São Paulo

A Polícia Militar do Estado de São Paulo (PMESP) é uma das instituições mais antigas do Brasil, tendo sido criada em 1831, com o nome de "Força Pública". Desde então, a instituição passou por diversas mudanças em sua estrutura e funções, acompanhando as transformações da sociedade brasileira ao longo dos anos (Pimentel, 2006).

No início, a Força Pública era responsável por manter a ordem pública e combater as rebeliões que surgiam no interior do estado. Durante a Revolução Constitucionalista de 1932, a instituição teve um papel de destaque, lutando ao lado dos paulistas contra o governo federal. A partir da década de 1960, a Polícia Militar começou a atuar também no combate ao crime comum, além das funções tradicionais de policiamento ostensivo e preservação da ordem pública (Barros, 2011).

A história da Polícia Militar de São Paulo reflete as transformações sociais e políticas do Brasil ao longo dos anos, com mudanças significativas em sua estrutura e funções. Apesar dos desafios, a instituição desempenha um papel importante na proteção da sociedade e no combate ao crime no estado. A trajetória da Polícia Militar de São Paulo evidencia adaptações relevantes em sua organização e atribuições ao longo do tempo. Nesse contexto, a atuação das Ouvidorias de Polícia emerge como um mecanismo essencial para monitorar e corrigir desvios de conduta dentro da corporação, contribuindo para o aprimoramento da sua função de salvaguarda da ordem pública e enfrentamento à criminalidade no estado (Ataíde *et al*., 2020).

A publicação dos relatórios pelas Ouvidorias de Polícia representa uma oportunidade significativa para apontar irregularidades nas condutas policiais experimentadas pela população em diversos estados. No entanto, para que a análise da efetividade desses documentos emitidos tanto pelas Ouvidorias quanto pelas próprias instituições policiais seja precisa, é imperativo superar obstáculos que comprometem tal avaliação. Entre esses desafios, destacam-se a insuficiência na qualidade das informações registradas, a inconsistência na frequência de divulgação e a dificuldade em estabelecer comparações padronizadas entre os relatórios, todos fatores que dificultam uma análise aprofundada e consequente melhoria nas práticas policiais (Ataíde *et al.,* 2020).

Além disso, a corrupção também é um problema recorrente na Polícia Militar de São Paulo. Em 2019, a Corregedoria da instituição instaurou 1.647 procedimentos para investigar denúncias de corrupção, sendo que 91 policiais militares foram presos em flagrante por envolvimento em esquemas de corrupção (PMESP, 2020). A corrupção na

Polícia Militar não só prejudica a imagem da instituição, como também compromete a segurança da população e a efetividade do combate ao crime.

A contínua desvalorização dos policiais militares, conforme destacado por Alves (2018), não só mina o moral da força como também compromete a implementação de métodos inovadores na segurança pública. Nesse contexto, surge o policiamento preditivo como uma estratégia emergente, visando a eficiência e a proatividade no combate ao crime. É importante abordar as questões de reconhecimento e valorização dos policiais para que se possa aproveitar plenamente o potencial das novas abordagens tecnológicas em segurança pública.

2.4 Policiamento Preditivo

O policiamento preditivo é caracterizado como o mais recente patamar da escala evolutiva das forças policiais, com raízes que podem ser rastreadas desde a década de 1950. Nesse sentido, nos Estados Unidos, foi criado o "*National Institute of Justice"* (NIJ)[5], que é uma agência de pesquisa, desenvolvimento e avaliação do Departamento de Justiça. Eles se dedicam a aprimorar o conhecimento e a compreensão das questões do crime e da justiça por meio da ciência, fornecendo conhecimento e ferramentas objetivas e independentes para aprimorar a tomada de decisões dos órgãos de justiça criminal, com o objetivo de reduzir o crime e promover a justiça, principalmente nos níveis estadual e local (National Institute of Justice, 2009a).

O *Predictive Policing* tem sido amplamente explorado para melhorar a eficácia das operações policiais. A partir da premissa de que modelos preditivos são capazes de transformar o universo empresarial, também pode-se explorar o potencial desta técnica para a área policial. Esta abordagem pressupõe que os elementos estatísticos presentes em dados históricos de criminalidade possam ser usados para prever o futuro comportamento criminoso. Assim, ao usar técnicas de inteligência artificial, aplicando modelos preditivos para predizer e prevenir a criminalidade, as unidades policiais podem obter resultados mais eficazes (Pearsall, 2010).

Um dos principais benefícios desta técnica é a capacidade de prever onde, quando e como a criminalidade pode ocorrer. A análise de dados históricos de criminalidade permite que sejam desenvolvidas estratégias de prevenção mais eficazes para a ocorrência de crimes. É possível identificar padrões de criminalidade, como áreas problemáticas e

[5] Disponível em: https://nij.ojp.gov/. Acesso em: 23 fev. 2023.

horários mais propensos à ocorrência de delitos, permitindo que as unidades policiais estejam mais preparadas para lidar com essas situações (National Institute of Justice, 2009a).

Clemente (2013, p. 152) destacou a importância da estratégia de prevenção da incivilidade, com um foco especial na videovigilância das áreas de maior incidência criminal ou de percepção de insegurança. Expandindo essa ideia, pode-se considerar que as técnicas de policiamento preditivo, embora não mencionadas explicitamente por Clemente naquela época, se alinham perfeitamente com a filosofia subjacente à sua abordagem. O uso de dados históricos criminais e a análise preditiva emergem como complementos naturais à videovigilância, permitindo não apenas a monitoração, mas também a previsão e prevenção mais eficaz de atividades criminosas. Esta abordagem integrada realça a visão de Clemente, sugerindo que as autoridades policiais podem aumentar a precisão de suas investigações e intervenções ao identificar tendências e padrões não comuns de crimes, otimizando assim os esforços para manter a segurança pública.

Em 1998, o NIJ abriu uma chamada pública[6] para que fossem desenvolvidos modelos preditivos, permitindo que as forças de segurança pudessem adotar uma abordagem proativa para a prevenção de incidentes criminais. No ano de 2002 foi publicado o resultado do trabalho, onde os autores concluíram que o desenvolvimento de modelos preditivos para a aplicação da lei tem um valor muito relevante para a sociedade.

> [...] o desenvolvimento de modelos de previsão da atividade criminosa é de enorme valor para a aplicação da lei. A utilização destes modelos para apoiar a tomada de decisões táticas no âmbito da aplicação da lei é óbvia: quanto melhor for a previsão da atividade criminosa, melhor será a alocação dos recursos de aplicação da lei para a combater [...] (Brown, 2002, tradução nossa).

Em contrapartida, o *Geospatial Technology Working Group (TWG)*, sendo um grupo de especialistas do NIJ, que desenvolve o programa *Mapping and Analysis for Public Safety* (*MAPS*), programa este que avalia as necessidades específicas em tecnologias geoespaciais e determina a prioridade na atribuição de recursos, indica que o policiamento preditivo não se limita a identificar incidentes criminais, mas também a identificar os fatores subjacentes à ocorrência de tais incidentes, bem como a identificar tendências iniciais. O grupo sugere que a previsão de crimes possui dois componentes. O primeiro componente diz respeito à predição de esforços que visam avaliar riscos mais amplos e

[6] NIJ Grant 984J-CX-KO10 (National Institute of Justice, 2009b).

tendências de longo prazo. O segundo componente envolve a previsão dos fatores de curto prazo que podem levar à ocorrência de crimes (Wilson *et al.*, 2009).

De acordo com Uchida (2009, p. 2), existem cinco elementos fundamentais no policiamento preditivo: (1) integração de informações e operações, que sugere que os trabalhos dos analistas devem corresponder às necessidades dos policiais e investigadores; (2) observação abrangente – a prevenção é tão importante como a resposta, tornando-se qualquer incidente numa importante fonte de informação; (3) análise e tecnologia de ponta, que requer que as unidades policiais aprendam a fazer melhor uso das ferramentas e tecnologia disponíveis; (4) ligação com o desempenho; e (5) adaptabilidade às condições de mudança, que remete à necessidade de organizações descentralizadas, formação sobre como adaptar estratégias baseadas em informações e elevados padrões profissionais.

O estabelecimento de uma conexão entre teoria e prática é fundamental para garantir o sucesso do policiamento preditivo. O domínio das teorias da criminologia e da prevenção criminal, bem como a análise de informações, são ferramentas essenciais para o direcionamento de estratégias e táticas policiais (Uchida, 2009).

Existem algumas características que foram definidas pelo NIJ (National Institute of Justice, 2009a), sendo elas: policiamento proativo, integração de paradigmas, tecnologias, participação da comunidade e seus direitos fundamentais, a qualidade e origem dos dados e rentabilização de recursos.

2.4.1 Policiamento proativo

A Análise Preditiva permite que as forças de segurança alterem a sua abordagem para além do que já aconteceu, para se focarem no que está para acontecer e como reagir antecipadamente. O maior benefício desta abordagem proativa é a descoberta de novos padrões de tendências que são desconhecidos. (Beck, 2009 *apud* Pearsall, 2010).

Clemente (2021, p.110) apresenta o conceito de "policiamento guiado pelas informações" como a essência de uma polícia inteligente que cumpre bem sua missão, enfatizando a máxima de conhecer para agir eficientemente.

Como forma de integração de paradigmas, o modelo "*Predictive Policing"* é uma fusão das melhores práticas policiais existentes, sem a intenção de substituir nenhum dos outros modelos de policiamento. Na verdade, essa abordagem se baseia nos princípios do *Problem-Oriented Policing (POP)*, do *Community Policing (CP)* e de outros modelos comprovados. Esta evolução foi possível devido à capacidade de coleta e análise de

informação, permitindo a previsão de ações (Bratton, 2009 *apud* Bureau of Justice Assistance, 2009; Beck; Bratton, 2009 *apud* Pearsall, 2010).

A relação entre teoria e prática é fundamental para o sucesso do policiamento preditivo. É necessário um conhecimento profundo das teorias da criminologia e da prevenção criminal, para permitir uma análise de informação mais precisa, objetivando a orientação de estratégias e táticas de policiamento (Uchida, 2009).

2.4.2 Tecnologias

O policiamento preditivo baseia-se na aplicação de novas tecnologias, processos e algoritmos para prevenir o crime e agir antes que ele ocorra. As bases tecnológicas derivam do mundo empresarial, particularmente do *business intelligence* e do *business analytics* e das suas técnicas analíticas, como *Hot Spots*, Mineração de Dados, previsão geoespacial, análise de redes sociais e probabilidades estatísticas (Uchida, 2009, p. 6). Nesse sentido, há que considerar a relevância dos dados.

Os dados coletados devem ser providos de confiança para o êxito da produção de informações. Uma integração eficaz de informações de diferentes fontes é essencial para assegurar que os resultados sejam de qualidade. Por meio de sistemas tecnológicos, as polícias têm acesso a um grande volume de dados, exigindo que exista uma supervisão para garantir a qualidade destes como sendo limpos e confiáveis (Uchida, 2009).

A utilização de meios tecnológicos permite um acesso a fontes de dados não tradicionais, tais como dados médicos, licenciamentos, dados escolares, dados de ocupação habitacional e dados censitários, os quais estão disponíveis localmente por meio de sistemas interoperáveis. A abordagem integradora adotada por essa prática, que busca transcender as paredes do departamento policial, é holística, e parte do desafio consiste em fundir esses dados e utilizá-los nas tomadas de decisões (Uchida, 2009).

2.4.3 Rentabilização de recursos

O Policiamento Preditivo tem sido destacado como uma valiosa ferramenta para apoiar departamentos policiais a tornarem-se mais eficazes em tempos de dificuldades econômicas, nas quais os orçamentos estão continuamente decrescendo. Além de ajudar a gerir orçamentos policiais, as predições permitem uma distribuição mais eficaz dos recursos, garantindo que os recursos certos estejam no local certo, no momento certo. No entanto, o Policiamento Preditivo vai além da abordagem *"cops-on-dot"* (pontos de estacionamentos de viaturas), ao prever as áreas onde crimes podem ocorrer, permitindo

aos decisores policiais implementar táticas para mitigar riscos. Embora não pretendendo substituir os analistas, o Policiamento Preditivo consegue analisar muitos dados e identificar relações que não são facilmente identificáveis. Desta forma, permite fazer mais com menos, ao permitir uma melhor distribuição de recursos policiais (Uchida, 2009).

2.5 Inteligência Policial Baseada em Dados

De acordo com Cukier e Mayer-Schönberger (2013), a análise de dados abertos pode ser usada para prevenir crimes por meio da identificação de padrões e tendências que possam indicar a ocorrência de crimes em determinadas áreas ou momentos. Os autores afirmam que a análise de dados abertos pode permitir que as autoridades policiais identifiquem padrões de comportamento criminoso, como horários, locais e *modus operandi*, que podem indicar a presença de criminosos em determinadas áreas. Essas informações podem ser utilizadas para planejar intervenções preventivas, como o aumento do policiamento ou a instalação de câmeras de vigilância em áreas de risco.

De acordo com Ratcliffe (2015), ao nível estratégico, a análise de dados pode ajudar a identificar os locais mais críticos de uma cidade, bem como as atividades criminosas que ocorrem nessas áreas, permitindo que as autoridades policiais adotem estratégias de policiamento mais eficazes, concentrando recursos onde são mais necessários e usando táticas mais adequadas para cada tipo de delito.

A coleta e análise de dados podem ser utilizadas pela Polícia Militar para prevenir crimes e otimizar os trabalhos de segurança pública. Isso ocorre porque a análise de dados pode fornecer informações valiosas sobre padrões e tendências de crimes, o que pode contribuir nas ações policiais preventivas e estratégias de policiamento inteligente (Vasconcellos, 2016).

Para coletar e analisar esses dados, são necessárias tecnologias e metodologias de análise de dados. Existem diversas ferramentas de análise de dados disponíveis, como algoritmos de *machine learning* (ML) e técnicas de mineração de dados (MD), que podem ser utilizadas para identificar padrões e tendências em grandes conjuntos de dados (Ratcliffe, 2015).

No entanto, existem desafios e limitações associados a essa abordagem. Um dos principais desafios é garantir a qualidade dos dados coletados e analisados, uma vez que pode haver erros ou imprecisões nos dados, que podem levar a análises incorretas ou enganosas (Vasconcellos, 2016). Alguns modelos podem ser citados como casos de sucesso, dentre eles, o estudo de Bruederle, Peters e Roberts (2017).

Além disso, a privacidade e a segurança dos dados também são preocupações importantes. A coleta e análise de dados abertos podem envolver informações pessoais e sensíveis, e é importante garantir que essas informações sejam protegidas adequadamente para evitar o uso indevido ou abusivo (Ratcliffe, 2015).

A Polícia Militar deve promover a transparência por meio da disponibilização de dados abertos, em formatos acessíveis e compreensíveis pela população, estabelecendo canais de diálogo com a sociedade para ouvir suas demandas e que sejam adotadas medidas de segurança para proteger os dados e evitar possíveis danos à segurança pública. Deve-se considerar que a disponibilização de dados abertos pela Polícia Militar pode levantar questões relacionadas à privacidade e à segurança das informações. É necessário, portanto, que sejam adotadas medidas de segurança para garantir a proteção dos dados, bem como que seja estabelecido um protocolo de compartilhamento de informações (Ratcliffe, 2015).

A disponibilização de dados abertos deve ser acompanhada de políticas de comunicação efetivas, a fim de que a sociedade tome conhecimento das informações disponibilizadas. É necessário, portanto, que sejam adotadas estratégias de comunicação que possibilitem a disseminação dos dados de forma ampla e que estimulem a participação da população na construção de políticas públicas, na área de segurança pública (Vasconcellos, 2016).

Outro aspecto relevante a ser considerado é a necessidade de capacitação dos profissionais da Polícia Militar, para a utilização de tecnologias de coleta e análise de dados. É importante que a instituição invista em treinamentos e capacitações, a fim de que os profissionais da área de segurança pública possam utilizar essas ferramentas de forma eficiente e eficaz, evitando conclusões equivocadas e ações ineficazes (Vasconcellos, 2016).

Além disso, é importante que a Polícia Militar trabalhe em conjunto com outros órgãos da administração pública, como prefeituras, institutos de pesquisa e universidades, para desenvolver e aprimorar as tecnologias e metodologias de análise de dados. Essa colaboração pode ser benéfica tanto para a instituição policial como para a comunidade em geral (Ratcliffe, 2015).

A transparência na forma como os dados são coletados, armazenados e analisados pela Polícia Militar é um fator importante a ser destacado. A instituição deve ser clara quanto aos procedimentos adotados, as fontes de dados utilizadas e as metodologias aplicadas. Isso ajuda a construir a confiança da sociedade nas atividades policiais e contribui para uma maior aproximação entre a polícia e a comunidade (Correia, 2021).

Outro desafio é a falta de padronização e de normas para a coleta e disponibilização de dados abertos. Isso pode dificultar a análise e a interpretação dos dados por parte da Polícia Militar, uma vez que diferentes fontes podem fornecer informações de maneira heterogênea. Portanto, é necessário que haja uma padronização na coleta e na disponibilização desses dados para poderem ser utilizados de forma mais eficiente (Correia, 2021).

É importante ressaltar também que a utilização de dados abertos não deve ser vista como uma solução única para o combate à criminalidade. A análise de dados deve ser complementada por outras estratégias de segurança pública, como o policiamento comunitário, a prevenção situacional, o policiamento preditivo, dentre outras. A Polícia Militar deve buscar uma abordagem integrada e multifacetada para a segurança pública (Vasconcellos, 2016). Por fim, é importante destacar que a disponibilização de dados abertos pela Polícia Militar deve ser acompanhada de uma política de comunicação clara e efetiva com a sociedade. A instituição precisa explicar como os dados são coletados e utilizados, a fim de garantir a transparência e a confiança da sociedade no trabalho policial. A Polícia Militar deve buscar uma relação mais próxima e colaborativa com a comunidade, envolvendo-a no processo de coleta e análise de dados abertos (Zuiderwijk; Janssen, 2014).

De modo geral, a análise de dados abertos pode contribuir significativamente para a prevenção e o combate à criminalidade, bem como para a aproximação e a confiança entre a Polícia Militar e a sociedade. No entanto, é importante considerar os desafios e limitações associados a essa abordagem, buscando formas de minimizá-los e de complementar a análise de dados com outras estratégias de segurança pública. A capacitação dos profissionais da Polícia Militar, a padronização na coleta e na disponibilização de dados, a proteção da privacidade e segurança dos dados, a interpretação adequada dos dados, uma abordagem integrada e multifacetada para a segurança pública e uma comunicação clara e efetiva com a sociedade, são elementos essenciais para uma análise de dados abertos bem-sucedida na área de segurança pública.

A partir da necessidade de uma análise de dados abertos bem fundamentada e eficaz na segurança pública, conforme discutido anteriormente, emerge a relevância do próximo capítulo, 'Dados Abertos'. Este capítulo aprofundará a discussão sobre como a disponibilização e a utilização estratégica de dados abertos podem servir como um pilar fundamental para o aprimoramento das políticas de segurança.

3 DADOS ABERTOS

Nesta seção, aborda-se o contexto sobre dados abertos, destacando sua importância, disponibilidade e acesso, potenciais benefícios e desafios, bem como as principais questões de governança, leis relacionadas ao seu uso, dentre outros temas de extrema relevância nesse contexto. O objetivo desta seção é fornecer uma visão geral das questões relacionadas aos dados abertos e seus impactos.

Os dados são usados para representar fatos, conceitos ou instruções que podem ser processadas por computador, usadas para tomar decisões baseadas em informações confiáveis, como estatísticas ou pesquisa de mercado, e podem ser organizados em tabelas, gráficos e diagramas. A grande disponibilidade de dados induz ao desenvolvimento de ferramentas avançadas, incluindo softwares e algoritmos de inteligência artificial, projetados para gerenciar e operar esses volumes de informações, visando solucionar problemas complexos. Conforme o *National Institute of Standards and Technology (NIST)*[7], os dados são informações qualificadas, como valores, conceitos ou símbolos, usadas para representar fatos, conceitos ou instruções que podem ser processadas por computador.

De acordo com Kitchin (2014a), dados são a matéria-prima a partir da qual a informação e o conhecimento são produzidos.

> Os dados são geralmente entendidos como a matéria-prima produzida pela abstração do mundo em categorias, medidas e outras formas de representação - números, caracteres, símbolos, imagens, sons, ondas electromagnéticas, bits - que constituem os blocos de construção a partir dos quais a informação e o conhecimento são criados (Kitchin, 2014a, p. 28, tradução nossa)

O autor afirma, ainda, que os dados apresentam natureza representativa, implícita ou derivada:

> Os dados são geralmente de natureza representativa (por exemplo, medições de um fenômeno, como a idade, altura, peso, cor, tensão arterial, opinião, hábitos, localização de uma pessoa), mas também podem ser implícitos (por exemplo, através de uma ausência em vez de uma presença) ou derivados (dados produzidos a partir de outros dados, como a variação percentual ao longo do tempo calculada através da comparação de dados de dois períodos de tempo), e podem ser registados e armazenados em formato analógico ou codificados em formato digital, sob a forma de bits (Kitchin, 2014a, p. 28, tradução nossa).

[7] Disponível em: https://www.nist.gov/. Acesso em: 20 fev. 2023.

O aumento significativo de dados produzidos à medida que mais dispositivos e sistemas se conectam à internet provoca uma profunda reflexão sobre o papel que a tecnologia desempenha na nossa sociedade. Com o aumento do volume de dados produzidos, os dados passam a ocupar um papel cada vez mais importante na economia em geral, na tomada de decisões e na forma como as pessoas se relacionam entre si. Isso leva a considerar a relevância dos dados não apenas como informação, mas como um recurso inerente ao nosso processo de tomada de decisão (Cukier; Mayer-Schonberger, 2013).

De acordo com Reis e Sá (2020), cerca de 2,5 exabytes de dados são produzidos diariamente em todo o mundo. Isso inclui dados gerados por redes sociais, sites de comércio eletrônico, dispositivos móveis, sistemas de monitoramento de câmeras, geolocalização, dentre outros.

Entretanto, a maioria desses dados não é disponibilizada para a sociedade, tampouco são organizados e estruturados para que os cidadãos os compreendam de modo a tomarem decisões, sejam elas corporativas ou cotidianas. Dessa maneira, inúmeras empresas, governos e instituições de pesquisa têm trabalhado para produzir tecnologias que possibilitem a produção e o consumo de dados visando acelerar a descoberta de novos conhecimentos e dar valor a qualquer informação disponível gratuitamente *online* (Isotani; Bittencourt, 2015).

De acordo com Buckland (2012), os dados estão relacionados à Ciência da Informação, uma vez que esta ciência se ocupa do estudo e da gestão dos dados, incluindo como eles são coletados, armazenados, organizados, analisados e disseminados. Os dados são conjuntos de informações que podem ser usadas para descrever ou avaliar um determinado fenômeno. A Ciência da Informação lida também com questões éticas e legais relacionadas à coleta, uso e divulgação de dados (Buckland, 2012).

Enquanto disciplina, a Ciência da Informação tem por objetivo auxiliar as pessoas a lidar com a quantidade crescente de informações de que dispõem. Como tal, ela oferece uma combinação única de habilidades técnicas e humanas, permitindo ao profissional da área desenvolver soluções inovadoras para o uso eficaz dos dados. Assim, a Ciência da Informação ajuda as organizações a identificar, selecionar, organizar, analisar e disseminar informações relevantes para tomadas de decisões eficazes (Buckland, 2012).

A Ciência da Informação oferece métodos para a criação, armazenamento, recuperação, análise e disseminação de dados, incluindo a criação de bancos de dados, a organização de informações em formatos acessíveis, a análise de dados e a criação de mecanismos de busca para facilitar o acesso a informações relevantes. Além disso, a

Ciência da Informação também está envolvida na criação e disseminação de dados abertos governamentais, que auxiliam na transparência, no acesso à informação por parte dos cidadãos e, por fim, desempenha um papel fundamental na criação, organização, análise e disseminação de dados (Buckland, 2012), permitindo que os dados sejam usados de forma eficaz para tomadas de decisão.

3.1 O que são Dados Abertos

O conceito de dados abertos surgiu em 2008, com o lançamento do *Open Access Data Policy,* pelo governo dos Estados Unidos (USA.Gov, 2018). Desde então, o conceito tem sido adotado por governos, empresas e organizações em todo o mundo, tornando-se cada vez mais importante para a promoção de inovação e transparência. Segundo West (2018), "governos a todos os níveis e em todas as partes do mundo poderiam beneficiar enormemente de uma utilização mais estratégica de dados abertos. E é do interesse do movimento de dados abertos ajudá-los".

Dados abertos são disponibilizados ao público de forma gratuita e podem ser usados, compartilhados e modificados pelos usuários. Eles podem incluir dados de governos, empresas e organizações sem fins lucrativos. A disponibilização de dados abertos é geralmente feita por meio de plataformas *online*, onde os usuários podem acessar, baixar e trabalhar com os dados (James, 2013).

Nos últimos anos, tem havido esforços para desenvolver métodos eficazes de gestão e análise de dados, dado o seu volume e a diversidade de fontes e formatos presentes na internet. O uso dos dados assume um papel vital na interação entre as pessoas e as empresas, na web (Kumar; Jain; Sharma, 2020).

Eles permitem que as pessoas tenham acesso a informações valiosas que podem ser usadas para tomar decisões. Por exemplo, os dados abertos podem ser usados para identificar padrões e tendências, criar mapas e gráficos, e até mesmo para desenvolver novas tecnologias e produtos. Além disso, os dados abertos promovem a transparência e a responsabilidade, permitindo que os cidadãos monitorarem o trabalho dos governos e das empresas (Kuhn, 2019).

Dados abertos têm sido utilizados de maneira muito útil em nossa sociedade. Sobre o transporte público, foram usados para criar aplicativos que auxiliam as pessoas a planejar suas viagens. Os dados abertos sobre a qualidade do ar foram usados para criar mapas em tempo real que mostram a qualidade do ar em diferentes regiões. E os dados abertos sobre

a saúde pública foram usados para identificar padrões e tendências em doenças, ajudando a orientar as políticas de saúde pública (Kuhn, 2019).

Embora os dados abertos sejam disponibilizados publicamente, isso não significa que eles estejam isentos de restrições. Alguns conjuntos de dados podem ter restrições de uso ou requisitos de atribuição, então é sempre importante ler as informações de licenciamento e termos de uso antes de trabalhar com os dados.

Neste sentido, existem algumas premissas relevantes sob a ótica de dados abertos. A primeira trata da disponibilidade e acesso, onde os dados devem estar prontamente disponíveis a qualquer momento e em qualquer plataforma, sendo eles acessíveis e modificáveis de forma direta; a segunda premissa relaciona-se com o reúso e redistribuição, onde os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a interconexão com outros conjuntos de dados; e a terceira, a participação universal, em que todos os usuários devem conseguir usar, reutilizar e redistribuir (Open Knowledge Foundation, [2010]).

3.1.1 Relevância dos Dados Abertos

Os dados abertos ajudam a promover a transparência, aumentam a eficiência e aumentam a qualidade das decisões tomadas, podendo ser usados para analisar tendências, comparar resultados e até mesmo criar aplicativos e serviços.

Um dos benefícios no uso de dados abertos está relacionado ao modo como as pessoas utilizam a informação disponível para melhorar suas vidas. Nesse sentido, cidadãos podem usar dados abertos para descobrir problemas de saúde pública, como surtos de doenças infecciosas. Os governos também podem usar os dados abertos para aprimorar os serviços públicos, como educação e saúde. Os dados abertos também ajudam a desenvolver soluções inovadoras para problemas existentes.

Segundo Costa e Gonçalves (2020), os dados abertos são importantes para a transparência do governo. Um governo aberto é uma forma de governo que permite ao público ter acesso à informação do governo. Isso ajuda a garantir que as decisões do governo estejam fundamentadas em evidências reais. O acesso aos dados abertos também pode permitir que os cidadãos monitorem o governo para garantir que os seus interesses estejam sendo representados.

Os dados abertos também são importantes para a inovação, pois permitem que os desenvolvedores de *software* criem produtos e serviços que melhoram a vida das pessoas.

Moreira e Malin (2015, p. 14) realizaram um estudo que identificou vários aplicativos, desenvolvidos por meio de iniciativa individual, tendo foco na fiscalização e nos instrumentos de utilidade pública, presentes no Portal Brasileiro de Dados Abertos.

Os dados abertos podem desempenhar um papel importante na promoção da ciência cidadã, que envolve a participação da comunidade na coleta e análise de dados. Conforme o artigo "Ciência cidadã e dados abertos: uma abordagem para a inclusão, a participação e a responsabilização" (Kuhn, 2019, p.118), a ciência cidadã deve ser vista como um meio para as pessoas usarem os dados abertos para monitorar, avaliar e solucionar problemas ambientais, sociais, de segurança e de saúde. Além disso, os dados abertos podem contribuir para que os governos sejam mais transparentes, inovadores, podem ajudar na educação e no desenvolvimento dos negócios (Kuhn, 2019, p.120).

Os dados abertos também são cruciais para ajudar as pessoas a utilizar informações disponíveis para melhorar suas vidas. Segundo o artigo "A ciência cidadã e os dados abertos: a promessa de melhorar o acesso à informação" (Chang, 2018), os dados abertos podem ajudar a desenvolver a consciência cidadã, permitindo que as pessoas possam colaborar com o governo para melhorar o uso da informação para o bem-estar social e promover a inclusão social. Além disso, os dados abertos podem melhorar a qualidade de vida de todos os cidadãos e promover a inovação (Chang, 2018).

3.1.2 Disponibilidade e acesso

O governo brasileiro disponibiliza um portal de dados abertos[8] por meio do site "dados.gov.br" para que qualquer cidadão possa acessar, consultar e recuperar dados das mais variadas fontes para uso e reúso. Seu acesso é totalmente gratuito e intuitivo sendo possível fazer *download* de arquivos em diversos formatos como *.xml* e *.csv*, por exemplo. A disponibilidade dos dados abertos oferece uma ampla gama de benefícios para a sociedade. A promoção da transparência significa que as pessoas têm acesso a informações úteis para tomar decisões e apoiar ações comunitárias de bem-estar público. Além disso, a disponibilidade de dados abertos possibilita o acesso a informações que permitem a criação de novos negócios e a geração de riqueza. A ampliação do acesso à informação também pode ajudar a melhorar a eficiência do governo e promover o uso responsável dos recursos direcionados à população (Kitchin, 2014a).

[8] Disponível em: https://dados.gov.br/home. Acesso em: 20 fev. 2023.

Para garantir que os dados abertos sejam usados de forma segura e responsável, é necessário desenvolver políticas e modelos de governança apropriados. Estas políticas devem garantir que os dados sejam usados de forma ética e responsável, simultaneamente, em que se mantêm os direitos de propriedade intelectual. Também é necessário criar um quadro de governança que estabeleça responsabilidades claras para a implementação de políticas relacionadas aos dados abertos, como a segurança da informação, a privacidade e a preservação dos direitos autorais (Beauchamp; McCurdy, 2016).

O livre acesso aos dados abertos também pode melhorar o debate público, pois as pessoas podem ter acesso a informações sobre tópicos diversos e discutir com base em fatos. Uma das principais vantagens do acesso a dados abertos é que ele permite que as pessoas acessem as informações que estão sendo compartilhadas. Além disso, o acesso a dados abertos pode ajudar a estimular a inovação e a criatividade. Os dados abertos podem ser usados para criar produtos e serviços, como aplicativos móveis, ferramentas de análise de dados, visualizações de dados, dentre outros. Além disso, o acesso a dados abertos também pode contribuir com os serviços públicos, como por exemplo, um melhor entendimento das atividades governamentais, um gerenciamento mais eficaz dos recursos públicos, um aumento da responsabilização e da prestação de contas, a integridade pública, a criação de comunidades mais seguras e uma maior participação dos cidadãos na gestão pública (Open Government Partnership, 2023).

A disponibilidade e o acesso a dados abertos oferecem muitos benefícios para a sociedade, como acesso a informações importantes, estimulação da inovação e criatividade, e melhoria da eficiência. É importante que os governos e as empresas continuem a investir em acesso a dados abertos para melhorar a transparência, a inovação e a competitividade (Borgman; Wallis; Enyedy, 2007).

3.1.3 Benefícios

Os dados abertos são um recurso cada vez mais importante para governos, empresas e outras organizações. Eles permitem compartilhar informações com o público de forma livre e transparente, proporcionando benefícios significativos para todos os envolvidos.

Nas organizações públicas e/ou privadas os dados abertos oferecem inúmeros benefícios. De acordo com Finnemore e Hollis (2020), os dados abertos podem ajudar as organizações a melhorar a sua eficiência operacional, reduzir os custos e aumentar a segurança. Além disso, eles também podem fornecer aos tomadores de decisão uma melhor compreensão das necessidades dos clientes, permitindo que eles melhorem seus serviços.

Segundo o estudo de Sørensen e Kocksch (2021), os dados abertos podem melhorar a transparência e a responsabilidade do governo, permitindo que qualquer cidadão acesse informações sobre as atividades do governo e melhore a sua qualidade de vida. Além disso, eles também permitem que organizações, sem fins lucrativos, criem soluções inovadoras para problemas sociais e de saúde globais.

Para a economia, os dados abertos podem ser uma força importante. Segundo o estudo de Rodrigo Klein, Deisy Klein e Edimara Luciano. (2008), os dados abertos podem estimular a inovação, pois permitem que os desenvolvedores criem produtos e serviços com base em dados reais. Eles também podem ajudar as empresas a entender melhor a concorrência e a melhorar seus produtos e serviços.

Na segurança pública, os dados abertos contribuem na garantia de que os órgãos de segurança pública estejam agindo de acordo com padrões aceitáveis e que a comunidade esteja ciente dos esforços que estão sendo feitos para garantir a lei e a ordem. Além disso, ajuda a promover a responsabilização dos órgãos de segurança pública, pois as informações estarão disponíveis para todos os envolvidos, otimizando a relação de confiança entre órgãos de segurança pública e a comunidade.

Os dados abertos beneficiam, também, a inovação empresarial, pois as empresas podem usar os dados para criar serviços e produtos inovadores. Por exemplo, os dados abertos podem ser usados para criar produtos de informação que ajudam as pessoas a tomar decisões informadas sobre finanças, saúde e outros temas (West, 2018).

Os dados abertos oferecem inúmeros benefícios para todos os envolvidos. Eles podem ajudar as organizações a melhorar a sua eficiência operacional, reduzir os custos e aumentar a segurança, além de melhorar a transparência e a responsabilidade do governo. Além disso, eles permitem que as empresas entendam melhor a concorrência. Com base nesses benefícios, é evidente que os dados abertos são uma força importante para a economia global.

Neste sentido, os dados abertos podem trazer muitos benefícios, alguns dos quais incluem:

- Promover a transparência e a responsabilidade do governo;
- Aumentar a eficiência na tomada de decisões;
- Melhorar a qualidade da informação e a eficácia do governo;
- Aumentar a inovação e a criatividade;
- Apoiar a descoberta de novos conhecimentos;

- Desenvolver novos produtos e serviços;
- Melhorar a segurança cibernética;
- Melhorar a qualidade da vida das pessoas.

3.1.4 Questões de governança relacionadas aos Dados Abertos

As principais questões de governança relacionadas aos dados abertos incluem a segurança, a privacidade, a confidencialidade, a transparência e a responsabilidade. As organizações devem estabelecer políticas e procedimentos claros para garantir o uso responsável dos dados abertos. Além disso, as organizações devem certificar-se de que os dados abertos sejam usados somente para fins permitidos pela lei.

Governança de dados abertos é o processo de lidar com a preservação, uso, compartilhamento e proteção de dados. Esta governança fornece orientação sobre como a organização deve utilizar os dados abertos para alcançar seus objetivos. A governança de dados abertos deve abranger uma variedade de aspectos, incluindo a segurança, a qualidade, a privacidade, a transparência e a responsabilidade (Hemmati; Buhl, 2018).

A governança dos dados abertos envolve as seguintes questões:

- **Segurança:** A governança de dados abertos deve incluir medidas que garantam que os dados abertos sejam protegidos contra acesso não autorizado, alterações indevidas e vazamentos. Os provedores de dados abertos devem fornecer aos usuários métodos de autenticação, criptografia, controle de acesso e outras medidas de segurança (Hemmati; Buhl, 2018).
- **Qualidade**: A governança de dados abertos deve garantir que os dados sejam precisos, relevantes, consistentes e completos. Os provedores de dados abertos devem seguir práticas de qualidade para garantir que os dados sejam de qualidade. Estas práticas incluem o uso de princípios de qualidade, o monitoramento da qualidade dos dados e o uso de ferramentas de qualidade (Hemmati; Buhl, 2018).
- **Privacidade**: A governança de dados abertos deve assegurar que os dados pessoais sejam protegidos de acesso não autorizado e uso indevido. Os provedores de dados abertos devem adotar métodos que promovam a privacidade, tais como, o uso de tecnologias de anonimização, o cumprimento de regulamentos de privacidade e o estabelecimento de medidas de segurança (Hemmati; Buhl, 2018).

- **Transparência**: A governança de dados abertos deve garantir que os dados sejam acessíveis, compreensíveis e legíveis. Os provedores de dados abertos devem divulgar os dados de forma clara e transparente para que os usuários possam entendê-los facilmente. Esta transparência também deve incluir informações sobre como os dados foram coletados, como foram armazenados e como podem ser usados (Hemmati; Buhl, 2018.
- **Responsabilidade:** A governança de dados abertos deve garantir que os provedores sejam responsáveis pelo uso e manipulação dos dados, devendo assegurar que os dados sejam usados de forma ética e responsável, estabelecendo medidas para reduzir as vulnerabilidades e os riscos associados ao uso dos dados (Hemmati; Buhl, 2018).

3.1.5 Desafios

Os dados abertos oferecem aos governos, empresas e organizações a oportunidade de compartilhar e disseminar informações de maneira mais eficaz. No entanto, como qualquer tecnologia, os dados abertos também apresentam seus próprios desafios.

O primeiro desafio relacionado aos dados abertos é a sua **coleta**, pois é necessário que os dados sejam coletados e organizados de maneira que possam ser facilmente compreendidos. O segundo desafio é a **qualidade** dos dados, pois os dados devem ser precisos, consistentes e confiáveis. O terceiro desafio é a **segurança** dos dados, pois é necessário proteger os dados contra ações maliciosas e uso indevido. O quarto desafio é a **disponibilidade** dos dados, pois é necessário garantir que os dados estejam disponíveis para uso quando necessário (Khan; Sharma, 2020).

Outro desafio relacionado aos dados abertos é a **interoperabilidade**. A interoperabilidade é importante para garantir que os dados possam ser compartilhados entre diferentes sistemas, bem como para permitir que diferentes usuários acessem os dados. Isso é particularmente importante para organizações que usam diferentes sistemas para armazenar e processar dados (Khan; Sharma, 2020).

Além disso, os dados abertos também podem enfrentar desafios relacionados à sua **licença**. É importante que as organizações estabeleçam e cumpram as licenças dos dados abertos para garantir que os usuários compreendam e respeitem os direitos relacionados ao uso dos dados.

Também existem desafios relacionados à **governança** e à **adoção**. A governança é importante para garantir que os dados sejam usados de forma apropriada e segura, enquanto a adoção é importante para garantir que os dados sejam compreendidos e usados (Lukoff, 2016).

Para superar esses desafios, é importante que as organizações envolvidas desenvolvam um plano de dados abertos. Este plano deve incluir processos, procedimentos e políticas para coletar, gerenciar, armazenar, proteger e distribuir os dados abertos. Além disso, é importante que as organizações desenvolvam ferramentas e tecnologias para gerenciar e analisar os dados abertos.

Em suma, os desafios em dados abertos são muitos e complexos. No entanto, com o desenvolvimento de planos de dados abertos adequados e o uso de ferramentas e tecnologias adequadas, é possível superar os desafios e desfrutar dos benefícios dos dados abertos (Unesco, 2018).

3.2 Legislação relacionada aos Dados Abertos

Como um movimento global, os dados abertos possibilitam que informações geradas pelo governo sejam acessíveis ao público. A partir de 2017, o governo brasileiro investe em uma legislação que promove a transparência de dados públicos.

A gestão da informação é um processo elementar nos Estados, mas só recentemente, tanto no cenário nacional como internacional, que a atividade de possibilitar acesso à informação à população tem sido implementada. No Brasil, a Constituição Federal de 1988 estabeleceu o acesso à informação como um direito fundamental (Brasil, 1988).

A Lei nº 12.527, de 2011, também conhecida como Lei de Acesso à Informação, regulamentou efetivamente este direito. Como resultado desta lei, diversas plataformas, meios e políticas públicas foram desenvolvidos para promover o acesso à informação, como portais de transparência, manuais, instruções normativas e bases de dados. A referida lei garante o acesso à informação pública, em todas as esferas, conforme apresentado no artigo 1º.

> Art. 1º Esta Lei dispõe sobre os procedimentos a serem observados pela União, Estados, Distrito Federal e Municípios, com o fim de garantir o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal. (Brasil, 2011).

Estabelece, também, que o poder público viabilize tecnologias que permitam seu livre acesso.

> Art. 3º Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes:
> II - utilização de meios de comunicação viabilizados pela tecnologia da informação; (Brasil, 2011).

Esta lei também prevê mecanismos para garantir a segurança dos dados públicos, como a proteção de dados pessoais e a garantia de que as informações sejam usadas somente para fins públicos. Foi criada para incentivar a transparência e a responsabilidade do governo na divulgação de dados.

> Art. 6º Cabe aos órgãos e entidades do poder público, observadas as normas e procedimentos específicos aplicáveis, assegurar a:
> I - gestão transparente da informação, propiciando amplo acesso a ela e sua divulgação;
> II - proteção da informação, garantindo-se sua disponibilidade, autenticidade e integridade; e
> III - proteção da informação sigilosa e da informação pessoal, observada a sua disponibilidade, autenticidade, integridade e eventual restrição de acesso. (Brasil, 2011).

A Lei de Acesso à Informação prevê que os órgãos do governo devem fornecer ao público informações a respeito de suas atividades e decisões. Ela estabelece também que os órgãos do governo devem responder às solicitações feitas pelo público em um prazo determinado.

Segundo o seu art. 7. É vedado:

> Art. 7º O acesso à informação de que trata esta Lei compreende, entre outros, os direitos de obter:
> IV - informação primária, íntegra, autêntica e atualizada;
> V - informação sobre atividades exercidas pelos órgãos e entidades, inclusive as relativas à sua política, organização e serviços;
> VI - informação pertinente à administração do patrimônio público, utilização de recursos públicos, licitação, contratos administrativos; e
> VII - informação relativa:
> a) à implementação, acompanhamento e resultados dos programas, projetos e ações dos órgãos e entidades públicas, bem como metas e indicadores propostos;
> ;b) ao resultado de inspeções, auditorias, prestações e tomadas de contas realizadas pelos órgãos de controle interno e externo, incluindo prestações de contas relativas a exercícios anteriores (Brasil, 2011).

O Decreto nº 8.777, de 11 de maio de 2016 (Brasil, 2016), que institui a Política de Dados Abertos do Poder Executivo Federal, é o marco legal que regulamenta o acesso a dados públicos no Brasil. Este Decreto estabelece que os dados devem ser disponibilizados de forma aberta, livre e gratuita, garantindo que a informação seja acessível a todos, conforme seus artigos 1º e 4º.

> Art. 1º Fica instituída a Política de Dados Abertos do Poder Executivo federal, com os seguintes objetivos:
> I - promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos;
> II - aprimorar a cultura de transparência pública;
> III - franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso;
> IV - facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação;
> V - fomentar o controle social e o desenvolvimento de novas tecnologias destinadas à construção de ambiente de gestão pública participativa e democrática e à melhor oferta de serviços públicos para o cidadão;
> VI - fomentar a pesquisa científica de base empírica sobre a gestão pública;
> VII - promover o desenvolvimento tecnológico e a inovação nos setores público e privado e fomentar novos negócios;
> VIII - promover o compartilhamento de recursos de tecnologia da informação, de maneira a evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações; e
> IX - promover a oferta de serviços públicos digitais de forma integrada.
> Art. 4º Os dados disponibilizados pelo Poder Executivo federal e as informações de transparência ativa são de livre utilização pelos Poderes Públicos e pela sociedade.O acesso aos dados abertos é amplo, gratuito e irrestrito (Brasil, 2016).

O referido Decreto permite que os cidadãos tenham acesso a informações públicas dos órgãos governamentais, como dados demográficos, informações sobre serviços públicos, licenças de trabalho, finanças e dados sobre políticas públicas. Essas informações devem ser fornecidas de forma rápida, oportuna e gratuita, conforme a lei, que estabelece que os órgãos governamentais devem manter os dados disponíveis em um formato digital, que seja aberto e acessível para todos (Brasil, 2016).

Este ato exige que os órgãos governamentais publiquem dados em seu site oficial, a fim de garantir que as informações sejam facilmente acessíveis para os cidadãos. Além disso, os órgãos governamentais devem fornecer mecanismos para que os cidadãos possam acessar e compartilhar os dados de forma segura (Marcolin, 2016).

A promulgação desse decreto foi uma grande conquista para a transparência e a participação cívica no Brasil. Ele abriu as portas para que os cidadãos tenham acesso às informações públicas, o que é fundamental para o envolvimento cívico e o desenvolvimento socioeconômico.

A Lei nº 13.709, de 14 de agosto de 2018 (Brasil, 2018), conhecida como Lei Geral de Proteção de Dados (LGPD), foi criada para estabelecer regras para o tratamento de dados pessoais, proteger os direitos fundamentais de liberdade e de privacidade das pessoas, conforme seu artigo primeiro.

> Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. (Brasil, 2018).

Esta lei estabelece que os dados pessoais devem ser tratados de forma adequada e garante que as informações sejam usadas somente para fins específicos.

> Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
> I - finalidade: realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades;
> II - adequação: compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento (Brasil, 2018).

Em 8 de julho de 2019 foi promulgada a Lei nº 13.853, que altera a Lei nº 13.709/2018 (LGPD) e dispõe sobre a proteção de dados pessoais e cria a Autoridade Nacional de Proteção de Dados.

A LGPD estabelece também que os dados pessoais devem ser armazenados de forma segura e protegidos contra uso indevido. Além disso, a lei prevê que as organizações devem fornecer ao cidadão a possibilidade de acessar, verificar e corrigir os seus dados pessoais.

Ela regulamenta a coleta, o tratamento, o compartilhamento, a transferência e o armazenamento de dados pessoais, definindo as responsabilidades de todos os agentes de tratamento de dados. A Lei também estabelece direitos dos titulares dos dados, como o direito de acesso aos dados, o direito de retificação e o direito de exclusão.

Além disso, a LGPD estabelece regras para que os dados sejam tratados com segurança, protegendo os dados contra perda, danos ou uso inadequado. Por isso, é necessário que as empresas implementem medidas de segurança adequadas para proteger os dados pessoais e garantir que eles sejam tratados conforme a lei.

Estabelece sanções administrativas e penais para aqueles que descumprirem as disposições da lei. As penalidades previstas para aqueles que descumprirem a lei variam de advertências até multas de até 2% do faturamento bruto global da empresa.

Além disso, a LGPD também estabelece a obrigatoriedade de um Encarregado de Proteção de Dados (EPD), sendo uma pessoa responsável por garantir que a lei seja seguida e por assegurar que os direitos dos titulares dos dados sejam respeitados. Com a promulgação da Lei Geral de Proteção de Dados, as empresas e os agentes de tratamento

de dados têm a responsabilidade de garantir a proteção dos dados pessoais e de respeitar os direitos fundamentais de liberdade e de privacidade das pessoas.

Conforme esta Lei nº 13.853/2019, para que os dados pessoais sejam tratados, é necessário que o titular dos dados dê seu consentimento explícito para que seu dado seja coletado e tratado. Além disso, também é necessário que existam mecanismos de segurança para proteger os dados pessoais e garantir que eles sejam tratados de forma correta e segura.

A Lei Geral de Proteção de Dados também estabelece direitos dos titulares dos dados, como o direito de acesso aos dados, o direito de retificação ou o direito de exclusão. Estes direitos garantem que os titulares dos dados possam ter acesso a seus dados, corrigi-los ou excluí-los, quando necessário.

Estas três leis estabelecem as regras para o uso, armazenamento e proteção dos dados públicos no Brasil. Elas asseguram que os dados sejam tratados de forma adequada e garantem que a informação seja acessível ao público. Estas leis também incentivam o governo a fornecer dados de forma aberta e transparente, melhorando a governança e a qualidade dos serviços públicos.

3.2.1 Direitos autorais e direitos de propriedade intelectual e industrial relacionados aos dados abertos

Os direitos autorais e os de propriedade intelectual são uma importante consideração para o compartilhamento de dados abertos. Os direitos autorais poderão ser aplicados a dados que contenham conteúdos criativos ou informações originais, bem como aos dados que sejam criados durante o processo de coleta de dados. Além disso, os direitos de propriedade intelectual podem ser usados para proteger os dados abertos, especialmente aqueles que possuem um valor comercial, que possam ser usados para fins lucrativos ou que sejam criados para esse fim (Koehler, 2018).

No entanto, algumas licenças de direitos de autor também permitem que os dados sejam usados e compartilhados de forma livre e aberta. A licença CC0 conhecida como "Dedicação ao Domínio Público" da Creative Commons, permite que criadores de obras intelectuais, como cientistas, educadores, artistas e outros detentores de direitos de autor ou proteções de base de dados, renunciem a esses interesses e coloquem suas obras no domínio público. Essa licença permite que outras pessoas construam livremente sobre o trabalho original, o aprimorem e o reutilizem para quaisquer propósitos sem restrições sob a lei de direitos autorais ou de bases de dados (Koehler, 2018).

Além disso, existem licenças de direitos de autor que podem ser aplicadas aos dados abertos, como a Open Data Commons Attribution License (ODC-BY) e a Open Data Commons Open Database License (ODBL).

A ODC-BY permite que os dados sejam usados e compartilhados de forma livre e aberta, desde que os autores sejam creditados (Koehler, 2018). A ODBL permite que os dados sejam usados para fins comerciais, desde que sejam atribuídos créditos aos autores (Koehler, 2018).

Portanto, é importante que os titulares de direitos autorais sejam conscientes de como os seus direitos autorais podem afetar o compartilhamento de dados abertos. Além disso, é importante que eles escolham uma licença de direitos autorais, que seja apropriada para os seus dados abertos, de modo a permitir que os dados sejam usados e compartilhados de forma livre e aberta.

É importante também que os titulares de direitos autorais tenham em mente que os direitos autorais não são infinitos. Eles são limitados no tempo, portanto, é importante que entendam quando seus direitos autorais expiram. Nos Estados Unidos, os direitos autorais duram por 70 anos após a morte do autor (Koehler, 2018). Além disso, os titulares de direitos autorais devem considerar a possibilidade de renunciar aos seus direitos autorais, permitindo que os dados abertos sejam usados de forma livre e aberta.

Os direitos de propriedade intelectual também são importantes para o compartilhamento de dados abertos. Por exemplo, a European Union Database Directive (Directiva da UE sobre Base de Dados) protege a base de dados na totalidade, ao invés de proteger cada elemento individualmente. Isso permite que os dados sejam compartilhados de forma livre e aberta, desde que os direitos de propriedade intelectual sejam respeitados (Koehler, 2018).

As leis de privacidade também devem ser consideradas ao compartilhar dados abertos. Por exemplo, a General Data Protection Regulation (GDPR) oferece proteção aos dados abertos, permitindo que os usuários compartilhem os dados com segurança. A GDPR também estabelece regras para o processamento de dados pessoais, como o direito de os usuários se oporem ao processamento de seus dados pessoais (Koehler, 2018).

Portanto, ao compartilhar dados abertos, é importante que os titulares de direitos autorais, titulares de direitos de propriedade intelectual e responsáveis pelo processamento de dados sejam conscientes dos seus direitos e responsabilidades. Ao considerar todas as

possíveis implicações legais, você pode garantir que os dados sejam compartilhados de forma segura e responsável.

3.2.2 Direitos de privacidade e proteção de dados

A privacidade e a proteção de dados são direitos fundamentais para as pessoas que desejam compartilhar informações pessoais. Esses direitos asseguram que os dados sejam protegidos e usados apenas para fins legítimos. Segundo a Declaração Universal dos Direitos Humanos, "Toda pessoa tem direito à privacidade, à liberdade de expressão e à segurança de seus dados pessoais" (United Nations, [1948], Artigo 12).

A privacidade e a proteção de dados também são fundamentais para o desenvolvimento de inovações baseadas em dados. A privacidade e a proteção de dados permitem que as pessoas compartilhem informações pessoais com a confiança de que seus dados estarão seguros.

A privacidade e a proteção de dados também são fundamentais para a criação de sistemas de dados abertos. Sistemas de dados abertos permitem que os usuários compartilhem infinitas quantidades de dados, mas esses dados precisam ser protegidos contra usos abusivos. A privacidade e a proteção de dados asseguram que os dados sejam usados de forma responsável e que os usuários sejam protegidos contra abusos e vazamentos.

A privacidade e a proteção de dados também são fundamentais para o desenvolvimento de novas tecnologias. Por exemplo, as tecnologias baseadas em dados, como a Inteligência Artificial (IA) e o Big Data, são fundamentais para o desenvolvimento de sistemas de tomada de decisão inteligentes. No entanto, essas tecnologias também precisam de privacidade e proteção de dados para funcionar corretamente.

Os direitos de privacidade e proteção de dados são fundamentais para a democracia. Isso porque os direitos de privacidade e proteção de dados permitem que as pessoas compartilhem informações pessoais com a confiança de que seus dados estarão seguros. Isso também permite que os governos e as empresas sejam responsáveis por seu uso de dados, ajudando a criar um ambiente mais seguro e democrático.

As leis de privacidade e proteção de dados são fundamentais para a proteção dos direitos de privacidade e proteção de dados. Essas leis criam direitos específicos para as pessoas, incluindo o direito de controlar seus dados, o direito de saber como seus dados estão sendo usados e o direito de exigir que seus dados sejam apagados ou corrigidos. Essas

leis ajudam a proteger os direitos de privacidade e proteção de dados e a assegurar que os dados sejam usados de forma responsável.

A privacidade e a proteção de dados são fundamentais para o desenvolvimento econômico e social. Os direitos de privacidade e proteção de dados permitem que as pessoas compartilhem informações pessoais com a confiança de que seus dados estarão seguros. Isso estimula o desenvolvimento de novas tecnologias, cria setores de emprego e ajuda as pessoas a obter melhores serviços.

De forma resumida, a privacidade e a proteção de dados são direitos fundamentais para as pessoas. Esses direitos asseguram que os dados sejam usados de forma responsável e protegidos contra usos abusivos. Eles permitem que as pessoas compartilhem informações pessoais com a confiança de que seus dados estarão seguros, o que é essencial para o desenvolvimento econômico e social.

3.2.3 Usos abusivos de dados abertos

A utilização de dados abertos não é necessariamente um processo seguro. Os usos abusivos de dados abertos, como roubo de dados, uso não autorizado de dados e acesso não autorizado, são preocupações comuns para aqueles que trabalham com dados abertos (Tandoc, 2018, p. 312). Embora seja possível usar dados abertos para fins inovadores, existem também muitas maneiras de usar os dados abertos de forma abusiva.

Os usos abusivos de dados abertos são um motivo de preocupação para aqueles que trabalham com dados abertos. Os usos abusivos podem incluir roubo de dados, uso não autorizado de dados e acesso não autorizado. É importante que as pessoas e as organizações que trabalham com dados abertos sejam conscientes dos riscos de uso abusivo e tomem as medidas necessárias para proteger seus dados.

O roubo de dados ocorre quando alguém se apropria de dados abertos sem autorização. Isso é feito normalmente para fins lucrativos, como a venda dos dados abertos para um terceiro. Isso também pode ser feito para fins mais sombrios, como extorsão ou espionagem.

Além disso, o uso não autorizado de dados abertos também é um uso abusivo comum. Esse tipo de uso abusivo ocorre quando alguém usa dados abertos para fins que não são autorizados. Por exemplo, alguém pode usar dados abertos para fins comerciais sem autorização. Isso pode ocorrer quando os dados abertos são usados sem o conhecimento ou consentimento dos proprietários dos dados.

Por fim, o acesso não autorizado de dados é outra forma de uso abusivo de dados abertos. Esse tipo de uso abusivo ocorre quando alguém acessa dados abertos sem autorização. Isso pode incluir acesso a dados confidenciais ou sigilosos.

O uso abusivo dos dados abertos pode ser evitado usando tecnologias de segurança. Por exemplo, a criptografia de dados pode ser usada para garantir que os dados abertos sejam protegidos contra usos não autorizados. A autenticação de usuários também pode ser usada para garantir que apenas aqueles que têm autorização para acessar os dados abertos possam acessá-los. Além disso, o controle de acesso pode ser usado para garantir que os usuários não acessem dados confidenciais ou sigilosos sem autorização.

Além disso, existem também tecnologias desenvolvidas especificamente para aproveitar os dados abertos. Por exemplo, o Linked Data é uma arquitetura de rede que permite que os dados abertos sejam compartilhados e acessados de maneira mais fácil e eficiente. O Interledger Protocol (ILP) é outra arquitetura de rede que permite a transferência de valores usando dados abertos.

É importante destacar que o uso abusivo de dados abertos pode ter consequências graves para aqueles que trabalham com os dados abertos. Por exemplo, o vazamento de dados pode levar ao roubo de identidade e à exposição de informações confidenciais. Além disso, o uso abusivo de dados abertos pode levar a multas e sanções por violações de privacidade.

Portanto, é importante que aqueles que trabalham com dados abertos compreendam os riscos do uso abusivo e adotem as medidas necessárias para proteger seus dados. Isso inclui o uso de tecnologias de segurança, a implementação de políticas de segurança de dados e a adesão às leis e regulamentos relacionados à privacidade.

Também é importante que as pessoas e as organizações que trabalham com dados abertos se familiarizem com as tecnologias relacionadas a dados abertos. Isso inclui: Linked Data, Interledger Protocol, Big Data, Inteligência Artificial, APIs, Internet das Coisas, Web 2.0 e computação em nuvem. Ao compreender essas tecnologias, as pessoas e as organizações podem aproveitar melhor os dados abertos e usá-los de forma segura.

Os usos abusivos dos dados abertos são um motivo de preocupação para aqueles que trabalham com essa tecnologia. Por essa razão, é importante que estes profissionais estejam conscientes dos riscos do uso abusivo e tomem as medidas necessárias para proteger seus dados. Além disso, também é importante que as pessoas e as organizações se

familiarizem com as tecnologias relacionadas ao uso e reúso de dados abertos para aproveitá-los melhor e de forma segura.

3.2.4 Segurança de dados

A segurança de dados é fundamental para a adoção de dados abertos. Como afirma Costa (2021, p. 3), "[...] a segurança dos dados abertos é o mecanismo que garante a proteção dos dados e impede o uso abusivo ou inadequado". Os mecanismos de segurança, portanto, são essenciais para assegurar que os dados abertos sejam protegidos de usos abusivos.

De acordo com Costa (2021), as principais técnicas para a segurança de dados abertos são a criptografia, a autenticação e a confidencialidade. A criptografia é usada para proteger os dados e impedir o acesso não autorizado. A autenticação garante que os usuários acessem e compartilhem seus dados e a confidencialidade assegura que apenas usuários autorizados possam acessar e usar os dados.

Além disso, a segurança dos dados abertos também envolve a gestão adequada dos dados em relação à monitoração e à avaliação dos dados para garantir que correspondam com as políticas de segurança de cada organização. Além disso, a gestão adequada dos dados também inclui a implementação de medidas de segurança para proteger o acesso aos dados.

A segurança lógica e física dos dados é importante para a segurança de dados abertos. No que se refere à segurança lógica, inclui o uso de autenticação em dois fatores, o uso de senhas fortes e o uso de mecanismos de verificação de integridade. Já em relação à segurança física, abrange o uso de controles de acesso físico, como a utilização de câmeras de vigilância, a utilização de sistemas de alarme e a utilização de códigos de acesso. Além disso, o uso de mecanismos de *backup* também é importante para garantir que os dados sejam seguros mesmo em caso de falhas. Estas medidas permitem garantir que os dados estejam seguros e que apenas usuários autorizados possam acessá-los.

A implementação de práticas de segurança corporativas é outra medida importante para garantir a segurança dos dados abertos. Estas práticas incluem a implementação de políticas de segurança, o treinamento dos usuários, a implementação de um sistema de monitoramento para detectar ameaças à segurança, o uso de proteções de rede, incluindo o uso de *firewalls*, filtragem de conteúdo, sistemas de detecção de intrusão e criptografia de rede, para evitar o acesso não autorizado aos dados. Estas práticas podem contribuir na

garantia que os dados abertos sejam seguros e que os usuários estejam cientes dos riscos associados ao uso destes dados.

A segurança dos dados abertos é uma questão contínua, assim, é importante monitorar os dados e os mecanismos de segurança constantemente para garantir que eles concordem com as políticas de segurança corporativas e ficar atento às novas tendências do setor e às novas vulnerabilidades que possam surgir.

3.3 Gerenciamento e compartilhamento de dados abertos

O gerenciamento e compartilhamento de dados abertos é um tema de grande importância para a sociedade. Esta tecnologia permite que os dados sejam acessados, compartilhados e usados de maneira segura. O gerenciamento e compartilhamento fornecem aos usuários ferramentas para o acesso e uso seguros dos dados. Além disso, também permitem que os usuários compartilhem os dados de forma segura e eficaz.

Ao compartilhar dados abertos, os usuários podem aproveitar vantagens como a transparência e o acesso a informações. Isso aumenta a eficiência e a produtividade de um sistema, simultaneamente, em que aumenta a segurança dos dados. No entanto, é necessário que os usuários entendam os direitos autorais e as questões de privacidade e proteção de dados, que estão relacionadas ao gerenciamento e compartilhamento de dados abertos.

De acordo com Gunn e Poulter (2019),

> [...] as leis de direitos autorais relacionadas a dados abertos são um dos principais fatores que influenciam o gerenciamento e compartilhamento de dados abertos. O Open Data Commons Attribution License (ODCAL) e o Open Data Commons Open Database License (ODCODL) são alguns dos principais direitos autorais relacionados a dados abertos. Estas leis protegem o uso dos dados e garantem que as informações sejam compartilhadas de forma segura (Gunn; Poulter, 2019, p. 91).

Além disso, as leis de privacidade e proteção de dados também desempenham um papel importante no gerenciamento e compartilhamento de dados abertos. Estas leis fornecem aos usuários direitos de proteção de dados e garantem que os dados sejam usados de forma responsável. Estas leis também protegem os direitos de privacidade dos usuários e garantem que os dados não sejam usados para fins abusivos.

Com isso, os usuários poderão aproveitar ao máximo os dados abertos e desfrutar dos benefícios que eles oferecem. “O gerenciamento e compartilhamento de dados abertos fornece aos usuários ferramentas para o acesso e uso seguros dos dados” (Moura, 2020, p. 3.8).

3.3.1 Armazenamento de dados abertos

O armazenamento de dados abertos significa armazenar e permitir acesso a informações públicas de forma segura. O armazenamento de dados abertos permite que organizações e governos disponibilizem informações para uso público, gerando conhecimento e maior transparência. Segundo a União Europeia, "os dados devem estar disponíveis em um formato estruturado, ou seja, em um formato que possa ser facilmente armazenado, processado e reutilizado" (European Union, 2023, p. 5).

O armazenamento de dados abertos envolve diversas tecnologias e ferramentas, tais como bancos de dados, serviços de nuvem, serviços de *streaming*, serviços de *backup* e armazenamento em *cache*. Os bancos de dados são usados para armazenar e organizar os dados, enquanto o armazenamento em cache é usado para aumentar o desempenho de aplicações que acessam dados frequentemente. Serviços de *streaming* e serviços de nuvem permitem acesso aos dados a partir de qualquer lugar, enquanto os serviços de *backup* permitem a recuperação de dados perdidos ou corrompidos.

Além disso, o armazenamento de dados abertos também deve envolver mecanismos de segurança para garantir que os dados abertos sejam protegidos contra usos abusivos, tais como criptografia de dados, autenticação de usuários e controle de acesso. O armazenamento de dados abertos deve ser feito de forma segura, para garantir que os dados sejam protegidos e não comprometidos.

Ao usar dados abertos, é importante que as organizações e governos usem as melhores práticas de segurança para garantir que os dados sejam armazenados de forma segura. Essas práticas incluem o uso de soluções de segurança criptográfica, tais como criptografia de dados, autenticação de usuários e controle de acesso. Além disso, é importante que as organizações e governos usem práticas de segurança para garantir que os dados sejam armazenados de forma segura contra usos abusivos, tais como roubo de dados, uso e acesso não autorizado.

Portanto, o armazenamento seguro de dados abertos é essencial para garantir que as informações sejam disponibilizadas de forma segura e protegida.

O armazenamento de dados abertos também é essencial para aproveitar os benefícios da análise que envolve a coleta, organização e análise de dados abertos para gerar informações importantes. Também é essencial para o desenvolvimento de novas tecnologias como Big Data, IA, APIs, IoT, Web 2.0 e computação em nuvem.

Portanto, o armazenamento seguro é uma parte importante da utilização permitindo que organizações e governos disponibilizem informações para uso público, gerando conhecimento e maior transparência.

É fundamental considerar a capacidade de armazenamento de grandes volumes de dados. À medida que a quantidade de dados abertos disponíveis aumenta, organizações e governos necessitam de soluções de armazenamento capazes de gerir o crescente volume de informações. Tais soluções devem ser flexíveis e escalonáveis, para facilitar a adição de novos dados de maneira simples.

O armazenamento de dados abertos deve incluir arquiteturas que otimizem o uso dessas informações, tais como o *Linked Data* e o *Interledger Protocol*. O *Linked Data* permite que os dados sejam conectados de forma segura, permitindo que os usuários compartilhem dados facilmente. O *Interledger Protocol* permite que os usuários interliguem redes de dados em diferentes locais, permitindo que os usuários acessem dados de várias fontes.

Além disso, o armazenamento de dados abertos também deve envolver auditorias de dados para garantir que os dados sejam armazenados de forma segura. As auditorias de dados podem ajudar as organizações e governos a analisar os dados para detectar potenciais problemas de segurança. As auditorias de dados também podem ajudar a identificar problemas de desempenho, como falhas de segmentação e armazenamento inadequado.

O armazenamento de dados abertos também deve incluir os princípios do movimento de Dados Abertos. Esses princípios incluem a disponibilização dos dados de forma aberta, o acesso aos dados por todos os usuários, a reutilização dos dados e a responsabilização dos usuários. Esses princípios ajudam a garantir que os dados sejam armazenados de forma segura e permitem que as organizações e governos acessem e compartilhem dados de forma segura.

O armazenamento de dados abertos significa armazenar e permitir acesso a informações públicas de forma segura. O armazenamento de dados abertos envolve diversas tecnologias e ferramentas, tais como bancos de dados, serviços de nuvem, serviços de streaming, serviços de backup e armazenamento em cache. Além disso, o armazenamento de dados abertos também deve envolver mecanismos de segurança para garantir que os dados abertos sejam protegidos contra usos abusivos.

3.3.2 Redes de dados

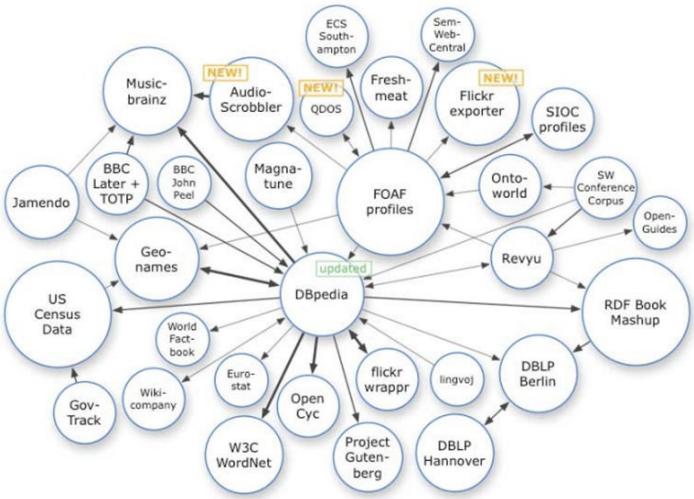
A apropriação de dados abertos é um tema importante atualmente, pois a disponibilidade desses dados pode oferecer benefícios substanciais para as organizações, sejam elas públicas ou privadas. Dentro deste contexto, alguns protocolos de rede de dados são fundamentais. De acordo com Mukherjee, Ghosh e Zhang, (2015, p. 276), "as arquiteturas de rede permitem o compartilhamento de dados entre diferentes sistemas, fornecendo às organizações a capacidade de aproveitar os dados abertos (...)".

Estes protocolos são fundamentais para aproveitar os dados abertos, pois permitem a conexão entre sistemas e acesso a informações. Existem diversos protocolos de rede que fornecem diferentes benefícios, dentre eles, pode-se destacar o *Linked Data* e o *Interledger Protocol*.

O *Linked Data* é uma arquitetura de rede que permite a organização dos dados em formato de grafos, possibilitando a interconexão entre os dados (Heath; Bizer, 2011). Esse protocolo é focado na conectividade e interoperabilidade dos dados, possibilitando que informações de diferentes fontes possam ser facilmente relacionadas, facilitando a recuperação de informações relevantes.

Ele se baseia na Web Semântica, dando suporte às técnicas de *Linked Data*, que permitem conectar informações diferentes mesmo entre fontes disjuntas. Dessa forma, ele ajuda a tornar os dados armazenados mais acessíveis, interconectados e reutilizáveis. Isso porque as informações são organizadas em formato de grafos, onde os "nós", representam, os dados e as "arestas" representam as ligações entre eles. A Figura 3 ilustra o formato de interconexão dos dados por meio do *Linked Data.*

Figura 3 – Interconexão dos dados por meio do *Linked Data*



Fonte: Bizer *et al.* (2008).

Vários serviços e aplicações podem se beneficiar do referido protocolo, dentre eles, os buscadores, que podem usar o *Linked Data* para melhorar o acesso aos dados, permitindo que os usuários possam encontrar informações relevantes mais facilmente. Outras aplicações, como as redes sociais, também se beneficiam desta arquitetura, pois ela facilita a gerência de dados e a criação de ligações entre informações de diferentes fontes. Além disso, ela usa o formato RDF para representar os dados, o que significa que os dados podem ser entendidos por todas as aplicações na Web.

O Interledger Protocol (ILP) é um protocolo de comunicação entre dispositivos muito utilizado pela tecnologia *blockchain*[9]*,* que permite a transferência de dados entre sistemas distintos. Desenvolvida pela Ripple Labs[10], essa arquitetura é usada para conectar várias redes de pagamento, permitindo que pagamentos sejam executados entre as mesmas (Krüger, 2020).

Mais do que isso, o ILP é uma série de protocolos que trabalham em conjunto para permitir que o dinheiro possa fluir entre várias redes distintas, independente do sistema usado. A arquitetura está construída sobre uma camada de protocolos que permitem transações entre partes diferentes, usando qualquer moeda digital (Krüger, 2020).

O principal objetivo do ILP é tornar os processos de pagamento mais rápidos e acessíveis. Ao permitir a conexão entre diferentes sistemas, ele proporciona maior flexibilidade e acessibilidade, permitindo que pagamentos sejam realizados de forma mais ágil e segura (Krüger, 2020).

A utilização desses protocolos e arquiteturas de rede permite às organizações aproveitar os dados abertos de forma mais eficiente, em comparação aos métodos tradicionais, pois proporcionam o armazenamento e compartilhamento de dados em forma de grafos, facilitando a interconexão entre os dados, viabilizando maior eficácia no uso dos dados.

Portanto, a utilização de protocolos de rede é fundamental para aproveitar os dados abertos, uma vez que possibilitam que as organizações usem os dados de forma mais eficiente e ampla, permitindo que eles sejam armazenados, compartilhados e interconectados.

[9] Blockchain é um banco de dados distribuído que armazena registros de transações criptografados e conectados em blocos que formam uma cadeia, tornando o sistema seguro contra alterações ou exclusões indevidas (Singhal; Dhameja; Panda, 2018).

[10] Disponível em: https://ripple.com/. Acesso em: 20 fev. 2023.

3.3.3 Tecnologias relacionadas a dados abertos

A crescente popularidade de dados abertos faz com que muitas tecnologias relacionadas a essa área sejam desenvolvidas. Essas tecnologias pretendem facilitar o acesso, o uso e o compartilhamento desses dados. Dentre elas, pode-se citar o Big Data, a Inteligência Artificial (IA), APIs (*Application Programming Interface*), Internet das Coisas (IoT), Web 2.0 e computação em nuvem.

As tecnologias de Big Data, Inteligência Artificial, APIs, IoT, Web 2.0 e computação em nuvem proporcionam novas formas de armazenar, analisar e compartilhar dados em escalas maiores e mais complexas. A integração entre essas tecnologias permite aos usuários obter informações mais relevantes, bem como identificar padrões e tendências com rapidez e precisão. Além disso, elas permitem que dados de várias fontes sejam reunidos e organizados de forma rápida e eficaz, contribuindo para que usuários possam tomar decisões mais assertivas.

Essas tecnologias também permitem que os dados sejam acessíveis a qualquer usuário em qualquer lugar do mundo, melhorando a colaboração entre equipes de diferentes departamentos e facilita a troca de informações entre eles. Os avanços na área dos dados abertos também contribuem para o desenvolvimento de tecnologias inovadoras que permitem aos usuários criar aplicações mais inteligentes e personalizadas, tornando a vida mais fácil.

Big Data é um termo usado para descrever o enorme volume de dados que estão sendo armazenadas no mundo todo. O Big Data se tornou uma importante ferramenta para a análise de dados abertos, pois permite aos usuários acessar e analisar grandes quantidades de dados. Além disso, ele também pode ajudar na tomada de decisões e na criação de novas soluções (Koch, 2018, p. 5).

O Big Data é um fenômeno que reflete a crescente quantidade de dados armazenados no mundo moderno. Esses dados variam de dados estruturados, tais como banco de dados e registros públicos, a dados não estruturados, como texto, áudio, vídeo, imagem e outros conteúdos multimídia. Esta vasta quantidade de dados disponíveis abriu disponibilizou novas possibilidades de análise, que nos permitem identificar padrões a partir de informações disponíveis. Além disso, o Big Data também é usado como uma ferramenta para a tomada de decisões, ajudando a criar soluções inovadoras baseadas em tendências atuais. Assim, o Big Data é um fenômeno que demonstra a importância crescente do processamento de grandes volumes de dados na nossa sociedade (Koch, 2018, p. 5).

Inteligência Artificial (IA) é outra tecnologia que tem sido usada para facilitar o acesso, uso e compartilhamento de dados abertos. A IA permite que os usuários criem e desenvolvam sistemas inteligentes que possam tomar decisões baseadas em dados abertos. Além disso, a IA é usada para analisar grandes quantidades de dados e identificar padrões e tendências (Koch, 2018, p. 6).

Outra tecnologia importante é a API, usada para acessar e compartilhar dados. As APIs permitem que os usuários criem aplicativos que podem acessar dados de diferentes fontes e compartilhá-los de forma eficiente. Isso permite que os usuários criem aplicativos personalizados usando os dados abertos (Koch, 2018, p. 7).

A Internet das Coisas (IoT) também tem sido usada para facilitar o acesso, o uso e o compartilhamento de dados abertos. A IoT permite que os usuários criem aplicativos que podem conectar dispositivos físicos à internet e compartilhar dados entre eles. Isso permite que os usuários obtenham dados em tempo real, o que pode ser usado para tomar decisões e criar soluções (Koch, 2018, p. 8).

A Web 2.0 também tem sido usada para facilitar o acesso, o uso e o compartilhamento de dados abertos. A Web 2.0 é uma plataforma que permite que os usuários criem e compartilhem conteúdos usando diferentes ferramentas e meios. Além disso, ela permite que os usuários acessem e compartilhem dados rapidamente, o que pode ser usado para criar soluções inovadoras (Koch, 2018, p. 9).

A computação em nuvem também tem sido usada para facilitar o acesso, o uso e o compartilhamento de dados abertos. A computação em nuvem permite que os usuários armazenem e compartilhem dados facilmente. Isso permite que os usuários acessem e compartilhem dados de forma segura e eficiente (Koch, 2018, p. 10).

Essas são algumas das tecnologias que estão sendo usadas para facilitar o acesso, o uso e o compartilhamento de dados abertos. O Big Data, a Inteligência Artificial (IA), APIs, Internet das Coisas (IoT), Web 2.0 e computação em nuvem são exemplos de tecnologias que estão sendo usadas para essa finalidade. Essas tecnologias estão ajudando a tornar os dados abertos mais acessíveis, úteis e compartilháveis.

3.4 Considerações relevantes sobre esta seção

Concluindo a seção, entende-se que o conceito de Dados Abertos é extremamente relevante para o desenvolvimento de qualquer área, pois possibilita que as informações fiquem disponíveis para todos. Apesar de terem diversos benefícios, existem algumas

questões de governança relacionadas aos dados abertos que precisam ser consideradas. Além disso, existem também leis e direitos que devem ser respeitados no que diz respeito ao seu uso, sendo também importante garantir que os mecanismos de segurança para dados abertos sejam aplicados corretamente para evitar usos abusivos.

Os dados abertos desempenham um papel fundamental no desenvolvimento de uma sociedade moderna. O uso de mecanismos de segurança eficazes, bem como o uso de tecnologias para gerenciar e compartilhar dados abertos, são fundamentais para garantir que os dados abertos sejam usados de forma apropriada. Com isso, os dados abertos tornam-se uma ferramenta poderosa para o avanço da sociedade.

Para que os dados abertos sejam efetivamente usados, é necessário que haja um compromisso com a ética, a transparência e a responsabilidade. A conscientização das pessoas sobre o uso responsável desses dados é fundamental para que eles possam ser usados adequadamente para o benefício de todos.

Na próxima seção, “Dados Abertos Conectados”, será abordado como as tecnologias podem ajudar a melhorar a disponibilidade, acesso, gerenciamento e compartilhamento de dados abertos. Estas tecnologias incluem armazenamento de dados, protocolos de rede e outras tecnologias relacionadas a dados abertos que possibilitam aproveitá-los de forma mais eficiente. Assim, o uso destas tecnologias para o gerenciamento e compartilhamento de dados abertos pode torna-los os dados mais acessíveis e úteis para as pessoas e organizações.

4 DADOS ABERTOS CONECTADOS

O movimento de dados abertos tem ganhado força nas últimas décadas, com governos e organizações em todo o mundo adotando políticas para tornar seus dados mais acessíveis e utilizáveis pelo público. No entanto, a evolução dos dados abertos tradicionais para os dados abertos conectados é uma das principais tendências recentes na área (Rautenberg *et al*., 2017). Os dados abertos conectados (DAC) se baseiam em um conjunto de melhores práticas para organizar, publicar, conectar e compartilhar dados na web de forma aberta e transparente, utilizando padrões da Web Semântica para conectar diferentes conjuntos de dados (Isotani; Bittencourt, 2015).

Esta seção apresenta os conceitos fundamentais relacionados aos DAC. Iniciando com a contextualização histórica e destacando a sua importância, a seção descreve os padrões e tecnologias utilizados para permitir a publicação, interconexão, uso e aplicabilidade dos DAC. Serão expostos exemplos de projetos e iniciativas bem-sucedidas, para demonstrar o potencial dos DAC para gerar novos conhecimentos para a sociedade. Compreender esses conceitos é fundamental para aproveitar ao máximo as possibilidades oferecidas pelos dados abertos e para desenvolver aplicações mais eficazes e inovadoras.

4.1 Definição e Contextualização

Os dados abertos conectados são uma evolução dos dados abertos tradicionais, que se baseiam em um conjunto de melhores práticas para organizar, publicar, conectar e compartilhar dados na web de forma aberta e transparente. A ideia de conectar dados abertos surgiu em 2006, quando Tim Berners-Lee, o inventor da *World Wide Web*, propôs a ideia de *Linked Data* (Bizer; Heath; Berners-Lee, 2009). A proposta consistia em utilizar padrões da Web Semântica para conectar diferentes conjuntos de dados e permitir que as máquinas entendessem o significado dos dados, materializando a Web Semântica (Santarém Segundo, 2018).

Em 2009, Berners-Lee lançou o Linked Open Data Cloud[11] (LODC), um conjunto de mais de 30 bilhões de links entre diferentes conjuntos de dados abertos. Desde então, a ideia de dados abertos conectados tem sido amplamente adotada por governos e organizações, em todo o mundo, como uma forma mais avançada e eficiente de compartilhar informações na web.

[11] Disponível em: http://cas.lod-cloud.net/. Acesso em: 1 abr. 2023.

Os dados abertos conectados devem estar disponíveis publicamente de modo interligado e acessíveis por máquinas. Esses dados devem ser fornecidos em um formato padronizado que permite sua fácil reutilização e conexão com outros dados para gerar novos conhecimentos. Os dados abertos conectados são informações que estão interconectadas e disponibilizadas em formatos legíveis por máquina para serem facilmente consumidos e reutilizados. A evolução desses dados levou ao desenvolvimento de tecnologias semânticas que permitem a interconexão e integração de dados heterogêneos (Bizer; Heath; Berners-Lee, 2009).

Ao conectar dados abertos, é possível criar visualizações de dados que permitem aos usuários identificar padrões e tendências que não seriam possíveis de serem identificados com dados isolados. Além disso, a conexão de dados abertos também permite aos usuários criar modelos de previsão e análise que podem contribuir para solucionar problemas cotidianos.

Santarém Segundo (2018) assevera:

> Os últimos anos têm sido bastante significativos em como as tecnologias da Web Semântica e as possibilidades propostas pelas práticas de Dados Ligados tem evoluído e refletido diretamente numa crescente necessidade de se publicar dados. Os dados governamentais de alguns países, disponibilizados em formato aberto e semântico, tem tido impacto perante a sociedade e despertado um conjunto de iniciativas pelo desenvolvimento de aplicações que possam efetivamente levar o cidadão a consumir esses dados para os mais variados propósitos no seu dia a dia (Santarém Segundo, 2018, p. 3).

Historicamente, o conceito de DAC tem sido objeto de estudo desde o início do século XXI, quando o Projeto da Web Semântica foi criado em 2001 (Bizer; Heath; Berners-Lee, 2009). Ao falar sobre Web Semântica, os autores se referem a um conjunto de tecnologias que visam aumentar a interoperabilidade, a padronização, a organização e o reúso de informações disponíveis na web. Além disso, a Web Semântica busca possibilitar inferências e a ocorrência de serendipidade, ou seja, a descoberta de informações relevantes de maneira inesperada (Santarém Segundo, 2018).

A Web Semântica é uma das principais tecnologias que suportam a publicação e o consumo de dados abertos conectados. Esta tecnologia permite a representação de dados em formato de grafos, facilitando a integração de dados heterogêneos. Além disso, a tecnologia semântica permite a criação de modelos de dados abertos conectados que podem ser compartilhados entre diferentes sistemas (Rautenberg; Burda; Souza, 2018).

A adoção de dados abertos conectados tem sido facilitada pelo desenvolvimento de ferramentas de publicação e consumo desses dados. Estas ferramentas permitem a

publicação de dados em formatos padronizados, permitindo a interoperabilidade entre sistemas. Além disso, estas ferramentas permitem a criação de aplicações que podem acessar e consumir dados de diferentes fontes (Bandeira *et al*., 2014).

A disponibilização de dados abertos conectados tem um grande impacto na sociedade. Estes dados podem ser usados para promover a transparência e a prestação de contas por parte das instituições públicas e privadas (Kitchin, 2014b). Eles podem ser usados para criar modelos de serviços, como aplicações inteligentes e ferramentas de visualização de dados. A capacidade de interligar dados em diferentes domínios do conhecimento também pode levar a descobertas significativas e inovadoras em diferentes áreas de pesquisa (Kitchin, 2014b).

Além disso, a conexão de dados abertos pode ajudar a melhorar a qualidade das atividades públicas, pois permite que os usuários acessem informações de diferentes fontes e as relacionem para obter resultados mais precisos. Esta conexão de dados também pode ajudar a melhorar a segurança, pois permite que os usuários acessem informações de diferentes fontes e as relacionem para identificar e prevenir ameaças (Kitchin, 2014b).

Nesse sentido, a disponibilização de dados abertos conectados tem um grande impacto na sociedade. Estes dados podem ser usados para promover a transparência, a prestação de contas, a tomada de decisão, a eficiência dos processos, a segurança, a inovação e a colaboração.

A compreensão do papel dos dados abertos conectados na sociedade é fundamental para o desenvolvimento de novos conhecimentos. O *Linked Open Data* (LOD) é uma iniciativa que visa fornecer acesso aberto a dados interligados e reutilizáveis. Esta iniciativa procura aumentar a transparência e a responsabilidade, bem como promover a inovação e a colaboração entre os usuários (Minichiello, 2017). A abordagem de *Link Open Data* (LOD) segue três princípios fundamentais: abertura, modularidade e escalabilidade. O princípio da abertura implica que os dados devem estar acessíveis e disponíveis para diversas aplicações. A modularidade garante que os dados estejam acessíveis sem a necessidade de planejamento prévio para a ligação com outros dados. A escalabilidade proporciona uma flexibilidade maior no acesso e inclusão de novos dados, uma vez que a entrada desses dados é facilitada quando existem dados em formato *Resource Description Framework* (RDF). A adoção da LOD vem crescendo cada vez mais, especialmente entre as instituições governamentais, que estão incorporando múltiplas informações de interesse público (Minichiello, 2017).

É importante ressaltar que, com a crescente quantidade de dados produzidos, é necessário garantir que os dados estejam estruturados e interoperáveis para obter a máxima utilização, além de estar em conformidade com as políticas e regulamentações de proteção de dados (Kitchin, 2014b).

4.2 Padrões de representação e tecnologias utilizadas em Dados Abertos Conectados

O modelo "5 estrelas" de Tim Berners-Lee é uma estrutura para a publicação de dados abertos conectados na web. Ele foi proposto por Tim Berners-Lee em 2006 e é amplamente utilizado para garantir que os dados sejam publicados de forma que possam ser facilmente acessados, compartilhados e reutilizados. O modelo é composto por cinco níveis, cada um representando um grau crescente de abertura e conectividade dos dados (Berners-Lee, 2009, p. 8).

No primeiro nível, os dados são publicados na web em qualquer formato, mas sem nenhum tipo de estrutura ou metadados. No segundo nível, os dados são publicados em um formato estruturado, como CSV (*Comma-Separated Values*) ou XML (*Extensible Markup Language*), mas ainda sem metadados. No terceiro nível, os dados são publicados em um formato estruturado e com metadados descritivos, como RDF (*Resource Description Framework*). No quarto nível, os dados são publicados usando padrões de dados abertos conectados, como o *Linked Data*, que permitem que os dados sejam conectados a outros dados na web. No quinto e último nível, os dados são publicados usando padrões de dados abertos conectados e também incluem *links* para outros dados relacionados na web (Heath; Bizer, 2011, p. 26; Victorino *et al*., 2017, p. 8)

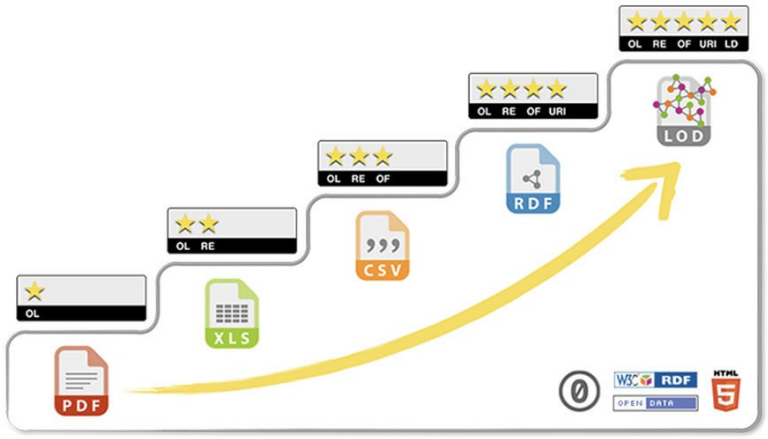
A implementação do modelo 5 estrelas pode trazer muitos benefícios para a publicação de dados abertos conectados, incluindo maior acessibilidade, interoperabilidade e reutilização dos dados. No entanto, a implementação completa do modelo pode ser desafiadora e requer um esforço significativo para garantir que os dados sejam publicados de forma consistente e padronizada (Rautenberg *et al*., 2017).

O processo de transição de Dados Abertos para Dados Abertos Conectados envolve uma série de transformações intermediárias, que aumentam a usabilidade e a interconexão desses conjuntos de dados. Conforme o esquema de classificação de 5 estrelas proposto por Tim Berners-Lee, essas transformações são estruturadas em uma progressão incremental que eleva o grau de abertura e conexão dos dados (Isotani; Bittencourt, 2015).

Inicialmente, os dados tornam-se disponíveis ao público sob uma licença aberta, sem restrições quanto ao formato de publicação. A partir daí, as etapas subsequentes envolvem a estruturação dos dados de maneira legível por máquina, a disponibilização dos dados em um formato aberto e não proprietário, a utilização de identificadores uniformes de recursos para nomear as entidades nos dados, e finalmente, a interligação dos dados com outros conjuntos de dados (Kitchin, 2014b)

Essa sequência de transformações pretende finalizar a criação de um ecossistema de Dados Abertos Conectados, em que os usuários possam navegar e descobrir informações de maneira intuitiva e eficiente (Berners-Lee, 2009). Tal ecossistema maximiza o valor dos dados ao permitir uma contextualização mais ampla e a descoberta de informações relacionadas, conforme ilustra a figura 4.

Figura 4 – Distribuição 5 estrelas para Dados Abertos



Fonte: 5 ★Open Data (2012).

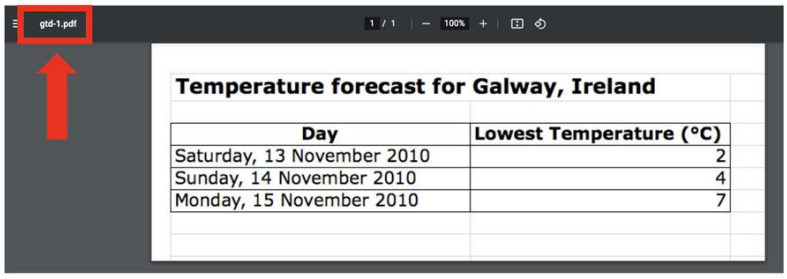
Desta maneira, cabe uma análise detalhada da taxonomia proposta por Tim Berners-Lee para a classificação de dados abertos. Entretanto, Isotani e Bittencourt (2015, p. 45) afirmam que, para alcançar um nível satisfatório de abertura, é recomendado que os conjuntos de dados atinjam, no mínimo, o patamar da terceira estrela na classificação proposta por Tim Berners-Lee. Para exemplificar os modelos de cada fase das 5 estrelas, é apresentado o clássico cenário, disponível em "5 ★ OPEN DATA"[12], de como os dados sobre a temperatura da cidade irlandesa de Galway podem migrar para o padrão aberto.

A primeira estrela é conferida aos dados que são disponibilizados ao público sob uma licença aberta (*Open License - OL*). Isso significa que os dados podem ser livremente acessados, usados, modificados e compartilhados por qualquer pessoa ou utilizados como

12 Disponível em: https://5stardata.info/en/. Acesso em: 9 maio 2023.

dados de entrada para outros sistemas. Geralmente, neste nível, os dados são disponibilizados em formato ".pdf" (*Portable Document Format*), como ilustra a figura 5, a seguir.

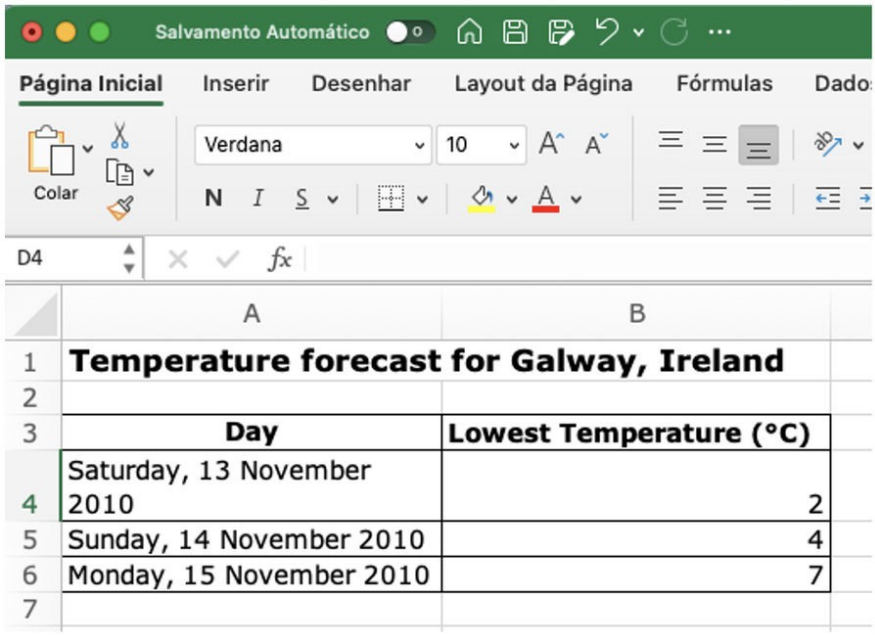
Figura 5 – Dados disponibilizados sob licença aberta



Fonte: https://5stardata.info/en/examples/gtd-1.pdf (2012).

A segunda estrela é atribuída a conjuntos de dados, publicados de forma estruturada e legível por máquina (*Readable Machine – RE*). Esses dados podem ser processados diretamente por *softwares* proprietários, como o Excel (.xlsx) e podem ser convertidos para outros formatos de maneira relativamente fácil. A Figura 6 ilustra esse modelo.

Figura 6 – Dados processados por softwares proprietários

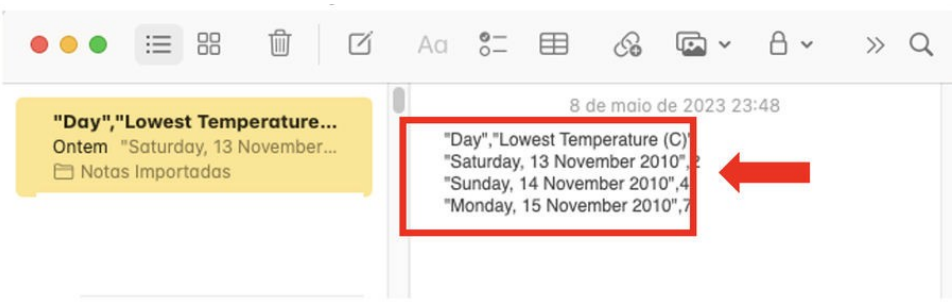


Fonte: https://5stardata.info/en/examples/gtd-2.xls (2012).

A terceira estrela é concedida quando os dados são disponibilizados em um formato aberto não proprietário (*Open Format - OF*), conforme ilustra a Figura 7. Isso permite a manipulação dos dados sem a necessidade de software proprietário, aumentando a

acessibilidade e a interoperabilidade dos dados. Uma das representações de dados mais utilizadas nesta fase são os arquivos com extensão ".CSV" (*Comma-Separated Values*).

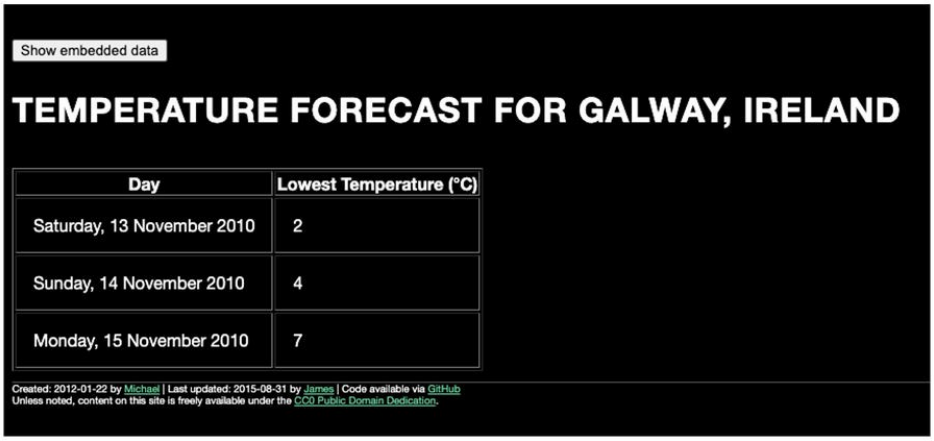
Figura 7 – Dados em formato aberto



Fonte: https://5stardata.info/en/examples/gtd-3.csv (2012).

A quarta estrela é designada quando os dados usam Identificadores Uniforme de Recursos (*Uniform Resource Identifier - URI*) para nomear entidades, conforme demonstra a Figura 8. Isso permite que outros usuários estabeleçam links para os dados, melhorando a usabilidade e a reutilização dos dados.

Figura 8 – Dados utilizando URI (*Uniform Resource Identifier*)



Fonte: https://5stardata.info/en/examples/gtd-4/ (2012).

A quinta e última estrela é concedida quando os dados são efetivamente conectados (*Linked Data - LD*) a outros conjuntos de dados. Isso permite a navegação entre diferentes conjuntos de dados e a descoberta de informações relacionadas, fornecendo um contexto mais amplo e enriquecendo o valor dos dados, conforme a Figura 9, a seguir.

Figura 9 – Dados conectados permitindo a descoberta de informações relacionadas

Fonte: https://5stardata.info/en/examples/gtd-5/ (2012).

O *Resource Description Framework* (RDF) é uma família de especificações da World Wide Web Consortium (W3C) originalmente projetada como um modelo de dados para metadados. No entanto, foi adotada como um formato geral de descrição de informações para representar informações na Web de maneira estruturada e semântica. No núcleo do RDF está a ideia de fazer declarações sobre recursos (no formato de sujeito-predicado-objeto) que podem ser expressas em notações de tripla. Essas declarações são conhecidas como triplas RDF. A força motriz do RDF reside na sua representação tripartida de informações, ou seja, na forma de triplas de sujeito-predicado-objeto. Essas triplas são flexíveis e podem capturar uma ampla gama de relações entre entidades, formando a base para a representação de conhecimento na Web (Heath; Bizer, 2011).

O RDF identifica os recursos (sujeito e objeto das triplas) e suas relações (predicado) usando identificadores uniformes de recursos (URIs). Esta capacidade de usar URIs para identificar tanto os objetos quanto as relações oferece um poderoso meio de representar estruturas de dados complexas, facilitando o entendimento e a interpretação entre diferentes sistemas e aplicações (Klyne; Carroll, 2006).

Complementando o RDF, a linguagem de consulta SPARQL (SPARQL Protocol and RDF Query Language) fornece uma maneira eficiente de acessar e manipular dados armazenados em RDF. A flexibilidade do SPARQL permite a execução de consultas em várias fontes de dados, independentemente do local de armazenamento dos dados ou do sistema de armazenamento subjacente. Esta característica é particularmente útil no

contexto de dados abertos conectados, onde os dados podem estar distribuídos em vários locais e formatos (Prud'hommeaux; Seaborne, 2008).

O SPARQL fornece um conjunto diversificado de funcionalidades de consulta, incluindo a capacidade de combinar dados de várias fontes, consultar padrões de dados e filtrar e processar os resultados da consulta. Além disso, oferece suporte para consultas agregadas e subconsultas, tornando-o uma ferramenta poderosa para análise de dados complexos. O SPARQL também define um protocolo para consultar e atualizar dados RDF através do protocolo HTTP, permitindo o acesso e a manipulação remota de dados RDF (Prud'hommeaux; Seaborne, 2008).

Além do RDF e do SPARQL, as ontologias desempenham um papel crucial na estruturação e interpretação de dados na Web. As ontologias são especificações formais de um vocabulário compartilhado, definindo tipos, propriedades e relações entre entidades. Elas são usadas para dar sentido aos dados, proporcionando um contexto e um significado mais ricos para a interpretação de dados. As ontologias podem ser expressas em várias linguagens, como a Web Ontology Language (OWL), sendo uma extensão do RDF projetada para representar relações complexas entre entidades (Gruber, 1993).

A aplicação de ontologias é essencial para a semântica dos dados, pois facilitam a interpretação e a inferência dos dados, permitindo que máquinas e sistemas interpretem melhor os dados. Por exemplo, uma ontologia pode definir que "Todos os humanos são mamíferos", permitindo que um sistema faça a inferência de que, se uma entidade é humana, ela também é um mamífero (Gruber, 1993).

O conceito de Linked Data, ou Dados Conectados, é uma abordagem para a publicação de dados cujo objetivo é facilitar a interligação de dados e a descoberta de informações. O Linked Data se baseia na ideia de usar URIs HTTP para identificar recursos de dados, proporcionando acesso a uma representação dos dados do recurso quando acessados por agentes de *software*, como navegadores ou *bots*[13] da Web (Bizer; Heath; Berners-Lee, 2009).

A abordagem do Linked Data para conectar dados permite a reutilização e a integração de dados de uma maneira descentralizada e escalável. Ao expressar os dados em RDF e usar URIs HTTP para identificar os recursos, os dados podem ser facilmente

[13] Software aplicativo programado para executar determinadas tarefas. Bots são automatizados, ou seja, atuam por conta própria, sem que um usuário humano tenha que iniciá-los manualmente todas as vezes (Bizer; Heath; Berners-Lee, 2009).

interligados e novas “ligações” podem ser adicionadas para conectar conjuntos de dados relacionados (Heath; Bizer, 2011).

O Linked Data é um dos principais facilitadores da Web Semântica, um conceito que visa tornar as informações na Web mais compreensíveis e úteis para os computadores. Ao conectar dados de diferentes fontes e contextos, o Linked Data aumenta o valor dos dados, permitindo a descoberta de novos insights e a criação de novos serviços baseados em dados (Berners-Lee; Hendler; Lassila, 2001).

Essas tecnologias não funcionam de maneira isolada, mas em conjunto, formando uma infraestrutura para dados abertos conectados. O RDF fornece a base para representar informações de forma estruturada e semântica, o SPARQL permite a consulta e manipulação desses dados, as ontologias dão significado aos dados e o Linked Data permite a interconexão de dados de diferentes fontes (Bizer; Heath; Berners-Lee, 2009).

Apesar do potencial dessas tecnologias, elas também apresentam desafios. Entre eles estão a necessidade de sistemas de gerenciamento de dados robustos para lidar com grandes volumes de dados, a necessidade de vocabulários padronizados para garantir a interoperabilidade e questões de privacidade e segurança que surgem com a publicação e ligação de dados (Hogan *et al*., 2010).

De modo geral, RDF, SPARQL, Ontologias e Linked Data são tecnologias fundamentais para a implementação de dados abertos conectados. Elas proporcionam os meios para representar, consultar, dar significado e conectar dados, facilitando a descoberta e a reutilização de informações, promovendo assim a visão da Web Semântica (Shadbolt; Berners-Lee; Hall, 2006).

4.3 Publicação em Dados Abertos Conectados

Uma nova metodologia que permite a conexão e a interoperabilidade de conjuntos de dados na web é a publicação de Dados Abertos Conectados (DAC). Esta abordagem desenvolveu-se a partir da ideia da Web Semântica e conta com padrões abertos como RDF, SPARQL e URIs para facilitar a integração de dados de várias fontes (Bizer; Heath; Berners-Lee, 2009).

4.3.1 Fontes de Dados Abertos Conectados

As fontes de DAC podem variar amplamente, desde bancos de dados governamentais até repositórios de pesquisa acadêmica e coleções de dados de empresas.

A estrutura aberta e conectada do DAC permite que essas várias fontes sejam combinadas e consultadas de maneira integrada, apresentando novas oportunidades para análise e descoberta de conhecimento (Heath; Bizer, 2011).

Fontes de dados governamentais, por exemplo, podem fornecer detalhes sobre população, geografia, clima, saúde, educação, segurança e muito mais. Repositórios de pesquisa acadêmica podem conter dados de estudos em uma variedade de campos, incluindo ciências naturais, ciências sociais e humanidades. *Clusters* de dados de negócios podem conter detalhes sobre vendas, marketing, operações e outros aspectos do negócio (Berners-Lee; Hendler; Lassila, 2001).

As fontes de DAC são diversas e abrangem múltiplos setores e disciplinas. Uma das principais fontes de DAC é a DBpedia (Auer *et al*., 2007), uma coleção de dados estruturados extraídos da Wikipedia. A DBpedia[14] permite que os usuários consultem e explorem a vasta quantidade de informações na Wikipedia usando tecnologias como SPARQL e RDF.

Outro exemplo significativo é o portal de dados do governo dos EUA, Data.gov[15] (Shadbolt; Berners-Lee; Hall, 2006), que disponibiliza uma infinidade de conjuntos de dados governamentais em um formato aberto e acessível. Esses dados abrangem uma ampla gama de tópicos, desde saúde e educação até clima e energia.

O DrugBank[16] é uma fonte valiosa de DAC na área da biomedicina. Ele fornece informações detalhadas sobre medicamentos, incluindo sua química, farmacologia, mecanismos de ação, interações e história clínica (Wishart *et al.,* 2018, p. 2).

Na esfera das ciências humanas, a Biblioteca Digital de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT)[17] oferece acesso a uma rica coleção de pesquisas acadêmicas de instituições de todo o Brasil.

O portal Europeana[18] é outra fonte importante de DAC, oferecendo acesso a milhões de itens digitalizados de bibliotecas, arquivos e museus em toda a Europa, incluindo livros, fotos, obras de arte e muito mais.

[14] Disponível em: https://www.dbpedia.org/. Acesso em: 9 maio 2023.

[15] Disponível em: https://data.gov/. Acesso em: 8 maio 2023

[16] Disponível em: https://go.drugbank.com/. Acesso em: 8 maio 2023

[17] Disponível em: https://bdtd.ibict.br/vufind/. Acesso em: 8 maio 2023

[18] Disponível em: https://www.europeana.eu/pt. Acesso em: 8 maio 2023

4.3.2 Processos de Publicação de Dados Abertos Conectados

O processo de publicação de DAC envolve várias etapas, como preparação de dados, conversão para o formato RDF, atribuição de URIs HTTP às entidades nos dados, conexão com outros conjuntos de dados e disponibilização dos dados *online*.

Para garantir que os dados sejam consistentes e livres de erros, a preparação de dados envolve a limpeza e normalização dos dados. A conversão para o formato RDF permite que os dados sejam estruturados de uma forma que facilita a integração com outros conjuntos de dados. A adição de URIs HTTP às entidades nos dados permite sua referência e acesso na web. A capacidade de se conectar a outros conjuntos de dados permite que os dados sejam contextualizados e enriquecidos com mais informações.

4.3.3 Ferramentas e Plataformas de Publicação em DAC

Na busca por uma gestão do conhecimento eficaz, o uso de dados abertos conectados tem se mostrado uma estratégia promissora. Como esses dados são disponibilizados em formato aberto, torna-se possível estabelecer conexões entre diferentes conjuntos de dados, aumentando a riqueza e a profundidade do conhecimento extraído desses recursos (Isotani; Bittencourt, 2015). No entanto, para que essa estratégia seja eficaz, é crucial a utilização de ferramentas e plataformas adequadas para a publicação desses dados.

A publicação de DAC é um processo complexo e de várias etapas que pode ser auxiliado pelo uso de várias ferramentas e plataformas. A natureza aberta e conectada do DAC cria oportunidades para a integração e a análise de dados, possibilitando a geração de ideias de projetos a partir de uma variedade de fontes de dados. No entanto, a qualidade do DAC publicado depende fortemente da qualidade dos dados de entrada e dos processos de preparação e conversão de dados (Bizer; Heath; Berners-Lee, 2009).

A plataforma Virtuoso[19] é uma opção que suporta a publicação de DAC, fornecendo recursos para o armazenamento, consulta e gerenciamento de dados RDF. A plataforma D2RQ[20] possibilita o mapeamento de bancos de dados relacionais para RDF, facilitando a conversão de dados já existentes para o formato RDF. A ferramenta Silk[21] é usada para

[19] Disponível em: https://virtuoso.openlinksw.com/. Acesso em: 15 maio 2023.

[20] Disponível em: http://d2rq.org/. Acesso em: 15 maio 2023.

[21] Disponível em: http://silkframework.org/. Acesso em: 15 maio 2023.

conectar conjuntos de dados, permitindo a identificação e a criação de links entre entidades semelhantes em muitos conjuntos de dados (Bizer; Cyganiak, 2006).

Ao publicar DAC, as organizações podem tornar seus dados mais acessíveis e úteis para uma variedade de outros usuários e aplicações, além de suas próprias operações internas. Isso pode levar a mais abertura, cooperação e inovação, o que beneficiaria tanto a organização que publica os dados quanto uma comunidade maior de usuários de dados (Heath; Bizer, 2011).

No entanto, a publicação de DAC também levanta uma série de questões, como aquelas relacionadas à privacidade e segurança dos dados, qualidade e confiabilidade dos dados e escalabilidade e desempenho dos sistemas DAC. Estes são desafios significativos que devem ser superados para garantir que o DAC possa ser usado de maneira eficaz e responsável (Berners-Lee; Hendler; Lassila, 2001).

As tendências futuras provavelmente mostrarão uma maior adoção e utilização de DAC à medida que mais organizações adotem sua publicação e mais ferramentas e plataformas se tornem disponíveis. Isso pode resultar em uma maior compreensão e apreciação do potencial do DAC, bem como novas oportunidades para análise e exploração de dados (Auer *et al*., 2007).

4.4 Conexão de Dados Abertos

A conexão de dados abertos tem se mostrado uma abordagem promissora para avançar a pesquisa e o compartilhamento de informações (Bizer *et al*.,2008). Esta ideia visa conectar coleções de dados abertos de fontes heterogêneas, permitindo uma compreensão mais abrangente e holística do conhecimento, atualmente disponível. Nesse contexto, pesquisadores de dados abertos conectados desempenham um papel crucial, explorando o potencial dessa interconexão para estimular avanços científicos.

Conforme a definição de dados abertos, esta informação pode ser livremente usada, reaproveitada e redistribuída, desde que a fonte original seja devidamente creditada e compartilhada segundo as mesmas regras (Berners-Lee, 2009). A troca de dados abertos tenta aproveitar essa liberdade ao conectar vários conjuntos de dados para gerar sinergias e permitir a análise e a fusão de dados de muitas fontes. Essa conectividade expande as possibilidades de novas possibilidades tecnológicas, incentivando a colaboração e a inovação na pesquisa.

É fundamental incluir a definição “Aberta” no contexto de conexões de dados abertos, que define dados abertos como qualquer dado que possa ser livremente usado, reaproveitado e redistribuído por qualquer pessoa. Esta abertura de dados remove barreiras e restrições ao acesso à informação, fomentando a transparência e a democratização do conhecimento (James, 2013).

Compreender o potencial desta abordagem para a pesquisa, explora sinergias entre vários conjuntos de dados, ampliando a compreensão e as possibilidades de descoberta. Além disso, ao aderir às diretrizes para dados abertos, contribui-se com o avanço da ciência, incentivando a colaboração e a reutilização de informações valiosas.

4.4.1 Mapeamento de dados abertos

O mapeamento de dados abertos refere-se ao processo de representação e conexão semântica de conjuntos de dados disponíveis publicamente. Essa atividade envolve a descoberta de relações entre entidades, o estabelecimento de *links* entre conjuntos de dados e a criação de ontologias para facilitar a interoperabilidade e a integração de dados (Heath; Bizer, 2011). Por meio do mapeamento, é possível criar uma rede de dados abertos, proporcionando uma visão abrangente e holística de informações dispersas.

A representação e a conexão semântica de conjuntos de dados em um mapeamento de dados abertos permitem a descoberta de relações e a compreensão da estrutura subjacente dos dados[22]. Essa abordagem não se limita apenas à simples vinculação de informações, mas busca estabelecer um contexto significativo entre os diferentes conjuntos de dados, proporcionando um entendimento mais profundo e completo (Heath; Bizer, 2011).

O mapeamento de dados permite a identificação de correlações e padrões ocultos, revelando assim conhecimentos valiosos. A conexão de dados sobre o clima com dados sobre segurança pública, por exemplo, pode analisar como as condições climáticas afetam os indicadores de criminalidade em diferentes regiões.

A seguir, estão alguns resultados da pesquisa que explicam a análise de correlação e como ela pode ser usada para entender as relações entre diferentes variáveis:

[22] [...] Tipo de padrões ou tendências que podem ser inferidos a partir dos dados. Padrões sazonais ou tendências de crescimento a longo prazo nos dados [...] (Heath; Bizer, 2011).

- A análise de correlação pode revelar complexos relacionamentos que existem frequentemente entre as variáveis em dados multivariados (Zhiyuan Zhang *et al*., 2015, tradução nossa).
- Quando observa-se mapas de duas variáveis relacionadas, muitas vezes notamos que eles se parecem: as áreas com altos níveis em um mapa tendem a ser as mesmas áreas com altos níveis no outro mapa. Isso é conhecido como correlação espacial (Minn, 2023, tradução nossa).
- A análise de correlação envolve a análise da relação espacial entre vários atributos ou temas. Em outras palavras, a análise de correlação tenta medir o quanto duas variáveis estão relacionadas entre si (Manson; Matson, 2017, tradução nossa)
- O mapa de correlação solar é um novo tipo de visualização que pode representar de forma bela e sucinta as matrizes para explorar as correlações (Zapf; Kraushaar, 2017, tradução nossa).

A análise de correlação é um método poderoso para descobrir como diversas variáveis em dados multivariados se relacionam entre si. Conectar dados de várias fontes, como estatísticas meteorológicas e de crime, torna possível examinar as interações entre muitas variáveis e obter predições importantes sobre sistemas complexos. Essa visão integrada dos dados contribui para a geração de conhecimento e o suporte à tomada de decisões.

A criação de ontologias é uma parte essencial do processo de mapeamento de dados abertos. As ontologias fornecem uma estrutura conceitual que define os termos, as relações e as propriedades dos dados em um domínio específico. Elas ajudam a padronizar a representação dos dados, permitindo que diferentes conjuntos de dados sejam interoperáveis e facilmente combinados. Além disso, as ontologias possibilitam a reutilização e o compartilhamento de conhecimento, estabelecendo uma base comum para a comunidade científica e os profissionais da informação (Isotani; Bittencourt, 2015).

Existem várias técnicas para o mapeamento de dados abertos, uma delas é a correspondência de esquemas (Cruz, 2015), que consiste em identificar elementos similares entre diferentes conjuntos de dados e mapeá-los para estabelecer relações semânticas. Outra técnica comum é o alinhamento ontológico (Basso *et al.,* 2017), que busca encontrar correspondências entre conceitos e propriedades em ontologias diferentes.

Além disso, o uso de algoritmos de aprendizado de máquina e processamento de linguagem natural também desempenham um papel importante no mapeamento de dados abertos.

O mapeamento de dados abertos tem uma estreita relação com a Ciência da Informação. Através do uso de técnicas e métodos de indexação, classificação e recuperação de informações, a Ciência da Informação contribui para a estruturação e organização dos dados abertos, tornando-os mais acessíveis e úteis para os usuários finais. Além disso, a Ciência da Informação fornece fundamentos teóricos e conceituais para a criação de ontologias e modelos de representação de conhecimento, essenciais no processo de mapeamento de dados abertos (Gayathri; Uma, 2018).

O mapeamento de dados abertos oferece diversos benefícios, como a promoção da transparência e da colaboração, o estímulo à inovação e a criação de novas oportunidades de pesquisa. A heterogeneidade dos dados, a falta de padronização e a necessidade de atualização contínua das ontologias e dos mapeamentos apresentam-se como desafios significativos. Superar esses desafios requer a aplicação de abordagens multidisciplinares que integrem conhecimentos da Ciência da Informação, da Ciência de Dados e de outras áreas relacionadas.

Dessa maneira, o mapeamento de dados abertos desempenha um papel crucial na exploração e na maximização do potencial dos dados abertos conectados. Ao estabelecer *links*, criar ontologias e promover a interoperabilidade, o mapeamento proporciona uma visão integrada dos dados, permitindo que pesquisadores e profissionais da informação extraiam conhecimentos significativos e promovam avanços nas áreas da Ciência da Informação e da Ciência de Dados. Com o contínuo desenvolvimento dessas disciplinas, espera-se que novas abordagens e técnicas aprimorem ainda mais o mapeamento de dados abertos, impulsionando a descoberta de informações relevantes.

4.4.2 Alinhamento e reconciliação de dados abertos

O alinhamento de dados abertos refere-se ao processo de identificar correspondências e relacionamentos semânticos entre diferentes conjuntos de dados, principalmente entre ontologias, de modo a reduzir redundâncias e ambiguidades (Gracia; Mena, 2012, p. 1). Por outro lado, a reconciliação de dados abertos envolve a resolução de heterogeneidade nos dados, visando a harmonização e a integração deles (Guimarães; Andrade; Baptista, 2022).

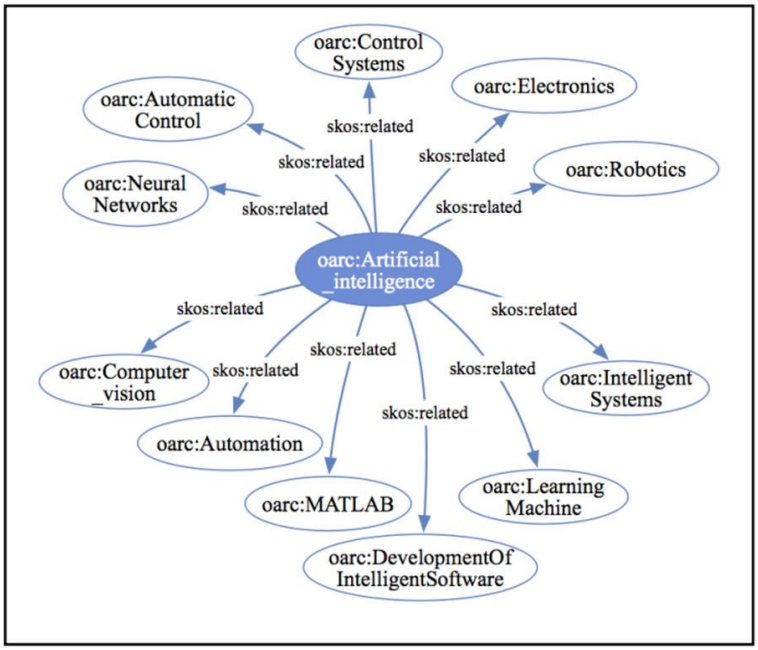
Importante destacar que o alinhamento de dados abertos pode ser realizado em diferentes níveis de granularidade, desde o alinhamento de instâncias individuais até o

alinhamento de esquemas e ontologias completas. No caso do alinhamento de instâncias, o objetivo é encontrar correspondências entre entidades específicas presentes nos conjuntos de dados, permitindo sua interconexão e enriquecimento mútuo. Já o alinhamento de esquemas e ontologias busca estabelecer correspondências semânticas entre os termos e conceitos utilizados em diferentes modelos de representação, facilitando a integração de dados em um nível mais abstrato (Gracia; Mena, 2012). O trabalho de Piedra *et al*. (2016, p. 7) apresenta modelos de enriquecimento e reconciliação de dados abertos conectados por meio de conexões RDF com recursos externos.

> Para melhorar a descoberta de recursos, foi necessário criar links RDF com recursos externos. Estes foram publicados na nuvem para dados vinculados abertos e foram conectados com repositórios externos. Os tópicos e palavras-chave foram vinculados a cada recurso digital, que consistia em conteúdo de vocabulário controlado e esquemas de classificação. Mais especificamente, links externos foram estabelecidos com a nomenclatura da UNESCO e conjuntos de dados da DBpedia inglês/espanhol/latino-americano (Piedra *et al*., 2016, p. 6, tradução nossa).

A Figura 10 ilustra o alinhamento dos DAC por meio das conexões RDF com recursos da Unesco e da DBPedia, ou seja, dados externos.

Figura 10 – Alinhamento de dados via conexões RDF



Fonte: Piedra et al., (2016, p. 7).

O prefixo 'oarc' identificado no esquema apresentado refere-se a uma convenção de nomenclatura dentro de uma ontologia específica ilustrada na figura 10, utilizada para categorizar e relacionar conceitos dentro do campo da inteligência artificial. Essa ontologia estrutura o conhecimento em categorias como 'Sistemas de Controle', 'Redes Neurais' e

'Robótica', dentre outras, destacando as interconexões e a natureza interdisciplinar da inteligência artificial. A utilização de tal prefixo permite a organização sistemática e a recuperação eficaz de informações, elementos cruciais para a pesquisa e desenvolvimento na área.

Gracia e Mena (2012, p. 61) propõem medidas que calculam o grau de similaridade e relação entre diferentes descrições semânticas. Os autores consideram técnicas como: correspondência de ontologia, agrupamento e desambiguação de sentidos, o que corrobora para o alinhamento semântico de dados abertos.

> A abordagem proposta baseia-se num estudo de medidas semânticas que calculam numericamente o grau de semelhança e de relação entre diferentes descrições semânticas. Para ultrapassar a redundância e a ambiguidade na Web Semântica, desenvolvemos um conjunto de técnicas baseadas nestas medidas: correspondência de ontologias, agrupamento de sentidos e desambiguação de sentidos (Gracia; Mena, 2012, p. 61, tradução nossa).

Diversas abordagens têm sido propostas para o alinhamento e a reconciliação de dados abertos. Uma abordagem comum é o uso de técnicas baseadas em correspondência de esquemas, que buscam identificar elementos comuns e estabelecer relações semânticas entre os dados. Além disso, técnicas de reconciliação baseadas em aprendizado de máquina e processamento de linguagem natural têm se mostrado eficazes na resolução de conflitos e discrepâncias nos dados. A aplicação dessas abordagens depende da disponibilidade de ontologias, vocabulários controlados e algoritmos sofisticados para realizar a tarefa de alinhamento e reconciliação (Basso *et al*., 2019).

A reconciliação de dados abertos, envolve a identificação e resolução de conflitos e discrepâncias entre os dados provenientes de diferentes fontes. Isso pode incluir a reconciliação de valores inconsistentes, como divergências de formatação ou unidades de medida, bem como a reconciliação de terminologias e vocabulários discrepantes. A reconciliação é um processo complexo que requer técnicas avançadas, como algoritmos de aprendizado de máquina e abordagens baseadas em regras, a fim de garantir a coerência e a consistência dos dados reconciliados (Guimarães; Andrade; Baptista, 2022).

A relação entre o alinhamento, a reconciliação de dados abertos e a Ciência da Informação é intrínseca. A Ciência da Informação oferece uma base sólida de conhecimento teórico e metodológico para lidar com a organização, a representação e a recuperação de informações contidas nos dados abertos. Através da aplicação de princípios e técnicas de indexação, classificação e recuperação de informações, a Ciência da

Informação contribui para a resolução de conflitos e discrepâncias, bem como para a integração e harmonização dos dados abertos (Rautenberg *et al*., 2017).

A Ciência da Informação fornece uma perspectiva crítica sobre questões éticas e legais relacionadas ao alinhamento e à reconciliação de dados abertos. Isso inclui a proteção da privacidade e a conformidade com regulamentos de proteção de dados, além da consideração de questões de propriedade intelectual, direitos autorais e interoperabilidade semântica, associados aos dados abertos (Rautenberg *et al.,* 2017).

A interoperabilidade e a integração semântica são conceitos fundamentais no que se refere aos dados abertos e à forma como estes podem ser compreendidos por humanos e *softwares*. Eles envolvem o acesso aos dados, agregação, correlação e transformação, de maneira que os dados possam ser processados de forma significativa e consistente (Muntean *et al*., 2010).

Contudo, a realização da interoperabilidade semântica e da integração semântica não é uma tarefa fácil. Para isso, se fazem necessárias técnicas de alinhamento e reconciliação de dados abertos. Essas técnicas são vitais para a qualidade, integração e interoperabilidade dos dados, pois possibilitam uma visão integrada e abrangente deles, melhorando a compreensão e utilização desses recursos (Muntean *et al*., 2010).

Uma pesquisa mais aprofundada na área da Ciência da Informação é crucial nesta situação, pois fornece os fundamentos teóricos, quadros metodológicos e ferramentas necessárias para abordar desafios relacionados ao alinhamento e reconciliação dos dados abertos. São necessárias pesquisas contínuas e a colaboração entre disciplinas para avançar neste campo, visando desenvolver abordagens inovadoras e soluções eficazes para melhorar a qualidade e integração dos dados abertos. Portanto, é possível afirmar que existe uma relação direta entre a interoperabilidade semântica, a integração semântica e o alinhamento e reconciliação dos dados abertos. Para melhorar a compreensão e o uso dos dados abertos, esses conceitos e técnicas devem ser devidamente implementados (Basso et al., 2019; Guimarães; Andrade; Baptista, 2022).

4.4.3 *Linking Open Data* (LOD)

A "linkagem de dados abertos" ou *Linking Open Data* (LOD) refere-se ao processo de conexão de conjuntos de dados criando relações semânticas entre eles (Berners-Lee, 2009). Essa abordagem envolve a descoberta e correspondência de entidades comparáveis em vários *clusters* de dados, possibilitando uma visão integrada e abrangente das informações contidas nessas fontes. O LOD é essencial para uma integração e

interoperabilidade (Souza; Alvarenga, 2004, p. 139) eficazes, facilitando a extração de conhecimento (Bizer; Heath; Berners-Lee, 2009).

Existem várias perspectivas para LOD. Uma delas é a correspondência de esquemas, que envolve encontrar semelhanças e diferenças entre os atributos e estruturas relacionais dos conjuntos de dados. Além disso, o uso de vocabulário controlado, ontologias e identificadores permanentes desempenham um papel significativo na ligação de dados abertos (Meij *et al*., 2011).

O conceito de LOD, intrinsicamente se conecta aos conceitos da Ciência da Informação. Esse contexto fornece os fundamentos teóricos e conceituais necessários para organizar, representar e recuperar as informações contidas nos dados abertos (Souza; Alvarenga, 2004). Ao utilizar técnicas para indexação, categorização e recuperação de informações, a Ciência da Informação contribui para a eficácia da ligação de dados abertos, possibilitando a descoberta de conexões e relações semânticas entre conjuntos de dados, bem como se compromete com os aspectos éticos, legais e de qualidade dos dados abertos, garantindo sua confiabilidade e reutilização (Victorino *et al.,* 2017; Rautenberg *et al*., 2017).

A ligação de dados abertos traz vários benefícios, como a promoção da interoperabilidade, descoberta de conhecimento oculto e obtenção de uma visão integrada dos dados. Esta abordagem facilita a pesquisa interdisciplinar, a colaboração institucional e o desenvolvimento de aplicações e serviços inovadores (Souza; Alvarenga, 2004).

Mas há desafios que a ligação de dados abertos também enfrenta, como a confiabilidade dos dados, identificação de conjunto de dados e suas distribuições, padronização de formato de dados e interoperabilidade (Jesus, 2021, p. 102). É essencial desenvolver estratégias e técnicas que superem esses desafios. Utilizar ontologias, vocabulários restritos e esquemas de metadados pode ajudar a padronizar e harmonizar os dados abertos, facilitando ligá-los. Além disso, o uso de correspondência e reconciliação de entidades, bem como técnicas de aprendizagem de máquina, pode ajudar a melhorar a precisão e a eficiência da ligação de dados abertos (Souza; Alvarenga, 2004).

Para avançar no campo do LOD, é crucial a colaboração interdisciplinar, neste sentido, a Ciência da Informação (CI) desempenha um papel importante, fornecendo os fundamentos teóricos, metodológicos e tecnológicos necessários para a organização e representação dos dados abertos. Algumas áreas dentro da CI contribuem com seus conceitos em indexação, classificação e recuperação de informações, essenciais para a eficácia do LOD (Souza; Alvarenga, 2004).

A conexão entre dados abertos e Ciência da Informação é uma área de estudo em constante desenvolvimento, assim, é possível superar os desafios relacionados à integração, interoperabilidade e qualidade de dados abertos através da aplicação de técnicas avançadas e da colaboração interdisciplinar. O uso eficaz da ligação de dados abertos promove colaboração científica, descoberta de conhecimento e tomada de decisões, impulsionando o avanço da Ciência da Informação.

4.5 Uso e aplicação de dados abertos conectados

A utilização de dados abertos e conectados permite que pesquisadores extraiam valor de dados que, em outra situação, poderiam ser deixados inexplorados. Dependendo das necessidades específicas do usuário, os dados abertos podem ser usados de várias maneiras. Eles podem ser usados para analisar tendências, detectar padrões, fazer previsões, desenvolver novos modelos e algoritmos e muito mais. Atemezing *et al*. (2013, p. 285) revelou que os dados abertos conectados foram efetivamente usados na análise de tendências climáticas do serviço público de meteorologia da Espanha[23]. Além disso, a disponibilidade de dados abertos conectados incentiva a colaboração aberta e transparente entre pesquisadores, profissionais da indústria e até mesmo cidadãos comuns. Isso torna possível que novas abordagens e ideias sejam rapidamente aprimoradas e compartilhadas, fomentando a inovação e a descoberta de conhecimento.

Esta ideia corresponde com os princípios do movimento da Ciência Aberta, que busca tornar o processo científico mais acessível, democrático e eficaz (Woelfle; Olliaro; Todd, 2011).

As aplicações para dados abertos e conectados são numerosas e diversas. Eles podem ser usados em vários campos, incluindo medicina, ciências sociais, engenharia, segurança pública, negócios e governo, para citar alguns. Isso destaca a importância do papel que os pesquisadores desempenham na utilização destes recursos, simultaneamente, em que sublinha a necessidade de mais pesquisas e treinamento nesta área emergente. O valor das conexões de dados abertos reside na sua capacidade de estabelecer conexões significativas entre vários conjuntos de dados. Na prática, essa interconectividade permite uma análise mais profunda e abrangente dos dados, o que poderia fornecer conhecimentos que, de outra forma, seriam difíceis de obter (Shadbolt; Berners-Lee; Hall, 2006).

23 Disponível em http://www.aemet.es/es/portada. Acesso em: 29 maio 2023.

Um pesquisador que estuda saúde pública pode comparar dados demográficos e de doenças para encontrar padrões e tendências que possam afetar a política de saúde. Além disso, as conexões de dados abertos têm o potencial de avançar significativamente a equidade de acesso à informação. O fato de que os dados estejam livremente disponíveis para o público reduz as barreiras de acesso à informação, permitindo que mais atores participem da pesquisa e tomada de decisões. Isso é especialmente importante no contexto de países em desenvolvimento, onde o acesso a informações de alta qualidade pode ser limitado (Gurstein, 2011).

Por fim, apesar dos evidentes benefícios dos dados abertos e conectados, existem desafios que devem ser superados para aproveitar plenamente este recurso. Devido às possíveis implicações éticas e legais de tornar alguns tipos de dados publicamente disponíveis, preocupações de privacidade e segurança são de importância particular (Zuiderwijk; Janssen, 2014).

Além disso, a qualidade e a compatibilidade dos dados podem variar, exigindo o desenvolvimento de padrões e protocolos para garantir a utilidade e a confiabilidade dos dados. A relevância dos dados abertos conectados está na capacidade de criar conexões significativas (semânticas) entre diferentes conjuntos de dados (Cruz, 2015).

Além disso, os dados abertos conectados também têm um forte potencial para promover a equidade no acesso à informação. Como os dados são gratuitamente disponíveis para o público, isso reduz as barreiras ao acesso à informação, permitindo que uma gama mais ampla de atores possa contribuir para a pesquisa e a tomada de decisões. Isso é particularmente relevante no contexto dos países em desenvolvimento, onde o acesso a informações de alta qualidade pode ser restrito (Gurstein, 2011).

Embora os benefícios dos dados abertos conectados sejam claros, também existem desafios que precisam ser abordados para aproveitar ao máximo esse recurso. Questões de privacidade e segurança são de particular importância, uma vez que a disponibilização pública de certos tipos de dados pode ter implicações éticas e legais (Zuiderwijk; Janssen, 2014). Além disso, a qualidade e a compatibilidade dos dados podem variar, requerendo o desenvolvimento de normas e protocolos para garantir a utilidade e a confiabilidade dos dados.

4.5.1 Análise de dados abertos conectados

A análise de DAC oferece oportunidades de avanço em várias áreas do conhecimento e, para assimilar todo esse contexto, se faz necessária uma compreensão

sofisticada de várias técnicas e ferramentas de análise de dados, bem como o conhecimento relevante necessário para avaliar e usar corretamente os resultados dessas análises (Halevy; Norving; Pereira, 2009).

A preparação inicial e pré-processamento dos dados são componentes cruciais na análise de dados abertos e conectados. Mesmo sendo valiosos, os dados abertos frequentemente possuem formatos inconsistentes e podem ter erros como entradas duplicadas ou dados ausentes. Portanto, é crucial a limpeza, organização e formatação dos dados de maneira apropriada para análise (Rahm; Hong Hai Do, 2000).

Após o procedimento de pré-processamento, é necessária uma análise eficaz de dados abertos conectados para uma seleção adequada de técnica. As técnicas de análise de dados que são mais frequentemente aplicadas a conjuntos de dados abertos e conectados são: análise de regressão, classificação, agrupamento e mineração de texto, cada uma das quais tem vantagens e desvantagens dependendo do tipo de questão de pesquisa que está sendo abordada (Côrtes; Porcaro; Lifschitz, 2002).

Além disso, depois que os dados abertos forem "ligados", a análise de rede pode ser muito útil para identificar e examinar relacionamentos entre vários *clusters* de dados. Pode-se visualizar e quantificar essas relações usando análise de rede, o que pode fornecer conhecimento sobre a estrutura e a dinâmica dos sistemas complexos que esses dados refletem (Côrtes; Porcaro; Lifschitz, 2002).

Outra técnica de análise que está se tornando cada vez mais aplicável a conjuntos de DAC é o aprendizado de máquina, que pode resultar no desenvolvimento de predições, identificar padrões e tendências e desvendar conhecimento escondido nos dados, treinando modelos de aprendizado de máquina nos dados abertos (Goodfellow; Bengio; Courville, 2016).

No entanto, apesar do potencial dessas técnicas, é crucial lembrar que a análise de DAC pode criar uma série de desafios. Uma vez que a análise e divulgação de alguns tipos de dados podem ter implicações significativas para o direito de privacidade de uma pessoa, questões de privacidade e éticas são particularmente proeminentes entre elas (Ohm, 2010). Outro desafio é o requisito de infraestrutura de dados e capacidade analítica apropriadas. Pode ser necessária uma grande capacidade de computação e uma coleção de habilidades altamente especializadas para a análise de DAC, e nem todos os pesquisadores ou organizações podem ter acesso a esses recursos (Ohm, 2010).

Além das técnicas de análise já mencionadas, é fundamental reconhecer a importância da visualização de dados, uma vez que fornece uma representação gráfica dos dados que pode ajudar os pesquisadores a entender padrões, tendências e relacionamentos, que não seriam imediatamente aparentes por meio de análise estatística independente (Keim *et al*., 2008).

Neste sentido, as ferramentas de visualização de dados fornecem um poderoso complemento às técnicas de análise, permitindo uma compreensão mais intuitiva dos dados. Estas análises têm significativas implicações para a reprodutibilidade do estudo. A disponibilidade dos dados e a publicação dos métodos de análise utilizados para gerar os resultados podem aumentar a transparência e permitir que outros pesquisadores confirmem, repliquem ou desafiem as conclusões (Peng, 2011).

É importante notar que a análise de DAC é uma área interdisciplinar. Isso significa que requer a cooperação de especialistas de várias disciplinas, incluindo, mas não se limitando à Ciência da Informação, Ciência da Computação, Estatística, Sociologia, Economia, e muito mais. Esta interdisciplinaridade da análise de DAC é um dos fatores que contribui para o seu potencial de produzir descobertas significativas e inovadoras em uma ampla gama de campos (Manyika *et al*., 2011).

Além disso, os DAC oferecem oportunidades para análises preditivas significativas. Os pesquisadores podem utilizá-los para prever comportamentos ou eventos futuros com base em dados históricos usando algoritmos de aprendizado de máquina. Isso pode ser muito útil em várias áreas, incluindo criminologia, previsão de tendências de mercado, previsão de doenças ou desastres naturais, previsão de resultados eleitorais e análises em tempo real (Manyika *et al*., 2011).

O desenvolvimento de computação em nuvem e tecnologias de *streaming* permitiram aos pesquisadores lidar e analisar grandes quantidades de dados em tempo real, permitindo-lhes agir e tomar decisões quase instantaneamente. Embora possam ser uma ferramenta poderosa para a geração de conhecimento, também é fundamental considerar o contexto e as limitações destes dados, visto que, podem apresentar lacunas, imprecisões ou erros que podem limitar sua aplicabilidade ou resultar em interpretações incorretas (Boyd; Crawford, 2012).

A análise de DAC é uma área que está em constante evolução. Considerando que novas técnicas e tecnologias estão sendo constantemente desenvolvidas, é fundamental que os pesquisadores estejam atualizados com esses desenvolvimentos. Isso não só permitirá

que eles aproveitem ao máximo essa tecnologia, mas também ajudará a garantir que sua análise seja conduzida de forma ética e responsável (Boyd; Crawford, 2012).

4.5.2 Visualização de dados abertos conectados

O uso eficiente de tecnologias para gerenciamento e modificação de dados é essencial para o avanço da transparência pública e do estímulo ao maior envolvimento democrático. No entanto, a lacuna entre a disponibilidade desses dados e seu consumo, de modo efetivo, continua sendo bastante grande. De acordo com (Gurstein, 2011), isso se deve à falta de acesso ao software ou tecnologia adequados, bem como à escassez de recursos financeiros e credenciais educacionais necessárias para utilizar esses dados de forma eficaz. A interoperabilidade e a interpretabilidade dos dados, ou mais especificamente, a capacidade dos usuários de perceber, identificar e interpretar os dados de maneira eficiente e precisa, devem ser aprimoradas por meio da adoção de várias iniciativas, uma dessas medidas é o uso de visualizações de dados (Gurstein, 2011).

A visualização de dados abertos é uma técnica que está se tornando cada vez mais significativa no campo da análise de dados conectados, ela pode tornar os padrões e tendências subjacentes aos dados mais acessíveis e compreensíveis para um amplo conjunto de usuários, transformando dados brutos em formas visuais intuitivas (Gurstein, 2011).

Dependendo do tipo e da complexidade dos dados, várias formas de visualização de dados abertamente conectados podem ser usadas. Por exemplo, gráficos de barras e gráficos de linhas podem ser úteis para representar tendências temporalmente ou comparações entre vários grupos, ou categorias. Por outro lado, gráficos de dispersão, diagramas de caixa e histogramas podem ser úteis para exibir a distribuição de variáveis contínuas. As visualizações da rede podem ser usadas para ilustrar as relações e interações entre várias entidades ou nós para dados abertos e conectados com uma arquitetura de rede (Macedo *et al*., 2020).

Utilizando tecnologias interativas, pode-se melhorar a visualização de dados abertamente conectados. Ferramentas de visualização interativa permitem que os usuários explorem os dados a partir de várias perspectivas, ajustem parâmetros, filtrem resultados e aprendam coisas que podem não ser óbvias a partir de visualizações estáticas (Macedo *et al*., 2020). A interatividade também pode facilitar para os usuários entenderem os dados, permitindo que eles “brinquem” com eles e aprendam por meio de exploração e experimentação.

A comunicação e disseminação de informações são grandes implicações da visualização de dados abertos e conectados. Uma visualização de dados bem projetada pode transmitir informações complexas de maneira clara e concisa, tornando-se uma ferramenta valiosa para apresentar os resultados de pesquisas, comunicar-se com o público ou tomar decisões baseadas em evidências. Também pode ajudar a aumentar a conscientização e o interesse do público em questões importantes, incentivar a competência em informação de dados e incentivar o envolvimento cívico e a participação, tornando os dados mais acessíveis e interessantes (Santos *et al.,* 2021).

No entanto, a visualização de DAC também vem com um conjunto de desafios. Um desses desafios é a necessidade de equilibrar a complexidade e a simplicidade da representação de dados. Embora uma visualização de dados simples possa ser mais fácil de entender para o público, se não bem planejada, corre o risco de simplificar ou distorcer os dados. Por outro lado, o público pode se sentir intimidado ou confuso com uma visualização de dados que é excessivamente sofisticada (Macedo *et al*., 2020).

Outro desafio é garantir a precisão e a ética na visualização de dados. Como as visualizações de dados têm o potencial de afetar percepções e decisões, é essencial garantir que elas retratem correta e justamente os dados. Isso inclui evitar a manipulação de escala, cores ou outros fatores que possam levar a erros ou enganar o público. Além disso, é fundamental respeitar a privacidade e confidencialidade dos dados, especialmente quando eles contêm informações sensíveis ou pessoais (Hullman; Diakopoulos, 2011).

A visualização de DAC é um campo que está se desenvolvendo rapidamente, com novas técnicas e ferramentas sendo criadas a todo momento. À medida que as tecnologias de visualização de dados avançam, é essencial que os pesquisadores "fiquem por dentro" das tendências e inovações mais recentes.

A visualização de dados abertos conectados envolve a utilização de ferramentas capazes de lidar com a complexidade e a interconexão desses dados. Aqui estão alguns exemplos de ferramentas que são particularmente adequadas para este propósito:

- **Gephi[24]:** é uma ferramenta de visualização e exploração de redes ideal para trabalhar com dados abertos conectados. Ela permite aos usuários visualizar e analisar redes de dados, descobrir padrões e tendências, e explorar as relações e interações entre diferentes entidades ou nodos (Antoniazzi, 2020, p. 27).

[24] Disponível em: https://gephi.org/. Acesso em: 16 maio 2023.

- **Linked Data Platform (LDP):** é uma plataforma que suporta a publicação e a interconexão de dados na Web. Ela permite a criação de visualizações interativas de dados abertos conectados que podem ser exploradas e manipuladas pelos usuários.
- **Cytoscape[25]:** é uma plataforma de software para visualização de redes complexas e análise de dados de rede. É amplamente utilizada na bioinformática para a visualização de redes moleculares e integração de dados de redes com perfis de atributos gerais (Antoniazzi, 2020).
- **Virtuoso[26]:** é um servidor de banco de dados de alto desempenho que pode serusado para armazenar e consultar dados abertos conectados. Ele suporta uma variedade de padrões e protocolos de dados abertos e fornece funcionalidades avançadas para a manipulação e visualização de dados abertos conectados.
- **Datawrapper[27]:** é uma ferramenta *online* de visualização de dados que permite a criação de gráficos interativos e mapas a partir de dados abertos. Embora não seja especificamente projetada para dados abertos conectados, sua capacidade de trabalhar com uma ampla variedade de formatos de dados e sua interface intuitiva a tornam uma ferramenta útil para a visualização de dados abertos.

4.5.3 Recomendação de dados abertos conectados

A recomendação de DAC oferece uma oportunidade significativa para melhorar a acessibilidade, usabilidade e eficácia dos dados na era digital. Este método envolve a busca por coleções de dados pertinentes às necessidades e interesses específicos de um usuário e à apresentação desses dados de forma simples de entender e usar (Sheth; Henson; Sahoo, 2008). A recomendação de DAC pode ser útil em ambientes de pesquisa, onde a capacidade de acessar e usar dados relevantes pode ter influência no sucesso de um projeto. Um pesquisador que estuda as mudanças climáticas pode se beneficiar de conjuntos de dados que incluem informações sobre temperaturas históricas, emissões de gases de efeito estufa ou padrões de precipitação. Da mesma forma, um pesquisador que estuda a economia pode se beneficiar da recomendação de conjuntos de dados que incluem detalhes como o PIB, a taxa de desemprego ou o comércio internacional (Berners-Lee; Hendler; Lassila, 2001).

[25] Disponível em: https://cytoscape.org/. Acesso em: 16 maio 2023.

[26] Disponível em: https://virtuoso.openlinksw.com/. Acesso em: 16 maio 2023.

[27] Disponível em: https://www.datawrapper.de/. Acesso em: 16 maio 2023.

A recomendação de dados abertos e conectados também pode ser usada para facilitar a tomada de decisões baseadas em evidências em muitos contextos. Por exemplo, os políticos podem se beneficiar com recomendações para coleções de dados contendo informações sobre saúde pública, educação ou bem-estar social. Da mesma forma, os gerentes de negócios podem se beneficiar da recomendação de conjuntos de dados que contêm informações sobre tendências do mercado, comportamento do consumidor ou desempenho financeiro (Halevy; Norgvig; Pereira, 2009).

A recomendação de dados abertos e conectados pode ser simplificada com o uso de técnicas de aprendizado de máquina e inteligência artificial para aprendizado. Para dar um exemplo, algoritmos de aprendizado de máquina podem ser usados para analisar os padrões de uso de dados de um usuário e prever quais conjuntos de dados podem ser interessantes para o usuário no futuro. Da mesma forma, os sistemas de inteligência artificial podem ser usados para compreender o conteúdo e o contexto dos dados e recomendar conjuntos de dados pertinentes às perguntas ou questões específicas que um usuário está tentando responder (Bizer; Heath; Berners-Lee, 2009).

Ao usar metadados padronizados e semânticos, a recomendação de dados abertos e conectados também pode ser melhorada. Os metadados podem incluir detalhes cruciais sobre os dados, como sua fonte, sua qualidade ou seu formato. Esses detalhes podem ajudar os usuários a avaliar a relevância e a utilidade dos dados e tomar decisões informadas sobre o seu uso. Da mesma forma, dados semânticos podem melhorar a capacidade de os sistemas de recomendação de reconhecer e fornecer dados que sejam pertinentes às necessidades específicas de um usuário, facilitando a interoperabilidade e a integração dos dados (Bizer; Heath; Berners-Lee, 2009).

A recomendação de dados conectados e abertos também está acompanhada de um número de desafios, dentre eles, a abundância de dados disponíveis é um dos principais. Isso pode tornar desafiador para os sistemas de recomendação reconhecerem e transmitirem efetivamente os conjuntos de dados mais relevantes. Além disso, a precisão e a utilidade das recomendações podem ser afetadas por variações significativas na qualidade e confiabilidade dos dados (Zaveri *et al*., 2012).

Outro desafio é a segurança e a privacidade dos dados. Embora a recomendação de dados abertos e conectados possa melhorar a acessibilidade e uso dos dados, também pode levantar preocupações quanto à proteção de informações pessoais e sensíveis. Isso é especialmente verdadeiro em situações em que os dados podem incluir informações sensíveis de saúde, financeiras ou outras. Portanto, é essencial garantir que a recomendação

de dados abertos e conectados seja realizada de maneira que respeite a privacidade e a segurança dos dados (Hogan *et al*., 2010).

Apesar destes desafios, a recomendação de dados abertos e conectados tem um significativo potencial para melhorar como os dados são acessados e usados. A recomendação de bancos de dados abertos e conectados pode facilitar a pesquisa, a tomada de decisões baseadas em evidências e uma ampla gama de outras atividades que dependem do acesso a dados precisos e atualizados, facilitando para os usuários encontrar e usar os dados relevantes para as suas necessidades específicas (Bizer; Heath; Berners-Lee, 2009). A recomendação de dados abertos e conectados oferece uma oportunidade significativa para melhorar a acessibilidade, usabilidade e eficácia dos dados na era digital. No entanto, também apresenta um número de desafios que devem ser considerados para garantir que a recomendação seja realizada de forma justa, eficaz e respeite a privacidade e segurança dos dados.

4.5.4 Outras aplicações de dados abertos conectados

Uma ampla variedade de aplicações para dados abertos conectados existe além de pesquisas e tomadas de decisão baseadas em evidências. Esses são usados em uma variedade de campos, desde Ciência da Informação até cuidados de saúde, passando pelo governo e educação (Heath; Bizer, 2011).

O campo da saúde é onde os dados abertos e conectados têm aplicações vitais. Eles podem ser usados para rastrear e analisar a propagação da doença, avaliar a eficácia de várias intervenções médicas e iniciativas de saúde pública e fornecer informações úteis a médicos, pacientes e pesquisadores. A iniciativa Linked Open Drug Data[28] (LODD) combina dados de várias fontes para fornecer informações detalhadas sobre drogas medicamentosas, seus efeitos e seu uso (Saeed; Chaki; Janev, 2019, p. 164)

Dados abertos conectados podem contribuir para o aumento e a responsabilidade na transparência do governo, incluindo a disseminação de informações sobre políticas, orçamentos e resultados do governo, bem como promover a democracia na participação cívica e a prestação de relatórios financeiros. O projeto europeu *Linked Government Data*[29] (LGD) trabalha para tornar os dados do governo acessíveis, compreensíveis e utilizáveis,

[28] Disponível em: https://www.w3.org/wiki/HCLSIG/LODD. Acesso em: 2 jun. 2023.

[29] Disponível em: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/linked-government-data-lgd. Acesso em: 2 jun. 2023.

permitindo que a administração pública, empresas e cidadãos compartilhem e reutilizem soluções e projetos em toda a Europa (Geiger; Von Lucke, 2012, p. 266).

O uso de dados abertos e conectados na educação pode melhorar o ensino e o aprendizado. Isso pode envolver a personalização da educação para atender às necessidades e interesses únicos dos alunos, facilitar o acesso a recursos educacionais e incentivar a colaboração e o envolvimento dos alunos. Por exemplo, o projeto *LinkedUp* incentiva o uso de dados abertos e conectados para apoiar a educação aberta e o aprendizado (Rozsa; Dutra; Nhacuongue, 2017).

Dados abertos e conectados têm aplicações significativas também na Ciência da Informação. Eles podem ser utilizados para melhorar a análise de dados, a organização do conhecimento e a recuperação de informações. Isso inclui o desenvolvimento de mecanismos de busca semânticos, sistemas de recomendação e outras ferramentas que podem ajudar os usuários a localizar, acessar e usar informações de forma mais eficaz (Shadbolt; Berners-Lee; Hall, 2006).

Neste contexto, a pesquisa de dados de acesso aberto e a Ciência da Informação desempenham um papel-chave. O potencial de dados abertos e conectados pode ser desbloqueado e seus inúmeros benefícios para a sociedade podem ser realizados pelos pesquisadores à medida que se aprimora o entendimento destes dados e se desenvolve novas técnicas e tecnologias para trabalhar com eles.

Existem muitas aplicações diferentes e extensas de dados abertos e conectados com uma significativa influência em muitas esferas sociais. No entanto, ainda existem muitos desafios a serem superados para maximizar o potencial destes dados. Estes incluem garantir a precisão e a confiabilidade dos dados, lidar com questões de privacidade e segurança e desenvolver ferramentas e infraestrutura para facilitar o acesso e o uso dos dados de forma eficaz (Kitchin, 2014b).

Podem ser usados para apoiar inovação e desenvolvimento econômico. Isso inclui o uso de dados para desenvolver novos produtos e serviços, melhorar a eficiência e a produtividade e incentivar a colaboração e a troca de conhecimentos. O projeto "Open Data 500"[30] estuda como os dados abertos são usados por empresas e desenvolve ferramentas para ajudá-las a maximizar o potencial dos dados abertos (Kitchin, 2014b).

A cooperação entre outras disciplinas e setores será crucial para o avanço do uso de DAC. Nesse sentido, pode-se desenvolver abordagens mais abrangentes e eficazes para

[30] Disponível em: https://thegovlab.org/project/project-open-data-500-global-network. Acesso em: 2 jun. 2023.

trabalhar com eles, combinando conhecimentos de campos como Ciência da Informação, Ciência da Computação, Estatística, Direito e Ética. No final, os DAC têm o potencial de mudar o acesso e o uso das informações, com implicações significativas para a ciência, a sociedade e a economia. À medida que se investiga e se aprimora o uso desses dados, criam-se oportunidades para avançar o conhecimento e melhorar a vida das pessoas em todos os lugares.

4.5.5 Aplicações de Dados Abertos Conectados em Segurança Pública

O crime não ocorre de maneira linear, ele é aleatório. Em mapas estatísticos não há uma distribuição uniforme dos locais onde ele acontece e essa informação serve para que as pessoas evitem certas regiões em determinadas horas do dia ou da noite (Kedia, 2016). Os órgãos de segurança pública[31] necessitam de tecnologias para obter vantagens no que se refere à prevenção do crime. Entretanto, na maioria dos países, esse fato não é uma realidade, visto que não estão capacitados para o uso, armazenamento, manutenção e recuperação da informação de modo eficaz, resultando em modelos de policiamento por intuição, empirismo e, na prevenção criminal, utilizam um método simples de tentativa e erro (Kedia, 2016, p. 2).

Na segurança pública, os DAC começam a apresentar uma influência significativa. Eles podem fornecer informações e predições valiosas para melhorar a prevenção de crimes, a resposta a emergências e a tomada de decisões estratégicas devido à sua capacidade de integrar e analisar grandes quantidades de informações de muitas fontes (Chainey; Ratcliffe, 2005).

A prevenção e detecção de crimes é uma área onde os dados abertos conectados estão sendo usados extensivamente. Por exemplo, o projeto *Data-Driven Approaches to Crime and Traffic Safety* (DDACTS), nos Estados Unidos, usa dados de crimes e trânsito para identificar áreas de alto risco e executar estratégias policiais baseadas em dados. Igualmente, a *Metropolitan Police* de Londres usa dados abertos conectados para analisar padrões de criminalidade e orientar suas operações de patrulha (Chainey; Ratcliffe, 2005).

Outrossim, DAC podem ser usados para melhorar os tempos de resposta a emergências. Autoridades podem obter uma compreensão mais precisa da situação, facilitar a coordenação entre várias agências e responder de forma mais eficaz às crises, reunindo e analisando dados em tempo real de uma variedade de fontes (Seltzer;

[31] O referido estudo ocorre na cidade de Faridabad, norte da Índia, entretanto, há uma proximidade significativa com o modelo de policiamento brasileiro (Kedia, 2016).

Mahmoudi, 2013). Contextualizando, é exemplificado o modelo “Change by US” que ocorre na Islândia, onde se utiliza a internet para permitir aos seus cidadãos, via redes sociais, um *feedback* sobre sua constituição.

> Alguns casos utilizaram simplesmente a Internet para chegar a uma ampla população, à semelhança da utilização das redes sociais ou de um instrumento de investigação baseado na Internet. Estes casos incluem: Islândia permitindo feedback sobre a sua nova constituição através do Facebook e Twitter; Change By Us solicitando sugestões para melhorias na cidade, mas sem um processo de seleção de soluções (Seltzer; Mahmoudi, 2013, p. 9, tradução nossa).

Orientar decisões estratégicas sobre segurança pública, onde as autoridades podem identificar tendências, avaliar a eficácia de várias políticas e programas, bem como tomar decisões sobre onde focar seus recursos, analisando dado a longo prazo sobre crimes, acidentes de trânsito e outros problemas, é um pressuposto viável para o uso dos DAC.

Ao tornar os dados sobre crimes, acidentes de trânsito e outras atividades de segurança pública disponíveis ao público, as autoridades podem melhorar a confiança e a cooperação do público, permitindo maior fiscalização e responsabilidade.

O conceito de “Dados Abertos Conectados” (DAC), bem como suas tecnologias, processos de publicação, métodos de interconexão e aplicações, foram abordados nesta seção. Foi possível entender a importância do DAC na promoção da transparência, eficiência e colaboração entre vários profissionais, como funcionários públicos, comunidades locais, empresas, desenvolvedores profissionais, bem como o terceiro setor.

Os padrões de representação e tecnologias utilizadas no DAC foram analisados, destacando a importância da teoria e pesquisa na identificação e resolução de problemas com foco nas fontes de DAC, processos de publicação, ferramentas e plataformas.

5 DADOS ABERTOS CONECTADOS NA SEGURANÇA PÚBLICA

A preservação da ordem pública configura-se como fator prioritário para o ser humano, pois está, intrinsicamente, relacionada ao direito inalienável à vida e a outras prerrogativas fundamentais como, saúde, educação, patrimônio, dentre outros. Esta constitui uma das bases essenciais de uma sociedade estável e próspera. Na sociedade, ocorrem constantes atos antissociais como, furtos, roubos, contravenções e demais atos ilícitos e, por conta do aumento da violência, há demasiada preocupação junto às autoridades governamentais (Costa, 2010, p. 130).

As mesmas autoridades têm por obrigação zelar pela proteção dos cidadãos contra crimes e pela manutenção da ordem pública, impactando diretamente a qualidade de vida, o crescimento econômico e a confiança nas instituições públicas.

Os problemas na área da segurança pública se relacionam a diversos fatores, sendo eles: desigualdade social, racial, política, econômica, desemprego e modelos defasados no sistema educacional e da saúde. Neste sentido, evidencia-se que toda a problemática que permeia a segurança pública não se relaciona apenas com a criminalidade, mas, também com os fatores acima descritos. Desta forma, o uso de informações para subsidiar atividades de inteligência de segurança pública, torna-se fundamental, especialmente quando se busca maximizar a eficiência e a eficácia das ações de prevenção criminal (Costa, 2010).

O uso do conhecimento produzido pelos serviços de inteligência pode apoiar os profissionais de segurança nas melhores tomadas de decisões, e, consecutivamente, elaborar políticas públicas para a redução dos índices de criminalidade, de letalidade e de ações arbitrárias por parte dos agentes de segurança pública (Silva; Rolim, 2017).

Um aspecto intrigante e frequentemente subestimado da segurança pública é o conhecimento tácito que os profissionais de segurança, especialmente os policiais, têm sobre os padrões de crimes. Este conhecimento, atestado por profissionais de segurança, manifesta-se como uma ferramenta de magnitude relevante para a salvaguarda da ordem pública. Tal conhecimento oferece várias vantagens, que incluem:

- Demarcação de Zonas de Vulnerabilidade: Com base em avaliações empíricas, especialistas em segurança podem discernir regiões com maior predisposição à

manifestação de delitos específicos. Esta identificação permite uma otimização na alocação de recursos, bem como a instauração de protocolos de prevenção.

- Criminalística Analítica: A ciência da criminalística enfoca na interpretação de dados pertinentes à criminalidade, visando fornecer informações relevantes às instituições encarregadas da ordem pública (Rocha, 2020). Especialistas de segurança conseguem enriquecer esta análise ao prover conhecimentos baseados em sua experiência, facilitando a identificação de padrões e tendências delitivas
- Direcionamento Estratégico: Administradores públicos encarregados da segurança devem adotar a criminalística analítica como instrumento-chave na tomada de decisões estratégicas, táticas e administrativas. Ao incorporar a perspectiva empírica de profissionais do setor, torna-se possível deliberar com maior acurácia e eficiência.
- Por fim, um exemplo notável é a observação da relação entre o clima e a incidência de crimes, foco desta pesquisa. Por gerações, os policiais têm notado como certas condições climáticas, como o calor excessivo ou períodos de chuva intensa, parecem correlacionar-se com mudanças na frequência e nos tipos de crimes cometidos.

De modo geral, a experiência prática acumulada por especialistas em segurança, especialmente no que tange à identificação de padrões criminológicos, desempenha um papel crucial no fortalecimento e refinamento das estratégias de segurança pública (Pinto, 2019).

Essas observações, no entanto, muitas vezes permanecem no domínio do conhecimento tácito, sem ser sistematicamente estudadas ou aproveitadas em sua totalidade. A questão, então, é como traduzir essa sabedoria empírica em estratégias práticas e eficazes de policiamento? Como direcionar o policiamento com base nas condições climáticas de uma forma que seja tanto estrategicamente sólida quanto sensível às complexidades sociais e ambientais da comunidade?

Alguns estudos relevantes podem contribuir com respostas para essas questões. O estudo de Brüderle, Peters e Roberts (2017) avaliou o efeito do clima sobre a criminalidade na África do Sul. Os autores utilizaram estatísticas de curto e médio prazo com informações climáticas, mais especificamente a precipitação pluviométrica e a temperatura, e a relação desses índices com crimes violentos e crimes contra a propriedade. No referido estudo, são apresentadas algumas teorias que conceitualizam o pensamento sobre o crime no contexto

das ciências sociais, sendo elas, a teoria econômica do crime (Becker, 1968), a teoria da tensão, introduzida por Merton (1938) e a teoria da desorganização social, proposta por Shaw e McKay (1942).

A teoria da tensão e a teoria da desorganização social argumentam que algumas características sociais como desigualdade, pobreza e heterogeneidade racial resultam em uma maior propensão das pessoas a cometer crimes. A teoria econômica do crime sugere que, na África do Sul, os indivíduos avaliam os benefícios e riscos de cometer um delito. Eles ponderam a utilidade de se envolver em atividades criminosas contra a possibilidade de serem detidos e os custos de oportunidade associados. É essencial considerar como as consequências de curto e médio prazo influenciam essa decisão (Bruederle; Peters; Roberts, 2017).

As temperaturas e as precipitações pluviométricas podem selecionar esses componentes de três maneiras, duas das quais operam no curto prazo e, uma, no médio prazo: A primeira é que as altas temperaturas aumentam imediatamente a agressão entre as pessoas. Psicólogos observaram que as variações de temperatura podem ter um impacto direto no comportamento humano. Muitos acreditam que o desconforto causado pelo calor excessivo pode tornar as pessoas mais irritadas. Esta irritação pode conduzir a pensamentos e atitudes mais agressivos, aumentando potencialmente as chances de comportamentos violentos ou criminosos (Anderson, 2001, p. 3).

Em segundo lugar, no curto prazo, o clima pode influenciar diretamente as condições que tornam o crime mais ou menos provável. Determinadas condições climáticas podem aumentar ou diminuir as chances de criminosos serem pegos, afetando assim o risco percebido e o custo associado à prática de um delito. Jacob, Lefgren e Moretti (2007) asseveram:

> Quando chove forte ou as temperaturas são extremas, é comum ver menos pessoas nas ruas ou até mesmo uma diminuição nas rondas policiais. Por outro lado, um dia de tempo agradável, com muita gente circulando na área pública, pode paradoxalmente criar um cenário mais propício para ocorrências criminosas (Jacob; Lefgren; Moretti, 2007, p. 17).

O terceiro fator se relaciona ao setor agrícola. Um estudo realizado na Índia corrobora com essa afirmação. É evidente que o setor agrícola é vital para a produção de alimentos e é fortemente afetado por condições climáticas. Em períodos de seca, a produção e as oportunidades de trabalho podem cair drasticamente. Diante desse cenário, algumas pessoas podem se sentir compelidas a recorrer ao crime como uma forma de

compensar a perda de renda ou porque veem um menor risco associado a tais ações quando as oportunidades legais são escassas (Blakeslee; Fishman, 2015).

No Brasil, existem pesquisas que analisam a relação entre o clima e a ocorrência de crimes, especialmente no meio rural. Nessas áreas, os delitos tendem a aumentar quando as condições climáticas afetam negativamente a produção agrícola. Ishak (2022) investigou a influência das variações climáticas sobre a criminalidade, observando correlações tanto no aspecto temporal, relacionado ao clima, quanto no espacial, focando nas taxas de criminalidade. O estudo destaca que, durante as secas, com a consequente queda na produção agrícola, muitos trabalhadores rurais são demitidos. Diante dessa situação, alguns podem recorrer ao crime como meio de sobrevivência. Já em períodos chuvosos, quando a produção agrícola prospera e a economia local se aquece, a tendência é oposta.

O estudo ainda aponta que as flutuações no rendimento, sobretudo entre trabalhadores rurais e não qualificados, parecem ser o maior indicador das taxas de criminalidade, solidificando a ideia de que o que realmente impacta são as variações de renda e não necessariamente as condições socioeconômicas gerais da região. Fatores como orçamento municipal, desemprego, pobreza, desigualdade e aspectos psicológicos, curiosamente, não se mostraram tão determinantes para explicar a criminalidade violenta.

O objetivo deste estudo é apresentar a sazonalidade como fator que influencia determinadas ocorrências de crimes em áreas urbanas. Por meio da análise de dados públicos de segurança e registros climáticos, pretendeu-se validar o que muitos profissionais da segurança identificam na prática. Evidencia-se o crime de violência doméstica, visto que há um aumento significativo desses casos durante períodos chuvosos, uma vez que o agressor tende a ficar mais tempo em casa. Por outro lado, crimes que ocorrem externamente aos domicílios, como furtos, parecem diminuir quando as chuvas são intensas, provavelmente por esses delitos serem mais comuns em tempos mais secos.

5.1 Levantamento de fontes de dados relevantes

A identificação de fontes de dados pertinentes é um processo que exige uma meticulosa investigação *online* para discernir as bases de dados que serão empregadas e avaliar sua persistência, garantindo que os dados estejam consistentemente acessíveis. Nesta perspectiva, uma alternativa promissora é a incorporação de Dados Abertos Governamentais (DAG). Estes são caracterizados como “a provisão, por meio da Internet, de informações e dados governamentais de domínio público destinados à livre utilização pelo conjunto da sociedade” (Agune; Gregório Filho; Bolliger, 2010).

A implementação dos DAG é, de fato, uma estratégia para promover a transparência, incentivar a participação social e estimular a inovação na administração pública. Incontroversamente, as entidades governamentais incentivam uma investigação mais aprofundada e o envolvimento dos cidadãos, enquanto viabilizam o aparecimento de soluções tecnológicas que podem melhorar os serviços públicos e estimular oportunidades econômicas, tornando a informação governamental relacionada com o domínio público acessível e passível de reutilização (Gil-García *et al*., 2012).

Além disso, os DAG estabelecem-se como uma infraestrutura estável e acessível a investigadores, acadêmicos e especialistas de várias áreas, desde as ciências sociais às tecnologias da informação, tornando possível analisar tendências, configurações e relações em grandes quantidades de dados. Esta metodologia orientada para os dados tem um potencial inexplorado para clarificar percepções fundamentais, que podem influenciar a formulação de políticas públicas mais precisas e adequadas. Para garantir que as inferências retiradas são verdadeiramente substanciais, torna-se essencial assegurar a exatidão, a qualidade, a atualidade e a integridade dos dados (Araújo, 2017).

Para que um dado governamental seja classificado como DAG, ele deve aderir aos oito princípios para Dados Abertos Governamentais propostos por Tim O'Reilly[32] (Araújo, 2017, p. 28).

- **Completude**: não sujeitos a limitações de privacidade, segurança ou privilégios válidos;
- **Primariedade**: tal como recolhidos da fonte, com o mais alto nível de granularidade, não agregados e não modificados;
- **Oportunidade**: disponibilizados tão rapidamente quanto necessário, de modo a preservar o valor dos dados;
- **Acessibilidade**: disponíveis para a mais ampla gama de usuários e para diversos propósitos;
- **Processável por máquina**: estruturados de forma adequada para permitir o processamento automatizado;
- **Antidiscriminatoriedade**: disponíveis para qualquer pessoa, sem necessidade de identificação;

[32] Disponível em: https://public.resource.org/8_principles.html. Acesso em: 3 jun. 2023.

- **Não proprietário**: disponíveis em formato sobre o qual nenhuma entidade tenha controle exclusivo;
- **Livre de licença**: não estando sujeitos a qualquer direito autoral, patente, marca registrada ou regulação comercial. Quando adequados, podem ser permitidas restrições de privacidade, segurança e privilégios sobre os dados. Dados sobre os quais não se aplica nenhuma restrição, devem ser marcados de forma clara como sendo de domínio público.

A definição de um conjunto de dados governamentais como "aberto" engloba um conjunto de critérios rigorosos que visam não só a divulgação da informação, mas também a sua utilidade, acessibilidade e integridade. Em primeiro lugar, o critério de "Completude" sublinha que a informação deve ser abrangente, respeitando, no entanto, restrições legítimas de privacidade e segurança. A noção de "Primariedade" enfatiza a importância da essência inalterada dos dados, indicando que estes devem ser compartilhados de maneira minuciosa e autêntica, sem modificações ou adições que possam comprometer seu significado original (Araújo, 2017).

O princípio da "Oportunidade" enfatiza a relevância da temporalidade dos dados. Para que a informação seja relevante e tenha impacto, a sua divulgação deve ser em tempo adequado. A interligação entre os conceitos de "acessibilidade" e "antidiscriminação" é fundamental para assegurar a disponibilidade de dados a um amplo espetro de acesso dos usuários, sem quaisquer distinções ou obstáculos. Isso é reforçado pela exigência de que os dados sejam "Processáveis por máquina", ou seja, apresentados de maneira que facilite a sua manipulação e análise por sistemas computacionais (Araújo, 2017).

Os conceitos de "Não-proprietário" e "Sem licença" asseguram que os dados abertos permanecem livres de constrangimentos proprietários ou restrições legais que possam limitar a sua utilização e reutilização. No entanto, é crucial notar que, apesar do objetivo de manter os dados tão abertos quanto possível, há situações em que as restrições, como as relacionadas com a privacidade ou a segurança, são não só apropriadas como também essenciais. Por conseguinte, enquanto se dá prioridade à abertura dos dados, é imperativo que esta divulgação seja equilibrada com a necessidade de proteger informações sensíveis e direitos individuais (Araújo, 2017).

Na prática, muitas entidades encarregadas de disponibilizar os DAG não aderem a todos os princípios estabelecidos. No entanto, é essencial que busquem cumprir a maioria deles (Matheus; Ribeiro; Vaz, 2012). Analisando os oito princípios previamente mencionados, nota- se que um dado, mesmo estando em conformidade com o princípio de

ser isento de licença, pode, em determinadas circunstâncias, não atender completamente ao primeiro princípio, especialmente quando são necessárias restrições de segurança e privacidade.

Neste sentido, surgem alguns desafios que vão, desde aspectos técnicos, como a normalização e a interoperabilidade dos dados, até dilemas éticos e questões de privacidade relacionadas com a divulgação de informações potencialmente sensíveis.

Dentre esses desafios, se destaca a relação semântica entre dados ambientais, como precipitações e vazão pluviométricas e fenômenos sociológicos, como taxas de ocorrências criminais. Um dos conjuntos de dados utilizados neste trabalho, está disponível no *site* dados.gov.br[33] onde será integrado com dados de segurança pública para identificar e prever tendências criminais associadas a variáveis pluviométricas (Brasil, 2012).

5.1.1 Extração dos Dados Pluviométricos e de Vazão

No que se refere à extração de dados pluviométricos, após a identificação de variáveis relevantes, como quantidade de precipitação e taxas de vazão, os dados foram extraídos e submetidos a procedimentos de sanitização, com uso da metodologia ETL[34] (Extração, Transformação e Carga), para assegurar sua integridade. Com a utilização de ferramentas de ETL fica garantido que os dados extraídos são não apenas relevantes, mas também isentos de inconsistências.

Esta fase envolve a identificação e remoção de possíveis *outliers* ou dados que não estejam completos, garantindo que os conjuntos de dados que passarão para as fases subsequentes sejam de alta qualidade e confiabilidade.

Continuando com o processo meticuloso de tratamento dos dados, é importante ressaltar que a curadoria dos dados não termina apenas com a remoção de *outliers* e a verificação de consistência. Dados pluviométricos, por sua natureza, são suscetíveis a variabilidades sazonais e regionais. Assim, uma etapa de normalização foi implementada, assegurando que todas as medidas estavam em escalas compatíveis e que as variações estivessem adequadamente contabilizadas. Esta etapa é vital para evitar que tendências temporais ou padrões geográficos indesejados introduzam ruído, ou distorção nas análises subsequentes.

[33] https://dados.gov.br/dados/conjuntos-dados/comportamento-das-chuvas-e-vazoes-diferenca-nas-chuvas-2021

[34] ETL (Extract, Transform, Load) - processo de combinação de dados de várias fontes em um grande repositório central, chamado de Data Warehouse (Rodrigues; Maciel, 2022).

Além da normalização, a geolocalização necessitava ser precisamente mapeada. Sabendo-se que diferentes regiões podem ter distintas taxas de criminalidade e padrões climáticos, uma combinação precisa dos dados pluviométricos com as regiões geográficas correspondentes, foi crucial. Esta associação geoespacial permitiu uma análise mais granular, possibilitando a identificação de tendências ou anomalias específicas de certas regiões.

Foi essencial reconhecer que os dados pluviométricos, por mais detalhados que fossem, representavam apenas um lado da equação. Para construir um modelo preditivo robusto e significativo, a sinergia com outros conjuntos de dados foi essencial. Nesse contexto, a Integração com Dados de Segurança Pública se destacou, complementando e enriquecendo a complexidade e profundidade da investigação.

5.1.2 Integração com Dados de Segurança Pública

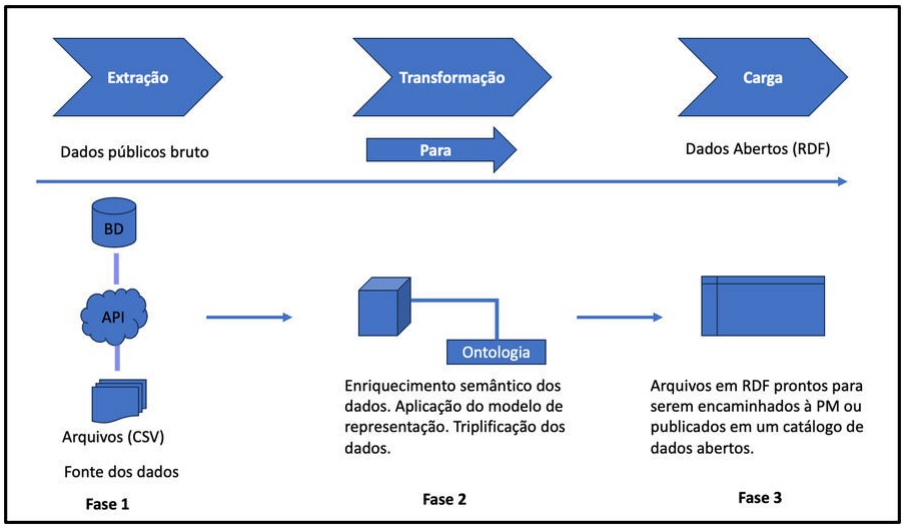
Da mesma forma, foi realizada coleta de dados de segurança pública, buscando analisar, a princípio, informações sobre crimes de violência doméstica e contra o patrimônio. Uma vez obtidos, os dados foram mapeados semanticamente para garantir uma integração coesa com os dados pluviométricos.

Nesta fase do trabalho o uso da metodologia ETL foi muito importante, pois foi realizada a mineração dos dados, transformando-os em dados abertos de qualidade, no formato RDF, para deixá-los prontos para publicação e entrega ao departamento de tecnologia da informação da Polícia Militar.

Neste sentido, foram considerados aspectos como fonte de dados, processo de extração e transformação dos dados, incluindo o enriquecimento semântico por meio do uso de ontologia para agregar valor aos dados. Também foi considerado o processo de triplificação dos dados (Sujeito, Predicado e Objeto) e a conversão para o formato RDF (Rodrigues; Maciel, 2022).

Baseando-se em Rodrigues e Maciel (2022, p. 3), é apresentada a Figura 11 representando um esboço para tal método.

Figura 11 – Estrutura do método



Fonte: Rodrigues e Maciel (2022, p. 3).

Para a extração dos dados criminais foram utilizadas as bases do site Dados Abertos, do governo federal[35], bem como as bases de ocorrências da Secretaria da Segurança Pública do Estado de São Paulo[36]. Essa extração ocorreu por meio de consumo de API (*Application Programming Interface*) das referidas bases de dados abertas e, consecutivamente, foram gerados dados em formato ".CSV".

No processo de coleta foi crucial a contextualização temporal e geográfica desses dados, uma vez que dados coletados em um determinado período ou local podem não ser relevantes em outro, ou, para observar tendências, padrões e mudanças, no decorrer do período, foi necessário saber quando os dados foram coletados (Jacob; Lefgren; Moretti, 2007).

Para assegurar que as correlações a ser estabelecidas posteriormente fossem válidas, foi necessário considerar o período específico de coleta de cada dado e a localização exata de sua origem. Esta etapa permitiu a correta sincronização com os dados de segurança pública, garantindo que as análises tenham sido baseadas em comparações válidas e temporalmente alinhadas.

Ao abordar o universo dos dados de segurança pública, aborda-se uma série de complexidades inerentes à sua natureza. Estes registros, frequentemente dispersos em várias fontes devido à descentralização dos órgãos de segurança, precisam ser coletados com uma abordagem sistematizada (Lima; Sinhoretto; Bueno, 2015).

[35] Disponível em: https://dados.gov.br/dados/busca?termo=crimes. Acesso em: 15 jan. 2024.

[36] Disponível em: https://www.ssp.sp.gov.br/estatistica/painel-estatistico. Acesso em: 2 jan. 2023.

Primeiramente, foi necessária uma identificação criteriosa das bases de dados confiáveis e atualizadas relacionadas à segurança pública. Isso envolveu uma seleção de fontes que documentam não apenas a frequência de crimes, mas também o tipo, localização e, idealmente, contextos ou causas associadas. A uniformidade dos dados é fundamental, portanto, os registros selecionados foram processados para seguir um formato padrão, facilitando a integração subsequente (Rocha, 2020).

Outra consideração essencial foi a temporalidade. Assim como os dados pluviométricos são sensíveis ao período de coleta, os registros de segurança pública também devem ser tratados com essa perspectiva. Isso significa que os dados criminais de um mês específico devem ser correlacionados com os registros climáticos do mesmo período (Rocha, 2020). Portanto, foi importante garantir que os intervalos de tempo em ambas as bases de dados se alinhassem perfeitamente, permitindo uma análise comparativa mais precisa.

A integração foi facilitada pelo uso de ferramentas e algoritmos especializados que pudessem mapear semanticamente os registros de chuva e vazão com os de criminalidade. Este processo, além de técnico, precisou ser guiado por uma abordagem teórica que justificasse e explicasse as possíveis relações entre os fenômenos observados.

5.2 Compreender e Identificar a estrutura dos dados

Uma vez concluída a fase de coleta e tratamento dos dados via ETL, foi necessária a compreensão estrutural dos dados, visto que esta fase é crucial, pois a estrutura correta dos dados determina, na maioria, sua utilidade, interoperabilidade e as possíveis abordagens analíticas que podem ser aplicadas.

A estrutura adequada do conjunto de dados é a base para uma acessibilidade eficaz e capacidade de integração a sistemas e processos de conexão de dados em tempos futuros, ou seja, para haver o reúso dos novos dados é necessária uma estruturação adequada para as análises, pois garante que os métodos e técnicas sejam apropriados ao tipo e natureza dos dados (Araújo, 2017).

Ao explorar a natureza fundamental da estrutura dos dados, é fundamental observar a relevância dos formatos de arquivo no contexto do armazenamento e manipulação desses dados. Desta forma, os formatos .CSV, .JSON e .XML desempenharam funções essenciais nesse contexto, cada um com suas características distintas que se adaptam a diversas necessidades e cenários.

O formato denominado CSV (*Comma-Separated Values*) é uma representação direta que utiliza vírgulas para separar valores. A sua estrutura plana lhe confere uma adequação para a manipulação de dados tabulares, tais como aqueles encontrados em planilhas ou bases de dados relacionais. A extração de dados de um arquivo CSV é fácil devido à sua simplicidade, tornando-o uma escolha preferida em muitos processos ETL. No entanto, a simplicidade do método pode apresentar uma desvantagem ao lidar com a representação de relações mais complexas entre os dados (Shafranovich, 2005).

O formato JSON (*JavaScript Object Notation*) é um formato leve de troca de dados interpretável para os humanos ler e escrever, e fácil para as máquinas analisar e gerar. É normalmente utilizado para representar dados estruturados e é particularmente vantajoso em cenários em que a estrutura de dados pode ser irregular ou os dados têm hierarquias complexas (Bray, 2014).

O formato XML, também conhecido como *Extensible Markup Language*, fornece uma estrutura abrangente para a descrição de dados. O *design* do sistema permite a personalização de estruturas de dados e é amplamente empregado em aplicações web, assim como em diversos processos de extração, transformação e carga de dados (Bray *et al*., 2006).

Portanto, ao abordar a compreensão e identificação da estrutura de dados, é necessário examinar o formato de armazenamento apropriado e a padronização dos dados. Cada formato tem seus pontos fortes e limitações, e a escolha dependerá das especificidades do conjunto de dados em questão e das necessidades da análise subsequente (Araújo, 2017).

5.3 Padronização de dados

A padronização, no âmbito da gestão de dados, funciona como uma espécie de "tradutor universal", garantindo que informações de diversas fontes possam ser comparadas e analisadas de forma coerente. Considerando a extensa gama de fontes de dados disponíveis, cada uma com suas peculiaridades e formatos, o processo de padronização surge como um facilitador, permitindo que analistas e sistemas computacionais compreendam informações de diversos contextos sem ambiguidades (Damasceno *et al.,* 2011).

Além disso, a padronização de dados também contribui significativamente para a qualidade geral dos conjuntos de dados. Quando os dados são padronizados, a probabilidade de ocorrência de erros resultantes de inconsistências ou ambiguidades

diminui significativamente (Damasceno *et al*., 2011, p. 2). Isso não apenas aprimora a confiabilidade das análises derivadas desses conjuntos de dados, mas também aumenta o valor dos resultados gerados, uma vez que eles se fundamentam em informações mais sólidas e coerentes.

A normalização pode poupar tempo e recursos. Embora o investimento inicial no processo de padronização possa parecer oneroso, os benefícios que se seguem, como a redução do tempo gasto na limpeza dos dados e na preparação para análises futuras, bem como a minimização de erros e retrabalho, tornam este processo indispensável. Em um contexto global cada vez mais influenciado pela análise de dados, é de extrema importância possuir conjuntos de dados que sejam bem estruturados e sigam padrões estabelecidos. Essa prática é essencial para otimizar a extração de valor das informações disponíveis.

5.3.1 Knowledge Discovery in Data Base

A padronização e a normalização de dados não apenas facilitam o processo analítico, poupam tempo e recurso, mas também são fundamentais para garantir a precisão e a relevância das descobertas no contexto do Conhecimento em Base de Dados (*KDD - Knowledge Discovery in Data Base*). A qualidade dos dados que entram em qualquer sistema de KDD tem uma influência direta sobre a validade dos padrões extraídos. Quando os conjuntos de dados são bem estruturados e aderem a padrões estabelecidos, eles se tornam mais confiáveis para análise e essa confiabilidade é crucial, pois os padrões extraídos através do KDD devem ser válidos, compreensíveis e úteis (Fayyad *et al*., 1996).

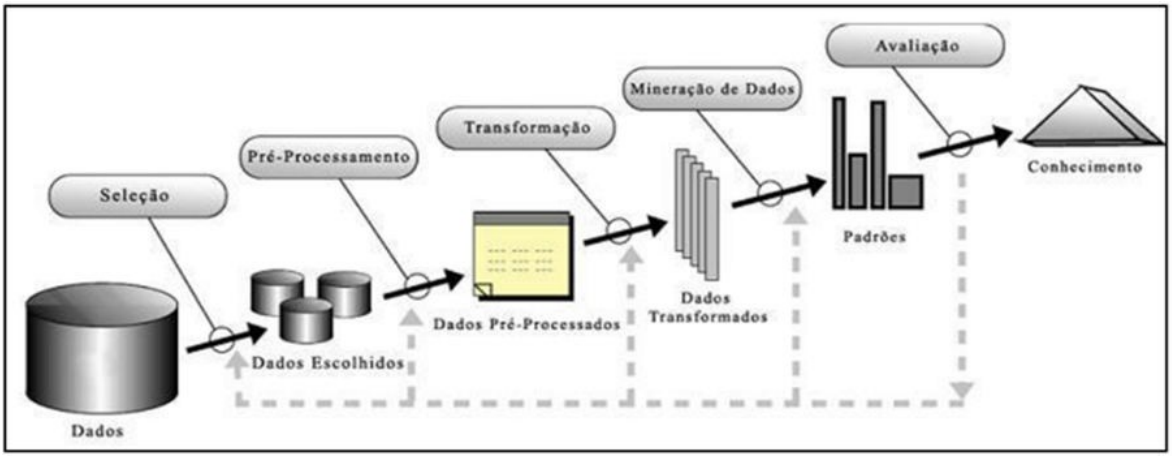
O KDD pode ser entendido como o processo não trivial de extração de padrões válidos de dados em grandes bases, demonstrando o conhecimento implícito ali apreendido e considerados potencialmente úteis. Tais padrões devem ser confiáveis, compreensíveis e úteis, para que o conhecimento obtido possa ter seu uso, científico ou comercial, aproveitado. São estabelecidas métricas com estimativas estatísticas para definir a utilidade desses padrões, tais como níveis de confiança, compreensão e utilidade (Fayyad *et al*., 1996; Han; Kamber, 2006).

O KDD é um processo que perpassa várias fases: seleção dos dados; pré-processamento e limpeza dos dados; transformação dos dados; Data Mining (DM); interpretação e avaliação dos resultados, sendo que, para o processo de padronização de dados, o DM tem papel de destaque (Fayyad *et al.,* 1996; Han; Kamber, 2006).

As fases que compõem o processo do KDD, como a escolha de uma amostra de dados, o pré-processamento, a transformação dessa amostra, a mineração por meio da

manipulação deles com a utilização de algoritmos e sua devida interpretação, culminam no conhecimento de uma informação antes não conhecida. O termo Mineração de Dados muitas vezes é empregado tanto para caracterizar a cadeia de processos do KDD quanto para designar apenas a fase de DM. Fayyad *et al.* (1996) propõem a distinção dos termos, especificando DM como os meios pelos quais se extraem e se identificam a série de padrões na base de dados. A Figura 12 ilustra as fases do processo KDD.

Figura 12 – Etapas do fluxo do processo KDD



Fonte: Fayyad et al. (1996, p. 84).

De acordo com Avelar, Rocha e Cruz (2017), a relação entre a DM e a padronização de dados é intrínseca e se manifesta em diversas etapas do processo analítico. Ambos são elementos essenciais para garantir a extração eficaz e precisa de informações de conjuntos de dados. Uma representação mais detalhada é apresentada a seguir, de acordo com Campos *et al*. (2020):

1) **Qualidade do Dado:** A mineração de dados envolve o processo de analisar grandes volumes de dados para descobrir padrões e informações anteriormente não conhecidas. Para que esse processo seja eficaz, os dados em questão devem ser de alta qualidade. A padronização contribui para essa qualidade, assegurando que os dados sejam consistentes e comparáveis em todo o conjunto.
2) **Preparação de Dados**: Antes de iniciar o processo de mineração, é comum que os dados passem por uma fase de pré-processamento. Esta etapa inclui limpeza, transformação e, muitas vezes, padronização dos dados. A padronização nesse contexto garante que os dados estejam em um formato uniforme, facilitando algoritmos de mineração a processá-los eficientemente e a produzir resultados mais confiáveis.

3) **Comparabilidade:** Algoritmos de mineração frequentemente fazem comparações entre diferentes pontos de dados. Se os dados não estiverem padronizados, essas comparações podem ser inválidas ou imprecisas.
4) **Integração de Múltiplas Fontes**: Em muitos projetos, os dados podem ser oriundos de fontes diferentes. A padronização é crucial para que esses diferentes conjuntos de dados possam ser combinados em uma única fonte coesa que pode ser minerada.
5) **Eficiência Computacional:** Dados padronizados geralmente exigem menos recursos computacionais durante o processo de mineração. Algoritmos podem operar mais rápido e eficientemente quando os dados estão em um formato uniforme.
6) **Validade dos Resultados:** A validade e confiabilidade dos padrões descobertos por meio da mineração são amplamente influenciadas pela qualidade dos dados de entrada. Dados padronizados e bem preparados aumentam a probabilidade de que os resultados da mineração sejam válidos e aplicáveis.

Para Berry e Linoff (1997), DM é considerada a parte mais importante do processo, pois nessa fase são definidas as tarefas e métodos, e aplicados os algoritmos escolhidos sobre os resultados dos dados transformados.

Segundo Herrera Varela (2006), em teoria, a DM pode ser aplicada a qualquer tipo ou quantidade de dados, mas são comumente aplicadas a grandes volumes, também é essencial em pesquisas científicas e técnicas, como instrumento de análise e descoberta de conhecimento, advindos das observações dos dados ou resultados dos ensaios.

Considerada uma área multidisciplinar, a DM agrega áreas como Inteligência Artificial (IA), Estatística, Banco de Dados (BD), dentre outras. Sua aplicação é diversificada, sendo utilizada em vários domínios do conhecimento como forma de encontrar conhecimento tácito, difícil de perceber em grandes quantidades de dados, por meio de algoritmos que traçam perfis e padrões para o *corpus* pesquisado (Han; Kamber, 2006).

Em muitas organizações, há a preocupação de se coletar grande quantidade de informações nas mais variadas formas. É fácil digitalizar informações, não é mais excessivamente caro armazená-las e, em princípio, acredita-se que os dados coletados podem ser úteis. É por isso que embora a mineração de dados possa ser aplicada a qualquer tipo de informação, variando apenas as técnicas a serem usadas em cada tipo de estrutura

de dados analisada, principalmente a extração de dados em bancos de dados relacionais, bancos de dados espaciais, banco de dados temporais, bancos de dados documentais e bancos de dados multimídia, também há uma forte tendência desde o advento da internet para extrair informações especialmente da World Wide Web (Herrera Varela, 2006, p. 125). Para Berry e Linoff (1997, p.7), DM "É a exploração e análise de grandes quantidades de dados para descobrir padrões e regras significativos". Para os autores, os instrumentos e métodos da DM são aplicáveis em vários campos, que vão desde a medicina a controle de processos industriais. Profissionais dessa área empregam processos emprestados da estatística, ciência da computação e Machine Learning, e a escolha de uma combinação específica de técnicas para se utilizar em um determinado cenário depende da natureza do trabalho de mineração, do tipo de dados disponíveis e das preferências do profissional.

Na realidade, a DM faz mais sentido quando há grandes volumes de dados, pois a maior parte dos algoritmos de mineração precisa processar grandes quantidades de informações para treinar e construir os modelos que serão usados para realizar: classificação, estimativa, predição, agrupamento por afinidade, clusterização, descrição e perfil, ou outras tarefas da DM (Berry; Linoff, 1997).

Segundo Lobaina e Suárez (2018), a DM é um processo revolucionário e a maneira mais rápida de se analisar grandes volumes de informações para se encontrar padrões e conhecimentos úteis implícitos nesses dados. É uma técnica já utilizada em muitas outras áreas, mas ainda incipiente na área das bibliotecas. Pode-se encontrar ainda na literatura o termo "bibliomineração", utilizado pela primeira vez por Nicholson (2004), que define o termo como a mineração de dados aplicada à biblioteca.

No entanto, para além da padronização, a estrutura semântica dos dados também ocupa uma posição central na otimização da análise de dados. Se, por um lado, a padronização estabelece uma base sólida para a aplicação eficaz de técnicas de Data Mining, por outro, o mapeamento da estrutura semântica usando tecnologias semânticas e dados conectados oferece uma compreensão mais profunda e enriquecida dos dados.

Neste sentido, dar-se-á continuidade ao assunto, na próxima seção, para investigar como essas técnicas podem ser empregadas para dar significado aos padrões descobertos e para criar conexões mais ricas entre os conjuntos de dados.

5.4 Mapeamento da estrutura semântica dos dados com uso de tecnologias semânticas e dados conectados

O advento da era digital foi acompanhado por um aumento significativo na quantidade e na diversidade de dados disponíveis (Manyika *et al*., 2011). No entanto, o simples acesso a uma vasta quantidade de dados não é suficiente para extrair valor e significado destes. Portanto, é essencial entender a estrutura e as relações inerentes entre diferentes organizações de dados. Nesta seção, será abordada a importância de usar técnicas e ferramentas semânticas.

5.4.1 Enriquecimento semântico dos dados por meio de ontologia

O termo "ontologia" deriva do grego "*onto*", que significa ser, e "*logia*", que se refere ao discurso escrito ou falado (Uschold; Gruninger, 1996). Historicamente utilizado na filosofia, o termo tem sido adaptado ao campo da Ciência da Computação para trabalhos relacionados à aquisição de conhecimento e raciocínio. Gruber (1993) definiu uma ontologia como "[...] uma especificação da conceituação [...]", um conjunto formal de conceitos e relações que existem para um agente ou comunidade de agentes.

De acordo com Uschold e Grninger (1996), o termo ontologia pode ser definido também como:

> [...] entendimento compartilhado de algum domínio de interesse que pode ser usado como uma estrutura unificadora para resolver os problemas acima descritos da maneira acima especificada. Uma ontologia necessariamente envolve ou incorpora algum tipo de visão de mundo em relação a um determinado domínio. Essa visão de mundo é frequentemente concebida como um conjunto de conceitos (por exemplo, entidades, atributos, processos), suas definições e suas inter-relações; isso é referido como uma conceitualização [...] (Uschold; Gruninger, 1996, p. 21).

Aplica-se também outro conceito à ontologia que é a importância da formalidade de um modelo conceitual, onde ocorre o acordo de certa comunidade de usuários sobre os termos da ontologia e seu significado, e tão somente este acordo garante a partilha dos conhecimentos (Erdmann; Decker, 2000).

Ontologia, além de ser um termo filosófico, pode também ser descrita nas mais diversas áreas que envolvem tecnologia, dentre elas: Web semântica, Engenharia de Software, Arquitetura da Informação.

A Ontologia, segundo seus componentes, descreve os indivíduos, classes, atributos e relacionamentos. As classes formam as generalizações de coleções, conceitos, tipos de objetos ou espécies de coisas, enquanto os indivíduos são suas realizações concretas,

havendo entre os indivíduos relacionamentos e cada indivíduo tem um conjunto de atributos.

5.4.2 Classificação de Ontologias

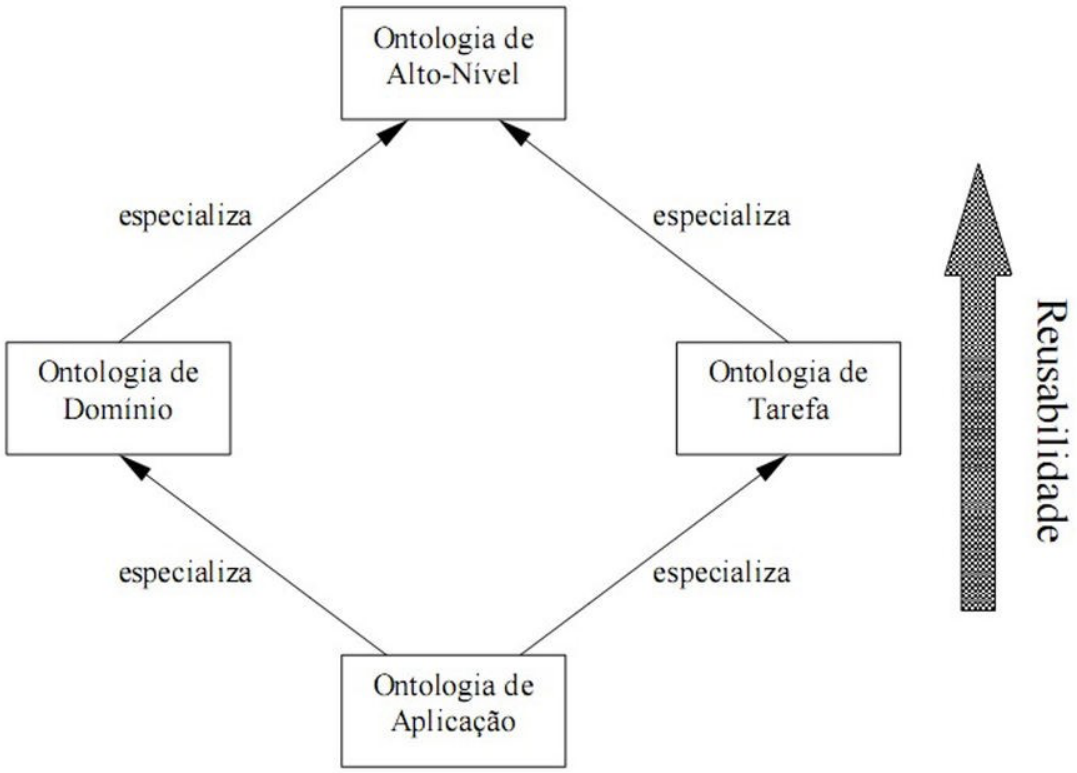
Existem diversas classificações de ontologias, porém, a que mais se relaciona com semântica, e a que se pretende utilizar nesta tese, é a que utiliza conceitualização como critério principal (Guarino, 1998).

As ontologias são divididas em quatro tipos: de alto nível, de domínio, de tarefa e de aplicação. O primeiro tipo proposto por Guarino (1998) são as ontologias de alto nível onde são descritos os conceitos mais genéricos como espaço, tempo, evento. Esses conceitos são livres de um problema ou domínio particular, assim, é bem comum encontrar grandes comunidades de usuários que empregam o compartilhamento de ontologias de alto nível. O segundo tipo sugerido são as ontologias de domínio, onde é descrito um vocabulário relacionado a um domínio genérico, através da especialização de conceitos inseridos nas ontologias de alto nível.

O terceiro tipo são as ontologias de tarefa que apresentam uma concepção relacionada a uma tarefa ou atividade genérica.

Por fim, o quarto tipo são as ontologias de aplicação que são mais específicas por serem utilizadas dentro das aplicações. Este tipo de ontologia especializa conceitos das ontologias de domínio e de tarefas, conforme demonstra a Figura 13.

Figura 13 – Tipos de classificação de ontologias e suas relações.



Fonte: Guarino, (1998).

Nota-se que as ontologias de alto nível são as que possuem maior capacidade de reuso, pois definem conceitos genéricos, já as ontologias de aplicação apresentam menor capacidade de reuso, pois definem conceitos relacionados a uma aplicação específica.

Como em todas as fases de construção de uma Ontologia, alguns critérios devem ser utilizados na estruturação do código de linguagem de programação, principalmente na distinção ou comparação das linguagens para o desenvolvimento. As linguagens de programação, em relação à ontologia, são, na verdade, códigos lidos em páginas web sendo interpretados ou compilados para serem compreendidos por agentes eletrônicos que vão pesquisar as informações (Guarino, 1998).

Para que a WEB[37] tivesse um melhor desempenho e que passasse a não mais ser uma WEB de documentos, puramente em páginas estáticas, e sim uma WEB de dados, foram propostas diversas tecnologias, dentre elas, as ontologias que, de acordo com (Pickler, 2007, p. 63), objetivava com essa tecnologia "[...] atribuir sentido e significado

[37] Rede que conecta computadores por todo mundo por meio do protocolo WWW (World Wide Web) (Bizer *et al.*, 2008).

ao conteúdo de documentos, atuando como ferramenta de representação do conhecimento [...].

Chateaubriant (1998, p. 12) sugere que a ontologia deriva da semântica, que, para o autor "[...] tem a ver com a relação entre linguagem e realidade e é a partir dessa ideia que, evidentemente, ontologia e semântica se conectam [...]".

Pickler (2007, p. 71) reafirma que "[...] se a semântica tem a ver com a relação entre a linguagem e a realidade, então a ontologia é o estudo da estrutura da realidade, dessa forma, havendo relação entre elas [...]".

Com o advento do serviço de internet, por meio do protocolo WWW (World Wide Web), no início da década de mil novecentos e noventa surgem, também, as grandes bases de conhecimento; assim, as tecnologias computacionais, dentre elas, a ontologia, precisaram migrar seus projetos para o gerenciamento e a organização dessas grandes bases, a fim de garantir interoperabilidade e estruturação eficazes (Pickler, 2007).

Pickler (2007, p. 72) assevera que

> [...] no contexto da web e da inteligência artificial, o termo ontologia foi adaptado e, para os profissionais dessas áreas, uma ontologia é um documento ou um arquivo que define formalmente as relações entre termos e conceitos, mantendo, nesse sentido, semelhanças com tesauros[38] utilizados para definição de vocabulários controlados [...].

Assim, as ontologias servirão como representação do conjunto básico de palavras e conceitos para facilitar a interação entre agentes e páginas WEB, estabelecendo as conexões semânticas entre diferentes conceitos.

Tendo estabelecido a importância das ontologias para representar conjuntos básicos de palavras e conceitos, para facilitar a interação semântica entre agentes e páginas web, é crucial examinar como essas ontologias são efetivamente implementadas e gerenciadas. Nesse contexto, a *Resource Description Framework* (RDF) surge como uma tecnologia-chave. RDF proporciona o arcabouço necessário para a formalização e compartilhamento de ontologias, agindo como uma camada subjacente que permite a interoperabilidade e a estruturação eficazes dos dados semânticos.

[38] Termos utilizados em indexação e na classificação de documentos, bem como modalidades de sistemas de organização do conhecimento (Moreira, 2019).

5.5 Resource Description Framework (RDF)

O *Resource Description Framework* (RDF) se apresenta como uma das ferramentas mais importantes, uma vez que, por meio do RDF, os dados são estruturados em triplas, formando uma base para representações semânticas. Estas triplas, consistindo em sujeito, predicado e objeto, permitem representar informações para estabelecer relações claras entre diferentes entidades e atributos. Em termos práticos, o RDF facilita a criação de uma rede de informações interconectadas, tornando os dados não apenas legíveis por máquinas, mas também contextualmente ricos (Cruz, 2015, p. 32).

Criar triplas RDF para ocorrências de violência doméstica, como exemplo, envolve a modelagem semântica de informações pertinentes a esses eventos. As triplas RDF são composições de "sujeito-predicado-objeto" (Rodrigues; Maciel, 2022) que descrevem relações entre entidades e valores.

É importante estabelecer quais aspectos criminais devem ser representados, uma vez que a precisão e a abrangência dessas informações determinam a eficácia das análises e interpretações posteriores. A escolha desses aspectos deve considerar não apenas o crime, mas também o contexto social, legal e geográfico em que ocorreu (Ishak, 2022). Além disso, é essencial ponderar sobre a sensibilidade de certas informações, garantindo que a privacidade das vítimas e demais envolvidos seja preservada, ao mesmo tempo em que se oferece uma visão holística e compreensiva do incidente para fins de investigação e políticas públicas (Cruz, 2015).

É apresentado, a seguir, um exemplo simplificado de como as triplas poderiam ser estruturadas:

Definição das entidades:

Pessoa (vítima, agressor)

Local (onde ocorreu o incidente)

Ocorrência (a instância específica de violência doméstica)

Definição das propriedades:

tipoDeViolência (física, psicológica, sexual, etc.) dataDaOcorrência

relaçãoEntreVítimaEAgressor (cônjuge, parceiro, familiar, etc.)

Exemplo das triplas RDF - Figura 14:

Figura 14 – Exemplo das triplas RDF

```
<#Ocorrência1> <#tipoDeViolência> "física".
<#Ocorrência1> <#dataDaOcorrência> "2023-08-10".
<#Ocorrência1> <#ocorreuEm> <#Local1>.
<#Ocorrência1> <#envolveVítima> <#PessoaMaria>.
<#Ocorrência1> <#envolveAgressor> <#PessoaJoão>.
<#PessoaMaria> <#éVítimaDe> <#PessoaJoão>.
<#PessoaJoão> <#agrediu> <#PessoaMaria>.
<#PessoaMaria> <#temRelação> "cônjuge".
```

Fonte: o próprio autor (2023).

Por apresentar uma natureza expansiva e flexível, uma das principais vantagens do RDF reside na sua capacidade de ser expandido e modificado sem a necessidade de reestruturar todo o conjunto de dados. Essa adaptabilidade é especialmente relevante no ambiente dinâmico da Ciência de Dados, onde novas informações e relações podem ser descobertas frequentemente, demandando uma incorporação contínua ao conjunto de dados existente.

Além disso, a estrutura de triplas do RDF é amplamente compatível com diversos sistemas e bancos de dados. Seu formato padronizado garante uma integração mais suave com outras tecnologias semânticas, facilitando a migração, compartilhamento e disseminação de informações. Neste contexto, é relevante mencionar sistemas de gerenciamento de triplas, como Virtuoso[39] e Jena[40], os quais são especialmente concebidos para armazenar, consultar e manipular dados RDF, reforçando assim sua aplicabilidade prática.

Dada a sua natureza granular e a multiplicidade de triplas que podem ser geradas, especialmente em grandes conjuntos de dados, pode haver implicações de desempenho durante a consulta e análise. Isto sublinha a necessidade de otimização e de ferramentas robustas de indexação. Apesar desses desafios, o valor proporcionado pelo RDF em representações semânticas detalhadas e interconexões ricas entre dados torna-o uma ferramenta insubstituível no arsenal do cientista de dados.

[39] Disponível em: https://virtuoso.openlinksw.com/. Acesso em: 12 ago. 2023.

[40] Disponível em: https://jena.apache.org/. Acesso em: 12 ago. 2023.

5.6 SPARQL

O SPARQL[41], dada a sua especificidade para consultas em estruturas baseadas em RDF, é uma ferramenta inestimável para aprofundar as análises em dados semânticos. Um de seus recursos mais notáveis é a capacidade de conduzir buscas federadas, permitindo que pesquisadores consultem múltiplas fontes de dados RDF simultaneamente. Isto significa que é possível colher informações de diversas bases, distribuídas geograficamente ou mantidas por diferentes entidades, em uma única consulta. Tal característica abre horizontes para investigações multidisciplinares e colaborativas, que seriam quase impraticáveis sem esta capacidade integrativa (Calvanese *et al*., 2016).

Outro ponto forte do SPARQL é sua adaptabilidade sintática e semântica. Através da definição de PREFIXOS, é possível simplificar e tornar as consultas mais intuitivas, permitindo uma leitura fluida e compreensível até mesmo para aqueles menos versados em linguagens de consulta. Além disso, a flexibilidade do SPARQL se estende ao retorno de dados, os quais podem ser formatados de várias formas, desde tabelas convencionais até formatos mais complexos como JSON ou RDF/XML. Essa multifuncionalidade garante que os dados possam ser diretamente integrados em diversas aplicações ou sistemas de análise posteriores (Banerjee *et al*., 2022).

Com os dados estruturados em RDF, surge a necessidade de uma linguagem de consulta que possa acessar e manipular essas informações interligadas. Aqui, o SPARQL (SPARQL Protocol and RDF Query Language) assume seu papel primordial. SPARQL permite que pesquisadores e cientistas de dados realizem consultas complexas dentro de bases de dados RDF, extraiam *insights* específicos e explorem relações semânticas que poderiam permanecer ocultas com métodos de consulta tradicionais (Banerjee *et al*., 2022).

No entanto, como qualquer tecnologia, o SPARQL também possui seus desafios. O mais proeminente é, possivelmente, a curva de aprendizado associada à construção de consultas eficientes e otimizadas. Uma consulta mal estruturada pode resultar em tempos de resposta prolongados ou em retornos excessivamente volumosos. Este desafio sublinha a necessidade de capacitação adequada e prática contínua para dominar a arte de consultar dados RDF com SPARQL, garantindo que se extraia o máximo valor dos dados semânticos disponíveis (Calvanese *et al*., 2016).

[41] Disponível em: https://www.w3.org/TR/rdf-sparql-query/. Acesso em: 12 ago. 2023.

5.7 *Extensible Markup Language* (XML)

A tecnologia XML (*Extensible Markup Language*) serve como um meio crucial para a representação e intercâmbio de dados estruturados. Sua natureza hierárquica e a capacidade de definir esquemas específicos para domínios particulares o tornam instrumental na integração e comunicação entre sistemas distintos. No contexto da Ciência de Dados, o XML muitas vezes serve como um ponto intermediário ou formato de exportação/importação entre sistemas semânticos e não semânticos (Tekli; Chbeir; Yetongnon, 2009).

O XML, ao longo dos anos, consolidou-se como uma linguagem padrão para a troca de informações entre sistemas heterogêneos. Seu design é fundamentado em marcadores, permitindo uma descrição detalhada e estruturada de dados. Essa capacidade de detalhamento e personalização permite que diferentes setores estabeleçam suas próprias especificações baseadas em XML (Tekli; Chbeir; Yetongnon, 2009).

Um ponto crucial do XML é sua capacidade de trabalhar com DTD (*Document Type Definition*) e XML Schema. Estas ferramentas permitem que os desenvolvedores definam regras e estruturas específicas para seus documentos XML, garantindo que os dados correspondam com um formato pré-estabelecido. Esta validação é vital, especialmente quando se trata da troca de informações entre sistemas, pois assegura que os dados recebidos ou enviados estejam em conformidade com o esperado, reduzindo assim os erros de integração e processamento (Almeida, 2002).

Apesar de suas múltiplas vantagens, o XML apresenta algumas desvantagens. Sua verbosidade e complexidade podem levar a maiores requisitos de armazenamento e tempos de processamento quando comparados a formatos mais leves, como o JSON.

Contudo, sua robustez, flexibilidade e capacidade de autodescrição compensam essas desvantagens, especialmente em contextos em que a precisão, validação e interoperabilidade são essenciais. Assim, no universo expansivo da ciência de dados, o XML continua a ocupar um lugar de destaque como uma ferramenta confiável para a modelagem e intercâmbio de informações estruturadas (Almeida, 2002).

Nesta seção, procurou-se explorar e apresentar a confluência entre a Ciência da Informação e a Prevenção de Crimes no contexto dos Dados Abertos Conectados, na Segurança Pública. A seção delineou desde o levantamento de fontes de dados relevantes, como dados pluviométricos e de vazão, até a sua integração com dados de segurança pública. A compreensão e a identificação da estrutura dos dados foram destacadas como

etapas fundamentais antes de se aprofundar em técnicas mais avançadas de padronização de dados através do KDD e Data Mining.

Destacou-se o mapeamento da estrutura semântica dos dados para extração relevante de conhecimento. A implementação das ontologias e suas diversas classificações proporcionam um enriquecimento semântico que permite o compartilhamento de conhecimento de forma eficiente e precisa. Tecnologias como RDF, SPARQL e XML foram discutidas como ferramentas essenciais para a implementação eficaz de soluções em dados abertos conectados.

5.8 Policiamento preditivo: sugestão de modelo informacional

Esta tese tem como objetivo geral "Desenvolver um modelo de policiamento preditivo que utilize dados abertos conectados e princípios da Ciência da Informação. Esse modelo buscará integrar e analisar os dados não conectados, considerando a sazonalidade e outros fatores relevantes, a fim de prever e identificar áreas com maior probabilidade de ocorrência de crimes".

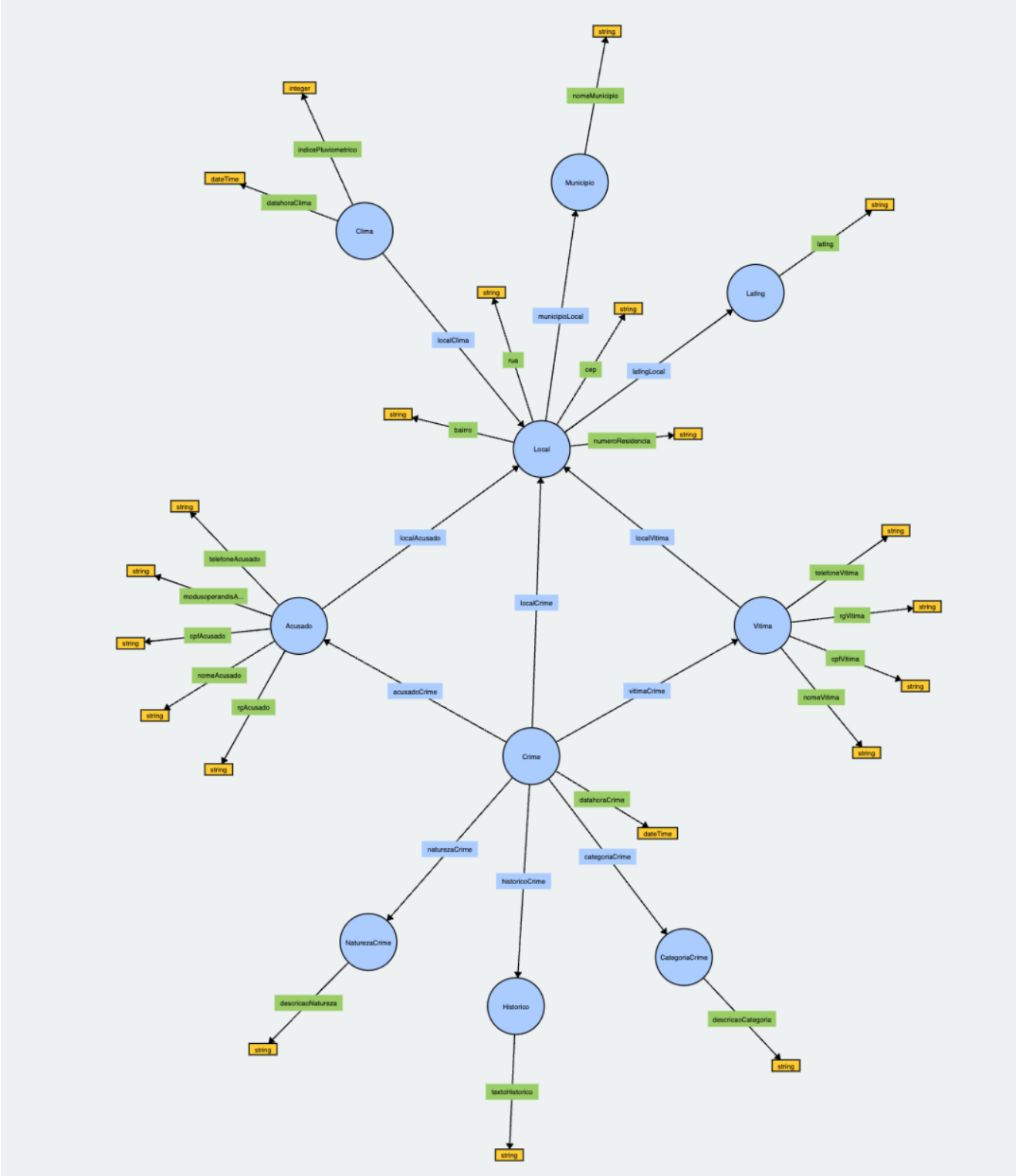
Em cumprimento a este objetivo geral, está sendo sugerido nesta subseção um modelo de ontologia, ilustrado na Figura 15, que descreve a base semântica para um sistema em desenvolvimento de predição de crimes e que está em fase de planejamento e implementação pela equipe de Tecnologia da Informação (TI) da Polícia Militar do Estado de São Paulo. A intenção foi desenvolver uma aplicação robusta que integrasse dados históricos e em tempo real sobre crimes e condições climáticas, permitindo que a equipe de TI desenvolva algoritmos de *Machine Learning* ou de outras técnicas estatísticas, para prever a probabilidade de crimes em diferentes locais e momentos.

Na fase atual, a equipe está trabalhando na coleta e estruturação dos dados, conforme definidos pela ontologia, garantindo que eles sejam precisos, relevantes e devidamente relacionados para permitir análises significativas. A fase de desenvolvimento também pode incluir a criação de interfaces de usuário para que os analistas de dados possam interagir com o sistema, realizar consultas e interpretar os resultados das previsões.

Além disso, a equipe de TI terá que trabalhar em estreita colaboração com os especialistas em segurança pública, para validar e refinar o modelo de predição, assegurando que as previsões sejam práticas e úteis para o planejamento estratégico e operacional da Polícia Militar. A aplicação também terá que cumprir com todas as leis e regulamentos de privacidade e proteção de dados, especialmente no que diz respeito à informação pessoal das vítimas de crimes.

Uma vez desenvolvido, o sistema de predição de crimes poderá ser uma ferramenta valiosa para aumentar a segurança em áreas de risco e prevenir a ocorrência de crimes.

Figura 15 – Proposta de Ontologia



Fonte: o próprio autor (2023).

A ontologia apresentada é uma representação de um modelo de dados que pode ser usado para a predição de crimes, considerando fatores relacionados ao clima e informações específicas sobre os crimes de violência doméstica e furto.

A seguir são descritas as classes e suas relações em formato de triplas RDF (*Resource Description Framework*), que é um padrão para a troca de dados na Web, o "Apêndice A" apresenta o código fonte desta modelagem.

As triplas RDF são expressas no formato "sujeito-predicado-objeto", representando, respectivamente, a entidade, a relação entre as entidades e o valor ou outra entidade relacionada.

Classe Clima: Armazena informações sobre o clima, que podem influenciar na ocorrência de crimes.

Clima - DataHora: Indica a data e a hora dos índices pluviométricos.

Clima - IndicePluviometrico: Relaciona o clima ao índice pluviométrico, que é um fator relevante na análise, representa a quantidade de chuva medida em um local e está associado à classe Clima. Essa classe também se relaciona com Local.

Classe Crime: Contém detalhes sobre o crime ocorrido.

Crime - DataHora: O crime está associado a uma data e hora específicas.

Crime - Local: O crime está vinculado a um local específico.

Crime - NaturezaCrime: A natureza do crime, que descreve o tipo de crime cometido.

Crime - Vitima: Relaciona o crime à vítima afetada.

Crime - DataHora: Indica que a data e hora que o crime ocorreu.

Classe Local: Representa o local onde um evento (como um crime) ocorreu.

Local - Municipio: O local está dentro de um município.

Local - Bairro: O local especifica um bairro, que é um tipo de *string*.

Local - Rua: O local especifica uma rua, também uma *string*.

Local – CEP: Representa o código de endereço postal.

Local – numeroResidência: Indica o número da residência

Local – Latlng: Representa a latitude e longitude onde ocorreu o crime.

DataHora - Clima | Crime: Indica que a data e hora pertencem ou estão associadas a um clima ou crime.

Classe NaturezaCrime: Descreve o tipo ou natureza do crime.

NaturezaCrime - DescricaoNatureza: A classe NaturezaCrime tem uma propriedade que descreve a natureza do crime em forma de texto.

Classe Vitima: Armazena informações sobre a vítima de um crime.

Vitima - nomeVitima: O nome da vítima do crime.

Vitima - cpfVitima: O CPF da vítima, que é um identificador pessoal no Brasil.

Vitima - rgVitima: O RG da vítima, que é um identificador pessoal no Brasil.

Vitima - telefoneVitima: A vítima pode possuir um número de telefone, geralmente uma *string*.

Classe Acusado: Armazena Informações sobre o acusado de ter cometido o crime.

Acusado – nomeAcusado: O nome do acusado de cometer um crime.

Acusado – cpfAcusado: O CPF do acusado, que é um identificador pessoal no Brasil.

Acusado - rgAcusado: O RG da vítima, que é um identificador pessoal no Brasil.

Acusado - telefoneAcusado: A vítima pode pssuir um número de telefone, geralmente uma string.

Classe Municipio: Representa o município onde o local está situado.

Municipio - nomeMunicipio: Descreve o nome do município.

Classe Histórico: Representa as informações sobre o histórico do crime.

Histórico - Crime: Esta relação indica que a classe Histórico está associada a um ou mais crimes. Cada instância de Histórico conteria informações sobre um evento criminal específico ou sobre uma série de eventos relacionados a esse crime.

Classe CategoriaCrime: Delimita a ocorrência do tipo Crimes contra o patrimônio ou crimes contra pessoas.

Classe Latlng: Representa a latitude e longitude do local em que o crime ocorreu.

O modelo de ontologia proposto é a espinha dorsal do sistema de policiamento preditivo que está sendo desenvolvido. Este sistema é uma tecnologia avançada que visa antecipar crimes antes que eles ocorram, utilizando a análise de dados complexos sobre o clima e ocorrências criminais passadas. A relação entre as condições climáticas e as ocorrências de crimes é um aspecto central deste modelo. Estudos anteriores como os de Perry *et al*., 2013; Brüderle, Peters e Roberts, 2017; Ishak, 2022 e Vu Thuy Huong Le *et al*., 2022 sugerem que fatores como a temperatura, precipitação e outras condições meteorológicas podem ter uma influência significativa sobre a probabilidade de crimes ocorrerem, como violência doméstica e furto.

A classe “Clima”, é uma fonte valiosa de dados que pode oferecer informações relevantes sobre padrões de crimes. A correlação entre o clima adverso e o aumento da incidência de crimes pode ser uma tendência observável. O índice pluviométrico, uma subclasse de Clima, pode ser particularmente útil; altos índices de chuva podem levar a um aumento nos crimes relacionados a acidentes ou inundações, enquanto clima seco e quente pode estar relacionado a outros tipos de delitos.

A classe “Crime” é central para a modelagem semântica, pois é aqui que os incidentes são registrados com detalhes específicos. Cada ocorrência é uma instância com atributos como data e hora, local e natureza do crime. Entender as características desses crimes ajuda a formular teorias preditivas. A análise de dados pode revelar que certos tipos de crimes são mais comuns durante certas condições climáticas ou em determinados horários.

A “NaturezaCrime” é uma classe que permite a categorização detalhada dos crimes. Esta categorização é essencial para o desenvolvimento de um modelo preditivo eficaz. Por exemplo, saber que a violência doméstica aumenta durante períodos de alta tensão emocional associados a certos eventos climáticos pode ajudar a polícia a prevenir esses

crimes através de campanhas de conscientização ou aumento do patrulhamento em áreas de alto risco.

A classe “Local” desempenha um papel crítico, fornecendo o contexto geográfico dos crimes. A análise geoespacial é um componente fundamental do policiamento preditivo, pois permite que a polícia identifique e monitore áreas com alta incidência de crimes. Esta análise pode ser aprofundada ao considerar dados de nível de rua e bairro, possibilitando uma resposta policial mais direcionada e eficiente.

O elemento temporal, representado pela classe “DataHora”, é outro componente crítico. O tempo não é apenas um registro de quando um crime ocorreu; ele pode também fornecer padrões quando cruzado com dados climáticos e criminais. Por exemplo, pode-se descobrir que certos crimes ocorrem mais frequentemente em determinadas épocas do ano ou que as condições climáticas extremas em um dia específico levaram a um pico inesperado de incidentes.

A classe “Vitima” oferece uma dimensão humana essencial para a análise. Os dados demográficos e pessoais podem ajudar a identificar se determinados grupos estão mais vulneráveis a certos tipos de crimes sob condições climáticas específicas. Esta informação pode ser usada para proteger proativamente essas populações através de políticas de segurança pública e programas de prevenção.

A classe “Acusado” é vital para o policiamento preditivo, por várias razões. Primeiro, ela permite que a polícia mantenha um registro dos suspeitos e suas características, o que é útil para a investigação e prevenção de crimes. Segundo a análise dos dados sobre os acusados pode revelar padrões de comportamento, como a tendência de cometer crimes em certas condições climáticas ou locais. Terceiro, o *modus operandi*, quando combinado com outros dados, pode ajudar a identificar e prever séries de crimes, permitindo que a polícia implemente medidas específicas de prevenção.

Além disso, a análise de dados da classe “Acusado” pode ser usada para aprofundar o entendimento das causas sociais do crime e desenvolver programas de reabilitação e prevenção mais eficazes.

A classe “Municipio” permite uma visão macro dos eventos, ajudando a discernir se existem fatores locais que contribuem para a criminalidade. Isso pode incluir aspectos como iluminação pública, presença policial e serviços sociais. Uma abordagem holística que considera tanto a micro quanto a macroescala é essencial para um sistema de policiamento preditivo abrangente.

O desenvolvimento desse sistema de policiamento preditivo, apoiado por uma ontologia bem estruturada, é um passo inovador no uso da tecnologia para a segurança pública. Ao conectar pontos de dados que anteriormente pareciam desconectados, as autoridades podem se antecipar aos eventos e agir de forma preventiva, e não apenas reativa. Isso não só pode ajudar a salvar vidas e propriedades, mas também a criar comunidades mais seguras e resilientes. A chave para o sucesso desta iniciativa será a contínua colaboração entre os especialistas em dados, os meteorologistas e os profissionais de segurança pública, garantindo que o sistema esteja sempre se adaptando e melhorando com base nas últimas evidências e tecnologias disponíveis.

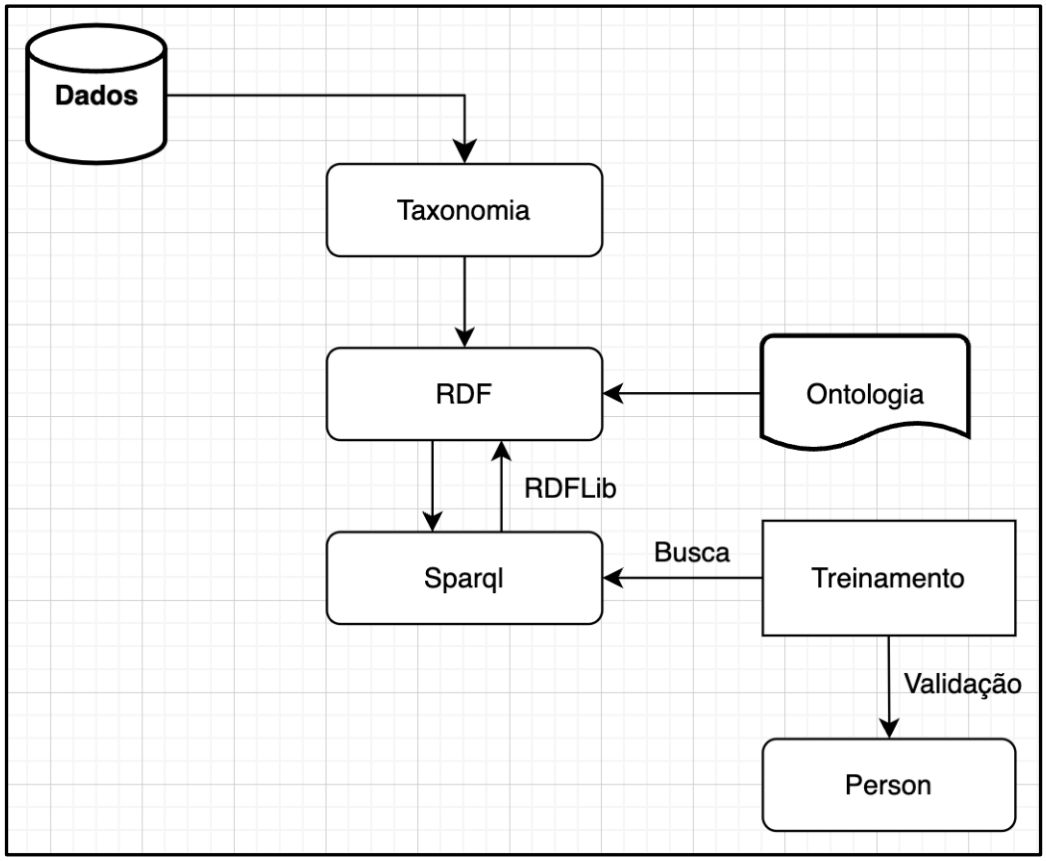
5.9 Considerações sobre esta seção

A convergência desses elementos não apenas enriquece a compreensão no contexto da segurança pública, mas também abre novas abordagens baseadas em dados para a prevenção de crimes. Ao aplicar métodos rigorosos d Ciência da Informação para padronizar, mapear e interconectar dados de diversas fontes, cria-se um ecossistema de informação mais robusto e confiável. Este ecossistema não apenas melhora a eficiência das autoridades em tomar decisões baseadas em dados, mas também fornece uma base para a inovação e a implementação de políticas públicas mais eficazes em segurança. Portanto, a interseção de Dados Abertos Conectados com a Ciência da Informação e a Prevenção de Crimes não é apenas promissora, mas essencial para a construção de um futuro mais seguro e informado.

6 DESENVOLVIMENTO DA APLICAÇÃO PARA A PREDIÇÃO DE CRIMES

Este capítulo apresenta o fluxo utilizado no desenvolvimento do sistema de predição de crimes, adotado pelo departamento de desenvolvimento de sistemas do décimo oitavo batalhão de Polícia Militar do estado de São Paulo, sediado no município de Presidente Prudente. A aplicação desenvolvida conecta dados abertos, combinados com uma ontologia projetada para classificar e interpretar esses dados de maneira eficiente e acurada. A ontologia foi enriquecida com taxonomia especializada, contribuindo significativamente para o treinamento de algoritmos de *Machine Learning* para predizer ocorrências. Este processo é descrito passo a passo, desde a coleta inicial dos dados até a aplicação prática na prevenção de crimes, conforme ilustra a figura 16.

Figura 16 – Fluxo do desenvolvimento do sistema



Fonte: o próprio autor (2024).

Esse processo inclui contribuições significativas da equipe de sistemas da polícia militar, evidenciando uma colaboração interdisciplinar na gestão de dados abertos. Para um

entendimento aprofundado sobre dados abertos, incluindo definições e relevância, recomenda-se a consulta ao Capítulo 3, que oferece um panorama detalhado sobre o tema.

6.1 Taxonomia de dados abertos em segurança pública

O fluxo para o processo de coleta de dados abertos em segurança pública envolveu a agregação sistemática de informações de fontes heterogêneas, como dados de ocorrências, dados climáticos e sazonais. Estes últimos incluíram eventos festivos e religiosos, períodos de safra e corte de cana de açúcar, saída de presos em feriados e datas comemorativas e variações no tráfego de automóveis, reconhecendo a influência significativa destes fatores na segurança pública. Utilizou-se, também, a biblioteca Gregorian, que apresenta todos os feriados nacionais, permitindo uma análise mais precisa da incidência de eventos e comportamentos em datas específicas.

Estas informações, após coletadas, foram processadas e padronizadas para garantir a consistência e a precisão dos dados. A criação de uma taxonomia, como instrumento de representação do conhecimento (AGANETTE, ALVARENGA, SOUZA, 2010, p. 77), foi necessária pois permitiu a classificação desses dados em categorias relevantes, facilitando a análise e sua organização. A compreensão eficaz da criminalidade é vital para uma resposta efetiva da segurança pública. A Figura 17 apresenta uma taxonomia genérica de crimes, metodicamente distribuída em categorias e subcategorias, que serve como um mapa para navegar pela complexidade dos delitos (Alves, Campos, 2014). Esta tabela categoriza os crimes em cinco grandes grupos, com cada grupo dividido em tipos específicos de crimes e seus respectivos subtipos.

Esta organização facilita a compreensão das diferentes naturezas dos crimes e oferece uma estrutura para a análise e a tomada de decisões no âmbito da segurança pública. Este arranjo sistemático não apenas refina a maneira como os dados são examinados e interpretados, mas também aprimora a coordenação e eficácia das medidas preventivas e estratégicas adotadas pelas autoridades.

Figura 17 – Modelo de Taxonomia genérica de crimes

Categoria de Crime	Tipo de Crime	Subtipo de Crime
Crimes contra a Pessoa	Homicídio	Doloso, Culposo, Infanticídio, Eutanásia
Crimes contra a Pessoa	Lesão Corporal	Dolosa, Culposa, Violência Doméstica
Crimes contra a Pessoa	Roubo	A mão armada, À pessoa, De veículo
Crimes contra a Pessoa	Sequestro	Com fins lucrativos, De menor, Relâmpago
Crimes contra o Patrimônio	Furto	Simples, Qualificado, De veículo
Crimes contra o Patrimônio	Roubo	A mão armada, À pessoa, De veículo
Crimes contra o Patrimônio	Estupro	De vulnerável, Com violência, Coletivo
Crimes contra o Patrimônio	Dano	Qualificado, Culposo, Ao patrimônio público
Crimes Cibernéticos	Ataques de Malware	Vírus, Spyware, Ransomware
Crimes Cibernéticos	Phishing	Simples, Spear phishing, Whaling
Crimes Cibernéticos	Invasão de Sistemas	Hacking, Cracking, Ataques de negação de serviço
Crimes Ambientais	Poluição	Do ar, Da água, Do solo
Crimes Ambientais	Desmatamento	Legal, Ilegal, Queimadas
Crimes Ambientais	Tráfico de Animais Silvestres	De animais exóticos, Em extinção, Para fins científicos
Crimes de Corrupção	Corrupção Passiva	Concussão, Peculato, Corrupção própria
Crimes de Corrupção	Corrupção Ativa	Oferecimento de vantagem indevida, Tráfico de influência, Lavagem de dinheiro

Fonte: o próprio autor (2024)

A Tabela 1 apresenta o processo de classificação dos dados, formando uma taxonomia que atendeu as necessidades para o desenvolvimento da tecnologia. A coluna “Furto” apresentam a classificação dos termos relevantes de acordo com cada Classe. Esta tabela representa a estrutura de taxonomia para a definição de crimes, proporcionando uma visão detalhada das categorias e subcategorias de furtos, junto com informações complementares que auxiliam na identificação e classificação das ocorrências no contexto da segurança pública.

Tabela 1 – Modelo de Taxonomia utilizado na aplicação

Classe	Sub-Classe	Furto	Complemento
Crime	Natureza	Simples, Qualificado	Celular, bicicleta, automóvel, motocicleta, roupas, jóias, gado, dinheiro, documentos
Crime	Histórico	Primeira Ocorrência, Reincidente	Tipo de furto, objetos subtraídos, *modus operandi,* local furto
Crime	Categoria	Contra o patrimônio	Descrição do patrimônio, local patrimônio
Acusado	-----	Identificado, não identificado, maior de idade, menor de idade	Identificação acusado, cpf acusado, RG acusado, endereço acusado, bairro acusado, telefone acusado
Vítima	-----	Individual, Corporativa	Identificação vítima, cpf vítima, cnpj vítima, RG vítima, endereço vítima, bairro vítima, telefone vítima
Local	-----	Urbano, Rural, Interior, Exterior	Logradouro, número, cep, rua, bairro, viela

Fonte: o próprio autor (2024).

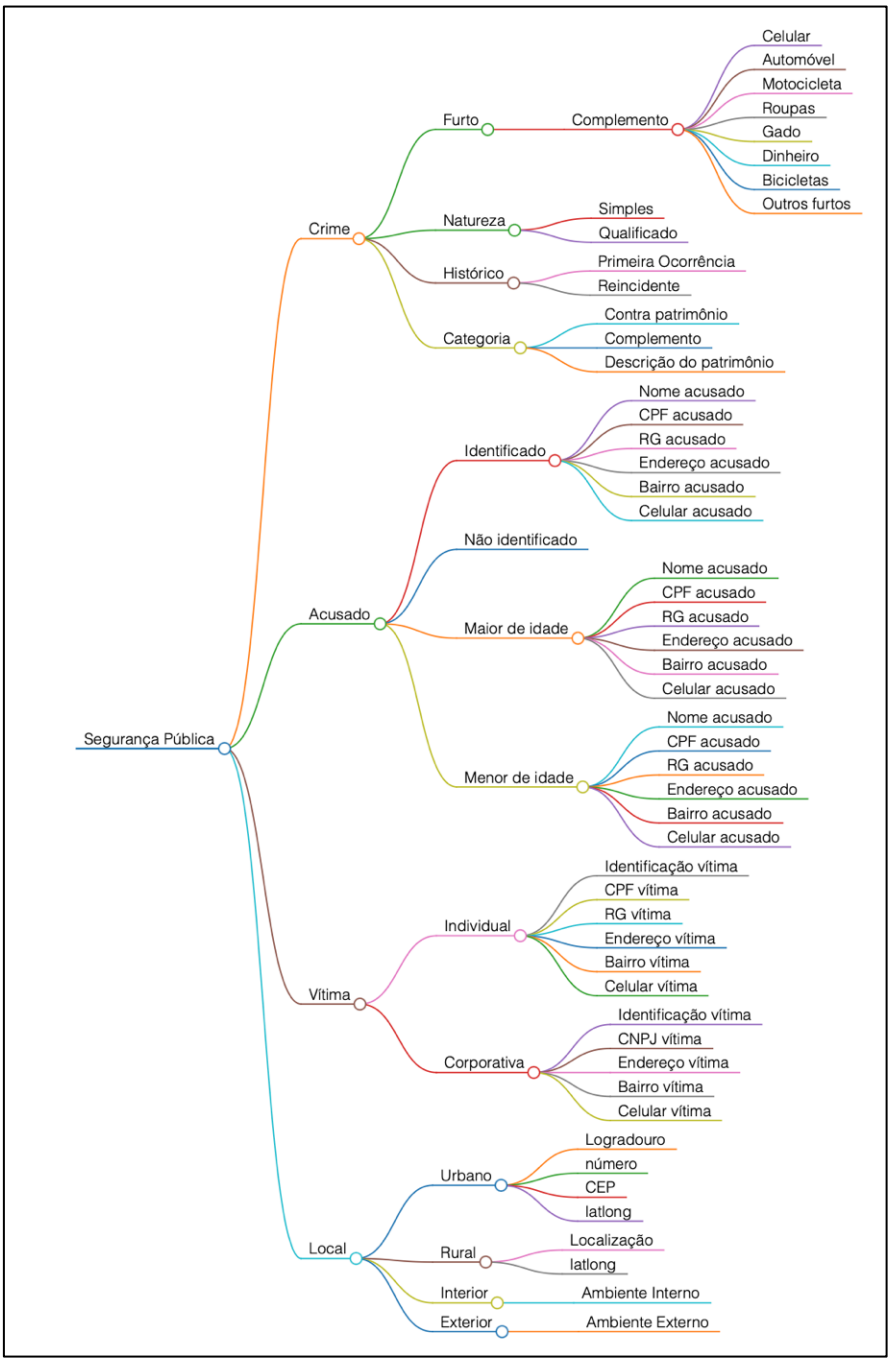
A Tabela 1 serviu como uma ferramenta para estruturar e simplificar o entendimento de dados complexos relacionados a crimes. A taxonomia, ou seja, o sistema de classificação utilizado, divide as informações em categorias e subcategorias claras que correspondem a aspectos específicos dos crimes de furto e violência doméstica, entretanto, mais naturezas de ocorrências poderão ser inseridas nesse modelo.

Na coluna "Furto", os dados são organizados de forma a distinguir os tipos de furto (Simples ou Qualificado), a frequência com que ocorrem (Primeira ocorrência ou Reincidente), e a categoria (Crimes contra o Patrimônio). Isso ajuda a identificar rapidamente a gravidade do crime, a vulnerabilidade e a predição de futuros crimes.

A organização dos dados dessa maneira facilita o acesso à informação por parte das autoridades e, também, permite uma análise mais eficiente que pode ser compartilhada de maneira compreensível, mesmo para aqueles que não são especialistas na área. Isso é fundamental para a tomada de decisões estratégicas, a alocação de recursos de segurança como policiais e viaturas próximos a incidência do crime e o desenvolvimento de estratégias de prevenção e intervenção.

A Figura 18 ilustra a taxonomia apresentada na tabela 1 com uma visão mais aprimorada de acordo com Silveira et. al. (2021).

Figura 18 – Taxonomia em formato de mapa mental



Fonte: o próprio autor (2024).

6.2 Benefícios do uso de taxonomia na segurança pública

Em qualquer cidade ou comunidade, a segurança de seus cidadãos é uma prioridade. Para manter essa segurança, uma grande quantidade de informações precisa ser coletada, desde detalhes de pequenos furtos até incidentes mais graves como casos de violência doméstica.

A taxonomia organiza as informações de segurança pública de maneira em que sua recuperação seja mais eficiente. Ela categoriza cada informação de forma lógica e intuitiva, permitindo que os policiais e analistas encontrem rapidamente o que precisam.

Quando os dados são mais bem organizados desde o início, com cada crime categorizado corretamente, os policiais no atendimento das ocorrências e os analistas no serviço administrativo podem registrar e recuperar informações sem confusão ou perda de tempo. Isso significa que se um oficial está procurando por padrões em furtos em uma área específica, ele pode facilmente recuperar todos os casos relevantes e ver o quadro completo. Da mesma forma, se uma equipe está tentando prever onde a violência doméstica pode ocorrer, eles podem analisar os dados passados para ajudar a prevenir futuros incidentes.

O uso de taxonomia também melhora a comunicação, uma vez que, todos os envolvidos na segurança pública usam o mesmo sistema de classificação, eles podem compartilhar informações rapidamente e com clareza. Isso é crucial em situações urgentes onde a velocidade é essencial para proteger as pessoas.

Os dados, quando organizados, não ajudam apenas na resposta a crimes, mas também na prevenção. Eles permitem que os formuladores de políticas públicas de segurança observem onde os problemas estão se concentrando e criem programas direcionados para essas questões. Com uma boa taxonomia, a segurança pública se torna mais que uma força de reação; ela se torna uma ferramenta proativa para construir uma comunidade mais segura e tranquila.

6.3 Geração do Modelo RDF

A ontologia, uma contribuição chave deste projeto, serve como o alicerce para a construção de um modelo RDF, vide anexo. Este modelo permite representar os dados classificados em um formato que é tanto rico semanticamente quanto flexível, facilitando a realização de consultas complexas e a extração de informações relevantes.

A construção do modelo RDF, apoiada pela ontologia desenvolvida, representa um marco fundamental deste projeto. A ontologia, meticulosamente desenhada e implementada, forneceu a base essencial para representar dados complexos e multifacetados de maneira estruturada e semântica. Este modelo semântico se mostrou uma ferramenta poderosa e flexível, capaz de capturar a riqueza e a profundidade dos dados de segurança pública e climáticos (STAAB, 2004).

Com o modelo RDF em mãos, a equipe de desenvolvimento da Polícia Militar recebeu um recurso valioso, que transforma dados brutos em informações acessíveis e analisáveis. As triplas RDF, que são a espinha dorsal desse modelo, permitem a ligação de conceitos e instâncias de uma forma que reflete com precisão a complexidade do mundo real.

Por exemplo, no modelo RDF, um crime específico é mais do que apenas um ponto de dados isolado; ele está conectado ao local onde ocorreu, ao histórico, aos dados do acusado e da vítima e até mesmo às condições climáticas do momento. Essas conexões são cruciais para o desenvolvimento de análises preditivas e estratégias de prevenção de crimes. Elas permitem que os analistas da polícia não apenas vejam o que aconteceu, mas também compreendam o contexto e as condições sob as quais os crimes ocorrem.

A ontologia, ao alimentar esse modelo RDF, viabilizou consultas complexas usando a linguagem SPARQL. Esse processo permitiu a extração de percepções que antes poderiam permanecer ocultos nos dados. Analistas e desenvolvedores agora podem formular perguntas específicas e receber respostas detalhadas, o que é crucial para a rápida tomada de decisões e para a alocação eficiente de recursos policiais.

O desenvolvimento desta ontologia e a subsequente geração do modelo RDF é uma das maiores contribuições deste projeto, demonstrando o poder da Ciência da Informação aplicada à segurança pública. Este trabalho não só aumenta o entendimento dos padrões de crime e clima, mas também fornece um caminho para a adoção de medidas preventivas mais eficazes, potencialmente salvando vidas e recursos.

O modelo RDF gerado a partir da ontologia criada foi um fator decisivo que habilitou a Polícia Militar a ampliar suas capacidades de análise de dados, promovendo uma segurança pública mais proativa. Este avanço representa um passo significativo em direção a um futuro em que a prevenção de crimes é tão científica e orientada a dados quanto possível.

6.4 Desenvolvimento do Sistemas Preditivo

No cerne do desenvolvimento da aplicação pela Polícia Militar está a utilização da tecnologia SPARQL para realizar consultas na base RDF. Esta base, alimentada pela referida ontologia e uma taxonomia detalhada, permite que os desenvolvedores e analistas analisem grandes conjuntos de dados com precisão e eficiência.

Na etapa crucial do desenvolvimento do sistema pela Polícia Militar, será feito o uso da linguagem de consulta SPARQL. Com a base de dados RDF, a linguagem SPARQL permite aos desenvolvedores e analistas formular consultas complexas e precisas. Esta capacidade de consulta é um aspecto fundamental da aplicação, permitindo uma exploração dos dados relacionados à criminalidade e condições climáticas. Ao utilizar SPARQL, é possível isolar incidentes de furto ocorridos sob condições climáticas específicas, em diferentes períodos do dia, ou em localidades variadas, fornecendo assim um retrato detalhado e contextualizado das ocorrências.

A biblioteca RDFLib, uma ferramenta essencial para o desenvolvimento da aplicação, foi utilizada para facilitar a manipulação e o processamento dos dados dentro do modelo RDF. Esta biblioteca, escrita na linguagem Python, oferece funcionalidades extensivas para trabalhar com RDF, permitindo à equipe de desenvolvimento da Polícia Militar ler, escrever, armazenar, consultar e realizar a análise semântica dos dados com eficiência GUPTA (2022).

Com o uso da biblioteca RDFLib, o processo de interligação dos dados torna-se mais ágil e menos suscetível a erros, garantindo que as consultas SPARQL sejam executadas com precisão e os resultados sejam interpretados corretamente. A adoção desta biblioteca conecta-se com as melhores práticas de engenharia de *software* e com a implementação de soluções confiáveis e escaláveis para o desafio da segurança pública GUPTA (2022).

A Figura 19 ilustra a chamada da biblioteca RDFLib para realizar a análise semântica dos dados.

Figura 19 – Chamada da Biblioteca RDFLib

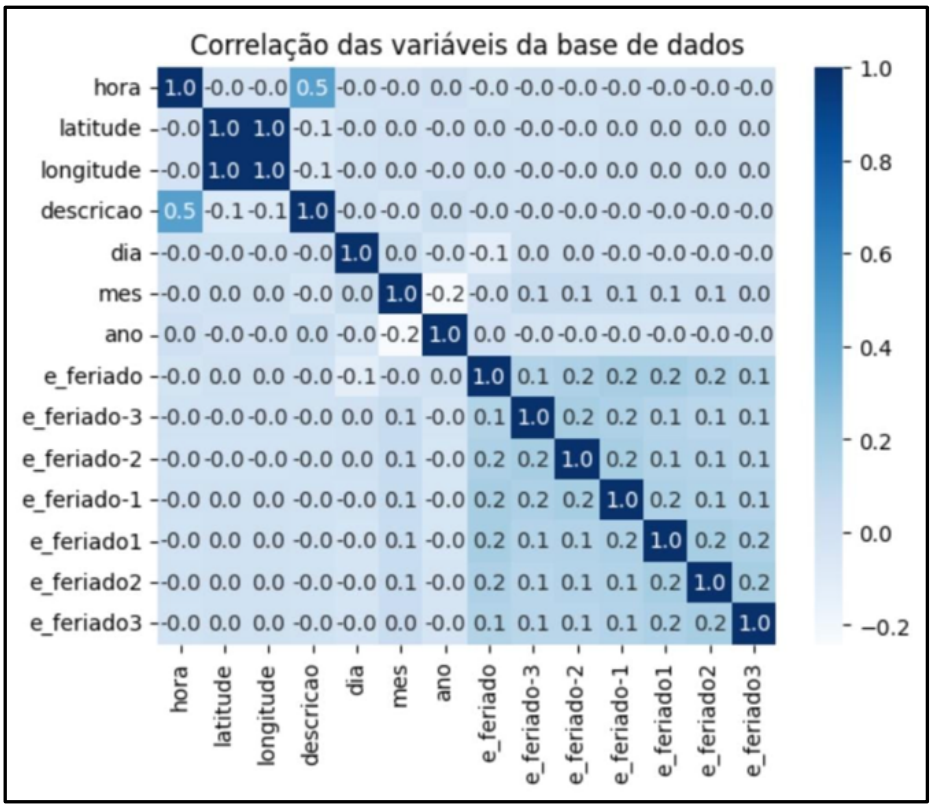
```
import pandas as pd
from rdflib import Graph, Namespace, Literal
from datetime import datetime
import requests

# Dados de ocorrências de janeiro
dados_JAN = [
  {"data_hora_fato": "2/1/2023 21:50", "latitude": "-23.524781143", "longitude": "-46.630798931",
"descricao": "FURTO"},
  {"data_hora_fato": "3/1/2023 01:50", "latitude": "-23.525287737527", "longitude": "-
46.629224769478", "descricao": "ROUBO"},
  {"data_hora_fato": "3/1/2023 03:31", "latitude": "-23.520465870702", "longitude": "-
46.635926993262", "descricao": "ROUBO"},
  # Adicione mais dados de ocorrências se necessário
]
```

Fonte: o próprio autor (2024).

Paralelamente às consultas, a aplicação beneficia-se do aprendizado de máquina para aprimorar a análise dos dados. O aprendizado de máquina, alimentado pelas descobertas obtidas através de SPARQL, não é apenas um repositório de informações, mas um sistema dinâmico que aprende, adapta-se e evolui. Com cada consulta SPARQL, os algoritmos de aprendizado de máquina recebem novos dados para treinamento, melhorando a sua capacidade de identificar padrões ocultos, prever tendências de criminalidade e, em última análise, propor ações preventivas eficazes.

O processo contínuo de treinamento e validação, apoiado pela robusta estrutura da correlação de Pearson, conforme ilustra a Figura 20, garante que o sistema não só responda às questões atuais, mas também se antecipe às futuras demandas de segurança pública.

Figura 20 – Resultado obtido com a correlação de Pearson



Fonte: o próprio autor (2024).

A Figura 20 apresenta um mapa de calor que ilustra as correlações de Pearson entre diversas variáveis em uma base de dados, revelando a força e a direção das relações lineares entre elas. O coeficiente de Pearson é uma escolha metodológica quando se busca quantificar o grau de relação entre variáveis que são supostamente influentes na predição de fenômenos complexos, como os padrões criminais. Valores próximos a 1 ou -1 indicam uma forte associação positiva ou negativa, respectivamente, enquanto valores em torno de 0 sugerem uma falta de relação linear. Esta análise preliminar é fundamental para a

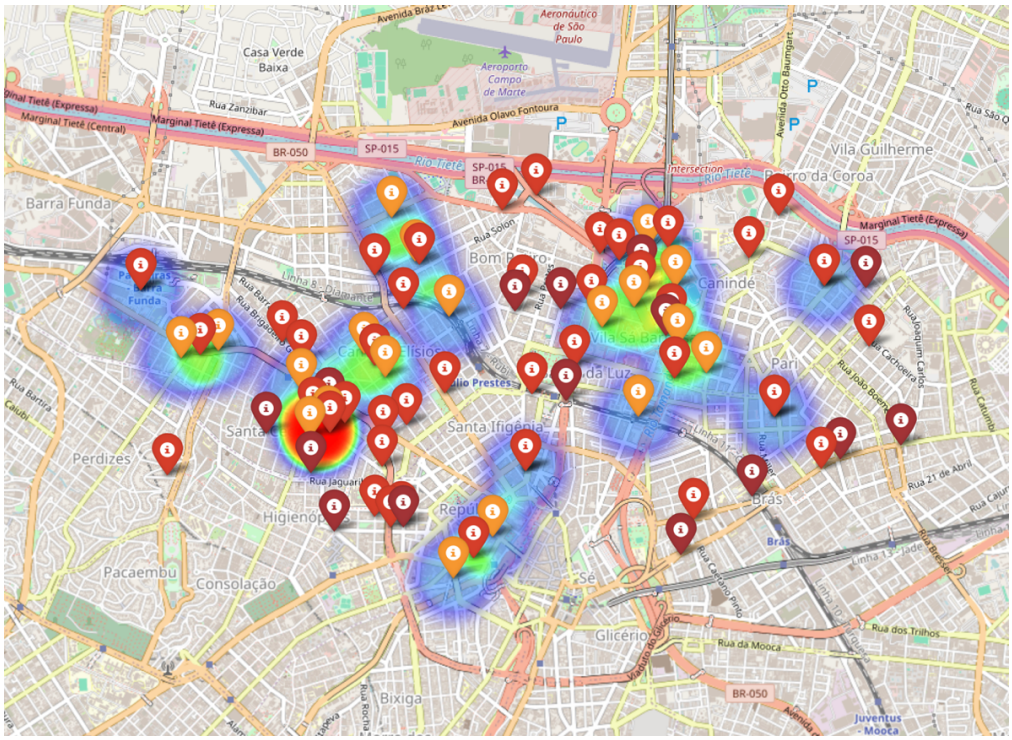
construção de modelos preditivos em aprendizado de máquina, pois orienta quais variáveis merecem atenção e potencial inclusão no modelo.

Optar pelo termo "Predição" reflete a transição de uma análise estatística pura para uma aplicação prática e orientada para a ação. Ao incorporar as descobertas da correlação de Pearson no desenvolvimento de algoritmos de aprendizado de máquina, a pesquisa transcende a mera catalogação de informações para abraçar uma abordagem proativa que busca não apenas compreender, mas antever e mitigar proativamente as incidências de criminalidade. A precisão da predição depende da qualidade e relevância dos dados fornecidos ao sistema, enfatizando a importância de um ciclo contínuo de treinamento e validação. Esse processo assegura que o modelo preditivo se mantenha atualizado e alinhado com as tendências emergentes, capacitando-o a ser uma ferramenta valiosa na prevenção do crime e no fortalecimento das estratégias de segurança pública.

6.5 Interfaces da aplicação

Nesta seção, apresentam-se algumas interfaces da aplicação desenvolvida para a predição de crimes, fundamentada em uma taxonomia de dados abertos em segurança pública e na geração de um modelo RDF. A Figura 21 ilustra a interface referente ao mapa de calor integrado com predições de ocorrências criminais.

Figura 21 – Mapa de calor de ocorrências e suas predições



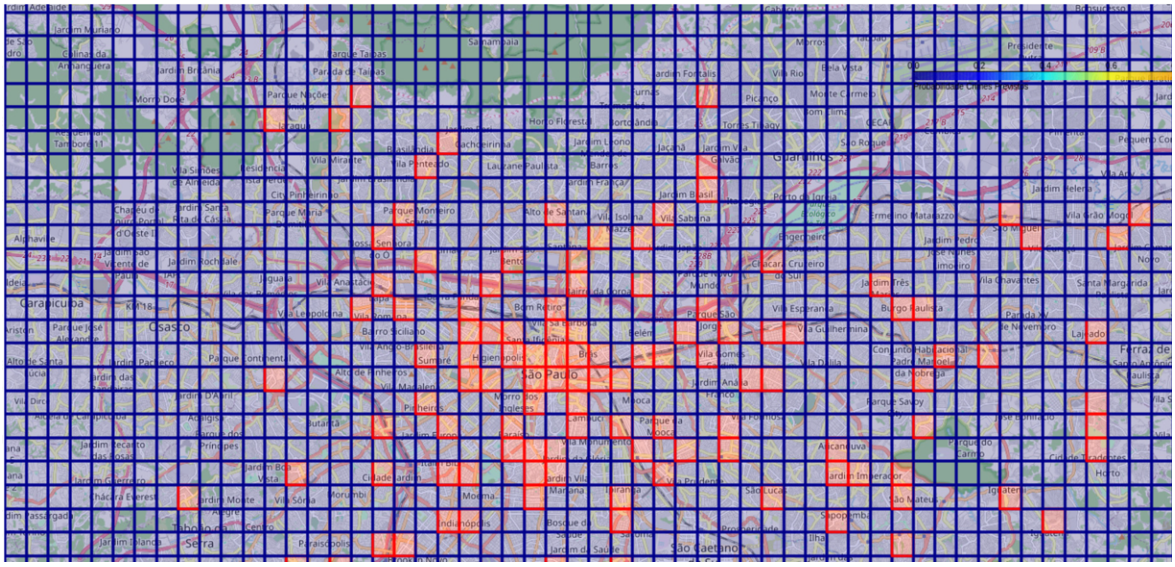
Fonte: o próprio autor (2024).

A imagem da Figura 21 apresenta uma interface gráfica que mostra um mapa de calor sobreposto a um mapa geográfico em uma área de teste. Os diversos matizes, que variam do azul ao vermelho, indicam a densidade de incidentes criminais ocorridos anteriormente. Marcadores de localização equipados com ícones descritivos são incorporados para assinalar locais específicos onde se prevê a ocorrência de crimes. Esse mapa de calor transcende sua função de ferramenta visual informativa, atuando também como um instrumento preditivo que direciona os esforços e os recursos da Polícia Militar de forma estratégica.

A funcionalidade deste mapa de calor oferece uma perspectiva dinâmica, possibilitando aos usuários examinar a distribuição espacial dos incidentes e identificar padrões e focos de atividade criminosa. Ele serve como um alicerce para o planejamento operacional e estratégico, fundamentando a tomada de decisão baseada em dados concretos e reforçando o apoio a decisões estratégicas. Além disso, promove uma alocação de recursos mais eficaz, melhorando a distribuição dos esforços policiais e encaminhando-os para regiões com maior risco de atividades ilícitas.

A Figura 22 ilustra um mapa em formato de *grid* demonstrando os pontos onde ocorreram crimes, estes representados na cor laranja.

Figura 22 – Mapa em grade com localização dos crimes ocorridos



Fonte: o próprio autor (2024).

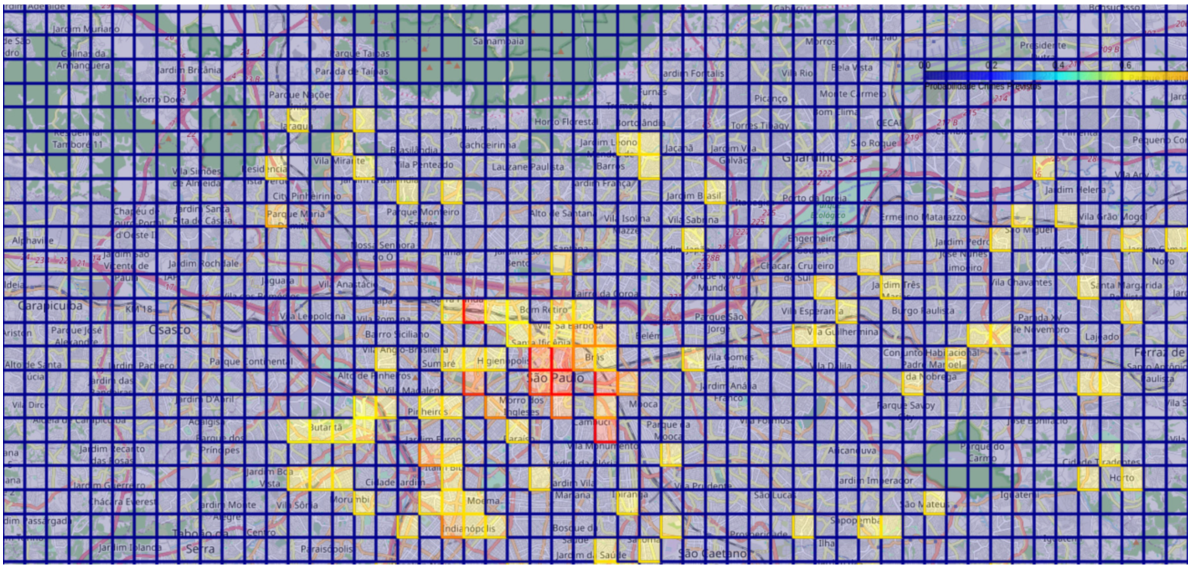
A Figura 22 apresenta um mapa organizado em formato de grade, evidenciando locais específicos de atividades criminosas, cada um marcado por pontos na cor laranja. Este método de representação facilita a visualização geográfica dos incidentes, bem como

destaca a frequência e a distribuição dos crimes dentro de uma área delimitada, permitindo uma análise detalhada dos padrões de criminalidade.

O mapa é parte integrante de uma estratégia de vigilância, operado pelo sistema de predição da Polícia Militar. Por meio da coleta e análise de dados históricos e atuais, o sistema busca identificar potenciais zonas de risco, otimizando assim a alocação de recursos e a prontidão das forças de segurança. Esta abordagem proativa visa não apenas responder aos crimes após sua ocorrência, mas também prevenir sua manifestação, garantindo uma maior segurança pública.

A Figura 23 apresenta os pontos onde, por meio do sistema de predição, os crimes poderão ocorrer. Este sistema baseou-se no mapa apresentado pela Figura 22.

Figura 23 – Mapa em grade com a predição criminal



Fonte: o próprio autor (2024).

A Figura 23 representa uma inovação na estratégia de combate ao crime, exibindo um mapa organizado em grade que sinaliza, através de pontos em laranja, as áreas onde crimes são previstos ocorrer. Esta previsão é fruto do sistema de análise preditiva, que, fundamentado na conexão de dados abertos e padrões de criminalidade anteriores ilustrados pela figura 22, emprega algoritmos e técnicas de aprendizado de máquina para antever locais com alta probabilidade de atividades ilícitas. Os locais marcados em laranja servem como alertas visuais, direcionando as autoridades a adotarem medidas proativas de segurança em regiões identificadas como potenciais pontos de risco.

Este mapa, portanto, não apenas demonstra a aplicação efetiva da tecnologia na prevenção do crime, mas também marca uma transição para uma abordagem mais preventiva por parte das forças de segurança. Ao destacar os futuros pontos de crime em

laranja, o sistema permite que a polícia se antecipe a possíveis incidentes, otimizando a alocação de seus recursos para focar nas áreas de maior necessidade. Assim, a iniciativa busca não só prever e prevenir a ocorrência de crimes, mas também fortalecer a sensação de segurança na comunidade, contribuindo para um ambiente mais seguro e tranquilo para todos.

A integração da ontologia, taxonomia e a subsequente geração do modelo RDF, aliada às capacidades de consulta e aprendizado de máquina, constitui uma das contribuições mais significativas deste projeto. Por meio desta aplicação, a Polícia Militar está redefinindo a maneira como os dados abertos são aproveitados no contexto da segurança pública. Este avanço tecnológico não só melhora as operações atuais, mas também abre novos horizontes para métodos preditivos e preventivos, alavancando a segurança pública em direção a um futuro em que as decisões são informadas por dados confiáveis e análises profundas.

7 CONCLUSÃO

Neste trabalho, exploraram-se os detalhes sobre o papel dos dados abertos conectados e suas implicações no contexto da segurança pública. Ficou evidente que essa abordagem pode ser uma promissora fronteira de pesquisa, que pode contribuir significativamente para novos estudos no que se refere à predição e prevenção de crimes nas suas mais variadas modalidades.

A análise detalhada dos conceitos fundamentais de democracia e transparência nos órgãos públicos de segurança e, principalmente, na tecnologia demonstrou que a democratização dessas instituições é essencial para o desenvolvimento de uma sociedade mais segura. Outrossim, o uso de tecnologias semânticas para analisar e mapear os dados abertos surgiu como uma abordagem inovadora, ampliando a compreensão dos fenômenos relacionados à segurança.

Do ponto de vista da Ciência da Informação, esta pesquisa contribui para o entendimento de como as estruturas de dados, a taxonomia e as ontologias podem ser aplicadas para resolver questões práticas e complexas em setores sociais críticos, como a segurança pública. A organização e a semântica dos dados não são apenas questões técnicas, mas também fundamentais para a construção de uma estrutura informativa que permita a interpretação adequada e a ação eficaz baseada em dados.

Assim, é evidente que este trabalho, ainda prematuro, se apresenta como um paradigma interposto no limiar de uma nova forma de conexão de dados abertos na segurança pública.

A aplicação das abordagens discutidas neste projeto pode abrir novas perspectivas para tomadas de decisões assertivas no que se refere à prevenção criminal, à otimização na gestão de recursos públicos e à promoção de comunidades mais seguras.

De modo geral, os dados abertos conectados têm o potencial de revolucionar as ações de segurança pública, promovidas com base em dados conectados com o uso de tecnologias semânticas, uma vez que estas são fundamentais nessa transformação. À medida que o trabalho evoluir, deve-se manter o comprometimento constante com a pesquisa, o desenvolvimento e a implementação prática dessas ideias com a finalidade de construir um futuro mais justo, democrático e seguro para todos.

A robustez do sistema proposto foi meticulosamente testada quanto à sua acurácia, utilizando a correlação de Pearson como um dos indicadores chave de validação. A figura

20 foi instrumental nesse processo, evidenciando associações positivas fortes, especialmente onde se observou um coeficiente de correlação próximo de 1. Esse valor indicativo de uma relação linear positiva perfeita entre determinadas variáveis foi fundamental para afirmar a confiabilidade do modelo preditivo. A escolha dessa metodologia estatística permitiu uma validação rigorosa do sistema, garantindo que as variáveis selecionadas para o modelo tivessem a maior influência possível sobre a variável alvo, que é a previsão de ocorrências criminais. Assim, foi possível delinear um perfil preciso de predição que se alinha estreitamente com as tendências observadas nos dados de segurança pública.

Este trabalho representou um marco inovador na intersecção entre a tecnologia de dados e a segurança pública, estabelecendo um precedente pioneiro para futuras investigações e aplicações no campo. A abordagem centrada na utilização de dados abertos conectados, aliada à aplicação de tecnologias semânticas, destacou-se não apenas pela sua originalidade, mas também pelo seu potencial transformador na maneira como as instituições de segurança pública operam e se engajam com a sociedade.

A demonstração de interesse por parte da Secretaria de Segurança Pública do Estado de São Paulo no sistema desenvolvido valida e reforça a relevância e o impacto potencial deste trabalho. Tal interesse por uma entidade governamental não apenas sublinha a aplicabilidade prática das pesquisas e tecnologias propostas, mas também sinaliza um reconhecimento crescente da importância da inovação tecnológica no aprimoramento das operações de segurança pública e na promoção da transparência e da democracia.

A inovação trazida por este estudo reside na sua capacidade de transcender as abordagens tradicionais, propondo uma nova forma de entender e lidar com os dados no contexto da segurança pública. Por meio da estruturação, organização e análise semântica de dados abertos, abre-se um leque de possibilidades para a prevenção e combate ao crime de maneira mais eficaz. Isso não apenas contribui para uma gestão mais eficiente dos recursos públicos, como também promove uma sociedade mais segura.

O caráter pioneiro deste trabalho também se reflete na sua contribuição para a literatura acadêmica e para a prática profissional na área de Ciência da Informação, especialmente no que tange ao uso de ontologias, taxonomias e estruturas de dados na solução de problemas complexos e práticos. A pesquisa estabelece um modelo para futuros trabalhos que busquem integrar dados abertos e tecnologias semânticas para resolver desafios sociais significativos.

Este trabalho não apenas avança o conhecimento acadêmico e prático na área de segurança pública, mas também sinaliza um caminho promissor para a integração de tecnologias avançadas no setor. A colaboração entre o setor público e os pesquisadores, exemplificada pelo interesse da Secretaria de Segurança Pública, é crucial para a implementação bem-sucedida dessas inovações. À medida que o projeto avança para a fase de implementação, permanece o compromisso com a pesquisa contínua, desenvolvimento e aplicação prática dessas ideias inovadoras, visando um futuro mais seguro, transparente e democrático.

REFERÊNCIAS

5 ★ OPEN DATA. 2012. Disponível em: https://5stardata.info/en/. Acesso em: 9 maio 2023.

AGANETTE, Elisangela; ALVARENGA, Lídia; SOUZA, Renato Rocha. Elementos constitutivos do conceito de taxonomia. Informação & Sociedade, v. 20, n. 3, 2010.

AGUNE, Roberto Meizi; GREGÓRIO FILHO, Álvaro Santos; BOLLIGER, Sergio Pinto. Governo aberto SP: disponibilização de bases de dados e informações em formato aberto. *In*: CONGRESSO CONSAD DE GESTÃO PÚBLICA, 3. Brasília, DF. **Anais eletrônicos [...].** Brasília, DF, 2010. Disponível em: https://www.consad.org.br/eventos/congressos/iii-congresso-consad-de-gestao-publica-brasilia-df. Acesso em: 14 jan. 2024.

ALEIXO, Diana Vilas Boas Souto. **O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil**. Orientador: Ricardo César Gonçalves Sant'Ana. 2020. 260 f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília, São Paulo, 2020.

ALMEIDA, Mauricio Barcellos. Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares. **Ciência da Informação,** v. 31, n. 2, p. 5-13, 2002. Disponível em: https://revista.ibict.br/ciinf/article/view/955/992. Acesso em: 15 jan. 2024.

ALVES, E. M., & de CAMPOS, M. L. M. (2014). Ontologies and Linked Data for the Semantic Web. Synthesis Lectures on the Semantic Web: Theory and Technology, 6(4), 1-154.

ALVES, Verlene Sousa de Castro. **Estresse laboral e suas consequências psicossociais em policiais militares no exercício de suas funções**. Orientador: Walberto Silva dos Santos. 2018. 89 f. Dissertação (Mestrado em Psicologia) - Programa de Pós-Graduação em Psicologia, Universidade Federal do Ceará, Fortaleza, 2018.

ANDERSON, Craig A. Heat and violence. **Current Directions in Psychological Science,** v. 10, n.1, p. 33–38, 2001.

ANTONIAZZI, Francesco. **Semantic driven agent programming**. 2020. 132 fl. Tese (Dottorato di Ricerca in Computer Science and Engineering) - Alma Mater Studiorum, Università di Bologna, 2020. Disponível em: https://amsdottorato.unibo.it/9197/1/tesi.pdf. Acesso em: 15 jan. 2024.

ARAÚJO, A. F. A Lei de Acesso à Informação como instrumento de gestão do- cumental no contexto brasileiro. **Informação & Sociedade: Estudos**, v. 27, n. 1, p. 1-14, 2017.

ARAÚJO, Narallynne Maciel de. **Dados abertos do governo brasileiro**: entendendo as perspectivas de fornecedores de dados e desenvolvedores de aplicações ao cidadão. 2017. Disponível em: https://www.academia.edu/66198337/Dados_abertos_do_governo_brasileiro_entendendo_as_perspectivas_de_fornecedores_de_dados_e_desenvolvedores_de_aplica%C3%A7%C3%B5es_ao_cidad%C3%A3o. Acesso em: 15 jan. 2024.

ATAÍDE, Mara Célia Ferreira *et al*. A divulgação de relatórios de atividades das ouvidorias de polícia no brasil: uma análise da transparência do serviço público à sociedade. **Revista de Gestão Pública: práticas e desafios**, v. 12, n. 2, 2020. Disponível em: https://periodicos.ufpe.br/revistas/gestaopublica/article/viewFile/242897/36945. Acesso em: 5 fev. 2024.

ATEMEZING, Gislain *et al*. Transforming meteorological data into linked data. **Semantic Web**, v. 4, n. 3, p. 285–290, 2013. Disponível em: https://www.semantic-web-journal.net/sites/default/files/swj281_0.pdf. Acesso em: 15 jan. 2024.

AUER, Sören *et al*. DBpedia: A nucleus for a Web of open data. **Lecture Notes in Computer Science**, v. 4825, 2007.

AVELAR, Cátia Fabíola Parreira de; ROCHA, Thiago Augusto Hernandes; CRUZ, Flávia Juliesse Soares. Mineração de Dados: uma revisão da literatura em Administração. **Revista Vianna Sapiens**, v. 8, n. 2, jul./dez. 2017.

BANDEIRA, Judson *et al*. Dados Abertos Conectados. *In*: JORNADA DE ATUALIZAÇÃO EM TECNOLOGIA DA INFORMAÇÃO, 3. **Anais [...].** São Paulo: EDUFAL, 2014.

BANERJEE, Debayan *et al*. Modern Baselines for SPARQL Semantic Parsing. *In*: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 45, July 2022. **Proceedings […].** Madrid: ACM, 2022. p. 2260–2265.

BARROS, S. A. **Polícia Militar e sociedade civil**: uma relação conflituosa. Monografia (Especialização em Políticas Públicas de Segurança Pública) - Universidade do Estado do Rio de Janeiro, 2011.

BASSO, Rafael; SCHMIDT, Daniela; SANTOS, Cassia Trojahn dos; VIEIRA, Renata. Alinhamento entre ontologias de topo e de dominio usando WordNet. *In*: BRAZILIAN ONTOLOGY RESEARCH SEMINAR, 10, August 2017, Brasilia. **Proceedings [...].** Brasília, DF. 2017. p.9-20 ⟨hal-02089251⟩ Disponível em: https://ceur-ws.org/Vol-1908/paper1.pdf. Acesso em: 19 maio 2023.

BEAUCHAMP, Tom L.; MCCURDY, Howard. **Data Governance**: how to design, deploy and sustain an effective data governance program. New Jersey: John Wiley & Sons, 2016.

BECKER, Gary S. Crime and Punishment: an economic approach. **The Journal of Political Economy,** v. 76, n. 2, p. 169-217, 1968.

BERNERS-LEE, Tim. **Putting government data online**. 2009. Disponível em: http://www.w3.org/DesignIssues/GovData.html. Acesso em: 15 jan. 2024.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**. May 2001. Disponível em: http://www2.ic.uff.br/~bazilio/cursos/sistweb/material/Barners-Lee-Scientific-American-May-2001.pdf. Acesso em: 15 jan. 2024.

BERRY, Michael J. A.; LINOFF, Gordon S. **Data Mining techniques**: for marketing, sales, and customer support. New York: John Wiley & Sons. 1997.

BIZER, Christian; CYGANIAK, Richard. **D2R Server** - Publishing Relational Databases on the Semantic Web. 2006. Disponível em: http://richard.cyganiak.de/2008/papers/d2r-server-iswc2006.pdf. Acesso em: 19 jan. 2024.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. Linked Data: the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1-22. July 2009. Disponível em: https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf. Acesso em: 15 jan. 2024.

BIZER, Christian; HEATH, Tom, IDEHEN, Kingsley; BERNERS-LEE, Tim. Linked data on the web (LDOW2008). *In*: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB 2008, 17, April 2008. **Proceedings [...].** New York: ACM, 2008. p. 1265–1266.

BLAKESLEE, David S.; FISHMAN, Ram. **Weather Shocks, Agriculture, and Crime**: Evidence from India. July 2015. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2428249. Acesso em: 15 jan. 2024.

BOBBIO, Norberto. **Liberalismo e democracia.** Tradução de Marco Aurélio Nogueira. São Paulo: Brasiliense, 2005.

BONSÓN, Enrique; TORRES, Lourdes; ROYO, Sonia; FLORES, Francisco. Local e-government 2.0: social media and corporate transparency in municipalities. **Government Information Quarterly**, v. 29, n. 2, p. 123-132, 2012.

BORGMAN, Christine; WALLIS, Jillian Claire; ENYEDY, Noel. Little Science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. **International Journal on Digital Libraries**, v. 7, n. 1-2, p.17-30, Oct. 2007.

BOYD, Danah; CRAWFORD, Kate. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon**. Information, Communication and Society**, v. 15, n. 5, p. 662–679, June 2012.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil**. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 15 jan. 2024.

BRASIL. **Decreto nº 8.777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo federal. Brasília, DF: Presidência da República, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm Acesso em: 1 fev. 2024.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. Brasília, DF: Presidência da República, 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 2 jan. 2023.

BRASIL. **Lei nº 13.709, de14 de agosto de 2018**. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 1 fev. 2024.

BRASIL **Lei nº 13.853 de 14 de agosto de 2019**. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Brasília, DF: Presidência da República, 2019. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13853.htm. Acesso em: 28 dez. 2022.

BRASIL. Ministério da Justiça. Secretaria Nacional de Segurança Pública. **Polícia Cidadã**: conceitos, princípios e ações. Brasília, DF: MJ, 2009.

BRAY. Tim. **The JavaScript Object Notation (JSON) Data Interchange Format**. RFC 7159. 2014. Disponível em: https://datatracker.ietf.org/doc/rfc7159/. Acesso em: 15 jan. 2024.

BRAY, Tim *et al*. **Extensible Markup Language (XML) 1.1.** (2nd Edition). 2006. Disponível em: http://www. w3.org/TR/2006/REC-xml11-20060816http://www.w3.org/TR/xml11Previousversion:http://www.w3.org/TR/2006/PER-xml11-20060614. Acesso em: 17 jan. 2024.

BROWN, Donald E. **Predictive Models for Law Enforcement**: Final Report. 2002. NIJ Grant 984J-CX-KO10. Disponível em: https://www.ojp.gov/pdffiles1/nij/grants/197634.pdf. Acesso em: 15 jan. 2024.

BRÜDERLE, Anna; PETERS, Jörg; ROBERTS, Gareth. **Weather and crime in South Africa**. 2017. Disponível em: https://www.rwi-essen.de/publikationen/wissenschaftlich/ruhr-economic-papers/detail/weather-and-crime-in-south-africa-1164. Acesso em: 15 jan. 2024.

BUCKLAND, Michael. What kind of science can information Science be? **Journal of the American Society for Information Science and Technology**, v. 63, n. 1, p. 1-7, Jan. 2012.

BUREAU OF JUSTICE ASSISTANCE. **Transcript**: Perspectives in Law Enforcement - The Concept of Predictive Policing: An Interview with Chief William Bratton. Entrevistadores: James H. Burch II e Kristina Rose. Entrevistado: William Bratton. Los Angeles, 2009. Podcast. Disponível em: https://bja.ojp.gov/sites/g/files/xyckuh186/files/publications/podcasts/multimedia/transcript/Transcripts_Predictive_508.pdf. Acesso em: 20 jan. 2024.

CALLAHAN, Richard. Governance: The Collision of Politics and Cooperation. **Public Administration Review**, v. 67, n. 2, p. 290-301, 2007.

CALVANESE, Diego *et al*. Ontop: answering SPARQL queries over relational databases. **Semantic Web**, v. 8, n. 3, 2016

CAMPOS, Aline de *et al*. Mineração de Dados Educacionais e Learning Analytics no contexto educacional brasileiro: um mapeamento sistemático. **Informática na Educação: Teoria e Prática**, Porto Alegre, v. 23, n.3, set./dez. 2020. Disponível em: https://seer.ufrgs.br/index.php/InfEducTeoriaPratica/article/view/102618/61035. Acesso em: 15 jan. 2024.

CHAINEY, Spencer; RATCLIFFE, Jerry. **GIS and Crime Mapping**. [S.l]: Wiley, 2005.

CHANG, H. A ciência cidadã e os dados abertos: A promessa de melhorar o acesso à informação. **Dados Abertos**, v. 4, n. 3, p.194-196, 2018.

CHATEAUBRIAND, Oswaldo. A filosofia, a linguagem e o mundo. *In*: BRITO, Adriano Naves de; VALE, Oto Araújo. **Filosofia, linguística, informática**: aspectos da linguagem. Goiânia: Ed. UFG, 1998.

CLEMENTE, Pedro José Lopes. Da ciência policial. **Lusíada: Política Internacional e Segurança**, n. 21/22, p. 103-119, 2021.

CLEMENTE, Pedro José Lopes. Rumos da Segurança em Portugal. **Revista de Direito e Segurança**, ano 1, n. 1, p. 143-163, jan./jun. 2013.

CLIFT, Steven L. **E-Government and Democracy**: representation ,and citizen engagement in the information age. Feb. 2004. Disponível em: http://www.publicus.net/articles/cliftegovdemocracy.pdf. Acesso em: 26 mar. 2023.

COLEMAN, Stephen; BLUMLER, Jay G. **The Internet and Democratic Citizenship**: theory, practice and policy. Cambridge: Cambridge University Press. 2009.

CORREIA, Rodrigo Borges. **Transparência ativa e Open Government Data**: uma proposta para a abertura de dados na Polícia Federal. Orientador: Douglas Dyllon Jeronimo de Macedo. 2021. 112 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2021.

CÔRTES, Sergio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sergio. **Mineração de Dados**: Funcionalidades, Técnicas e Abordagens. Rio de Janeiro: PUCRJ, 2002.

COSTA, A. C. C.; GONÇALVES, J. M. A importância dos dados abertos para um governo transparente. **Observatório**: Revista Brasileira de Ciência da Informação, v.13, n. 2, p. 4-13, 2020. https://doi.org/10.11606/issn.2176-1805.v13i2

COSTA, D. **Doutrina dos Dados Abertos**. São Paulo: Editora Juruá, 2021.

COSTA, Marco Antônio. Segurança Pública. **Revista Núcleo de Criminologia**, Paracatu, v. 7, p. 130-140, nov. 2010. Disponível em: http://www.atenas.edu.br/uniatenas/assets/files/magazines/Revista_Nucleo_Criminologia_07.pdf. Acesso em: 16 jan. 2024.

CRUZ, Jaderson Araújo Gonçalves da. **Mapeamento de bancos de dados para domínios semânticos**. Orientador: Cedric Luiz de Carvalho. 2015. 126 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia, 2015.

CUKIER, Kenneth; MAYER-SCHOENBERGER, Viktor. The rise of big data: How it's changing the way we think about the world. **Foreign Affairs,** v. 92, n. 3, p. 28-40, May/June 2013.

DAMASCENO, Carlos Diogo Nascimento *et al*. **SimCleaner-Sistema de Padronização de Bases de Dados utilizando Funções de Similaridade**. 2011. Disponível em: https://damascenodiego.github.io/assets/pdf/damascenoetal2011_erin.pdf. Acesso em: 16 jan. 2024.

ERDMANN, Michael; DECKER, Stefan. **Ontology-aware XML-Queries**. 2000. Disponível em: https://xml.coverpages.org/erdmann-semantic-xql-webdb00.pdf - Acesso em: 16 jan. 2024.

EUROPEAN UNION. **Improving data publishing by open data portal managers and owners**. 2023. Disponível em: https://op.europa.eu/en/publication-detail/-/publication/5af6fd02-582e-11ee-9220-01aa75ed71a1/language-en/format-PDF/source-304116231 Acesso em: 1 fev. 2024.

FAYYAD, Usama M.; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic; UTHURUSAMY, Ramasamy (ed.) **Advances in Knowledge Discovery and Data Mining**. California, USA: AAAI, MIT, 1996.

FERNANDES, C. S. A participação da sociedade na segurança pública: o caso do Conselho Comunitário de Segurança de Campinas. *In*: SOUZA, R. S.; SOUSA, F. O. (org.). **Segurança pública e direitos humanos**. Rio de Janeiro: Lumen Juris, 2018.

FINNEMORE, Martha; HOLLIS, Duncan B. Beyond Naming and Shaming: Accusations and International Law in Cybersecurity. **European Journal of International Law**, v. 31, n. 3, p. 969–1003, Aug. 2020.

FUKUYAMA, Francis. **The Origins of Political Order**: From Prehuman Times to the French Revolution. New York: Farrar, Straus and Giroux, 2011.

G1. **São Paulo tem maior número de policiais por habitante do país**. Disponível em: https://g1.globo.com/politica/noticia/2015/08/pais-tem-1-pm-para-cada-473-habitantes-diz-ibge.html#:~:text=Segundo%20a%20pesquisa%20do%20IBGE,de%20Janeiro%2C%20um%20para%20355. Acesso em: 22 mar. 2023.

GARVEY, William D. **Communication**: the essence of science. Oxford: Pegamon, 1979.

GAYATHRI, R.; UMA, Vijayasundaram. Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey. **ICT Express,** v. 4, n. 2, p. 69-74, June 2018. Disponível em: https://www.sciencedirect.com/science/article/pii/S2405959518300985. Acesso em: 16 jan. 2024.

GEIGER, Christian Philipp; VON LUCKE, Jörn. Open Government and (Linked) (Open) (Government) (Data) **JeDEM**: **eJournal of eDemocracy and Open Government**, v. 4, n. 2, p. 265–278, 2012. Disponível em: https://www.jedem.org/index.php/jedem/issue/view/8. Acesso em: 16 jan. 2024.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 2008.

GIL-GARCÍA, J. Ramon *et al.* (ed.). **ICEGOV 2012**: 6th International Conference on Theory and Practice of Electronic Governance. New York: ACM Press, 2012.

GOMES, Wilson. **A democracia no mundo digital**: História, problemas e temas. São Paulo: Edições SESC, 2018.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1. ed. Massachussets: MIT Press, 2016.

GRACIA, Jorge; MENA, Eduardo. Semantic Heterogeneity Issues on the Web, **IEEE Internet Computing**, p. 60-67, Sept./Oct. 2012. Disponível em: https://oa.upm.es/14460/1/04.06319296.pdf. Acesso em: 16 jan. 2024.

GRUBER, Thomas R. A Translation Approach to Portable Ontology Specifications. **Knowledge Acquisition,** v. 6, n. 2, p.199-221, 1993.

GUARINO, Nicola. Formal Ontology in Information Systems. I*n:* FOISí98, Trento, Italy, 6-8. **Proceedings [...].** Amsterdam: IOS Press, June 1998, p. 3-15

GUIMARÃES, João Pero Pereira; ANDRADE, Morgana Carneiro de; BAPTISTA, Ana Alice. Alinhamento de vocabulário de domínio utilizando os sistemas AML e LogMap. **Revista Digital de Biblioteconomia e Ciência da Informação**, v. 20, 2022. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8668437 Acesso em: 16 jan. 2024.

GUNN, W.; POULTER, A. **Open Data:** Governance, Legal and Social Challenges. [S.l.]. Routledge, 2019.

GUPTA, Rupal; MALIK, Sanjay Kumar. A classification using RDFLIB and SPARQL on RDF dataset. Journal of Information and Optimization Sciences, v. 43, n. 1, p. 143-154, 2022.

GURSTEIN, Michael B. Open data: Empowering the empowered or effective data use for everyone? **First Monday**, v, 16, n. 2, Feb. 2011.

HALEVY, Alon; NORVIG, Peter; PEREIRA, Fernando. The Unreasonable Effectiveness of Data. **IEEE Intelligent Systems**, v. 24, n. 2, p. 8-13, Mar./Apr. 2009.

HAN, Jiawe; KAMBER, Micheline. **Data Mining**: Concepts and Techniques. 2nd ed. New York: Elsevier, 2006.

HEATH, Tom; BIZER, Christian. **Linked Data**: Evolving the Web into a Global Data Space. New York: Morgan & Claypool, 2011.

HEMMATI, M. & BUHL, H. U. Governance of open data: Challenges and approaches. **Information Systems Management**, v. 35, n. 2, p. 107-118, 2018.

HERRERA VARELA, Ricardo. **Bibliomining**: minería de datos y Bibliotecário de conocimiento em bases de datos aplicados al ámbito Bibliotecário. 2006. Disponível em https://fdocuments.ec/document/bibliomining-mineria-de-datos-y-descubrimiento-de-crisp-dm-dentro-de.html. Acesso em: 03 jul. 2023.

HINDMAN, Matthew. **The Myth of Digital Democracy**. Princeton, NJ: Princeton University Press. 2009.

HOFFMAN, Jussara. **Avaliar para promover**: as setas do caminho. Porto Alegre: Mediação, 2001.

HOGAN, Aidan *et al*. Weaving the Pedantic Web. *In:* LINKED DATA ON THE WEB (LDOW2010), April 27th, 2010, Raleigh, North Carolina. **Proceedings [...].** Disponível em: http://events.linkeddata.org/ldow2010/. Acesso em: 22 jun. 2023.

HULLMAN, Jessica; DIAKOPOULOS, Nick. Visualization Rhetoric: Framing Effects in Narrative Visualization. **IEEE Transactions on Visualization and Computer Graphics,** v. 17, n. 12, p. 2231-2240, Dec. 2011.

HUNTINGTON, Samuel P. **The Third Wave**: Democratization in the Late 20th Century. Oklahoma: University of Oklahoma Press, 1993.

IBGE. **Cidades e Estados**. [2023]. Disponível em: https://www.ibge.gov.br/cidades-e-estados/sp/presidente-prudente.html. Acesso em: 19 jan. 2024.

ISHAK, Poebe W. Murder nature: Weather and violent crime in rural Brazil. **World Development**, v. 157, Sept. 2022. DOI https://doi.org/10.1016/j.worlddev.2022.105933.

ISOTANI, Seiji.; BITTENCOURT, Ig Ibert. **Dados abertos conectados**. São Paulo: Novatec, 2015. Disponível em: https://pgcl.uenf.br/arquivos/dadosabertosconectados_011120181613.pdf. Acesso em: 16 jan. 2024.

JACOB, Brian, LEFGREN, Lars; MORETTI, Enrico. (2007). The Dynamics of Criminal Behavior: Evidence from Weather Shocks. **The Journal of Human Resources**, v. 42, n. 3, p. 489-527, 2007.

JACOBI, Pedro Roberto. **Inovação no campo da gestão pública local**: novos desafios, novos patamares. Rio de Janeiro: FGV Editora, 2006. p. 131.

JAMES, Laura. **Defining Open Data**. Oct. 2013. Disponível em: https://blog.okfn.org/2013/10/03/defining-open-data/. Acesso em: 23 jan. 2024.

JESUS, Vanessa Marta de. **Dados abertos conectados a partir de catálogos online de bibliotecas**. Orientadora: Célia da Consolação Dias. 2021. 148 f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2021.

KEDIA, Pranav. **Crime Mapping and Analysis using GIS**. July 2016. Disponível em: https://www.researchgate.net/publication/309125859_Crime_Mapping_and_Analysis_using_GIS. Acesso em: 16 jan. 2024.

KEIM, Daniel *et al*. Visual analytics: Definition, process, and challenges. **Lecture Notes in Computer Science**, v. 4950, 2008.

KHAN, S., SHARMA, A. Open Data Challenges: An Overview. **International Journal of Computer Science and Mobile Computing**, v. 9, n. 3, p. 1-7. 2020.

KITCHIN, Rob. Big Data, new epistemologies and paradigm shifts. **Big Data and Society**, v. 1, n. 1, 10 July 2014b. Disponível em: https://journals.sagepub.com/doi/10.1177/2053951714528481 Acesso em: 16 jan. 2024.

KITCHIN, Rob. **The Data Revolution**: Big Data, Open Data, Data Infrastructures and their Consequences. London: Sage, 2014a.

KLEIN, Rodrigo Hickmann; KLEIN, Deisy Cristina Barbiero; LUCIANO, Edimara Mezzomo. Identificação de mecanismos para a ampliação da transparência em portais de dados abertos: uma análise no contexto brasileiro. Cadernos EBAPE. BR, v. 16, n. 4, p. 692-715, 2018. Disponível em: https://www.scielo.br/j/cebape/a/SbSdqx7HXRF4WKVTTRgCS4m/?format=pdf&lang=pt. Acesso em: 5 fev. 2024.

KLYNE, Graham; CARROLL, Jeremy. **Resource Description Framework** (RDF): Concepts and Abstract Syntax. Jan. 2006.

KNEUER, Marianne. E-democracy: A new challenge for measuring democracy. **International Political Science Review**, v. 37, n. 5, p. 666–678, 2016. DOI: 10.1177/0192512116657677.

KOCH, P. Big Data, Inteligência Artificial e Internet das Coisas: Como essas tecnologias estão sendo usadas para facilitar o acesso, uso e compartilhamento de dados abertos. **Novas Tecnologias**, v. 2, n. 1, p. 1–10, 2018.

KOEHLER, J. Direitos autorais e direitos de propriedade intelectual relacionados aos dados abertos. p. S-10. 2018. Disponível em: https://link.springer.com/book/10.1007/978-3-642-25160-3. Acesso em: 2 março 2023

KRÜGER, T. Interledger Protocol: How It Works and What It Does. 2020. Disponível em: https://www.investopedia.com/terms/i/interledger-protocol.asp.

KUHN, T. Ciência cidadã e dados abertos: Uma abordagem para a inclusão, a participação e a responsabilização. **Dados Abertos**, v. 5, n. 2, p.118-120, 2019.

KUMAR, S., JAIN, A.; SHARMA, S. Big Data Management: Challenges and Recent Tre;ds. **International Journal of Computer Applications**, v. 181, n. 5, 2020.

LATHROP, Daniel; RUMA, Laurel (eds). **Open Government**: Collaboration, Transparency, and Participation in Practice. Sebastopol: O'Reilly, 2010.

LIMA, Renato Sergio de. A produção da opacidade: estatísticas criminais e segurança pública no Brasil. **Novos Estudos CEBRAP**, mar. 2008.

LIMA, Renato Sergio de; SINHORETTO, Jacqueline; BUENO, Samira. A gestão da vida e da segurança pública no Brasil. **Sociedade e Estado**, v. 30, n. 1, p. 123–144, 2015. DOI https://doi.org/10.1590/S0102-69922015000100008.

LIMA, Suzinara Beatriz Soares de *et al.* Conflitos gerenciais e estratégias de resolução pelos enfermeiros gerentes. **Revista de Enfermagem da UFSM**, v. 4, n. 2, 2014.

LIMA, V. F. Democracia, participação e controle social: a construção da cidadania na gestão pública. **Revista do Serviço Público**, v. 68, n. 2, p. 347-364, 2017.

LIMA NETO, Joaquim Soares de; VIEIRA, Thiago Augusto. A estratégia de prevenção do crime através do desenho urbano. **Revista Ordem Pública**, v. 7, n. 1, 2014. Disponível em: https://rop.emnuvens.com.br/rop/article/view/67/66. Acesso em: 3 maio 2023.

LIJPHART, Arend. **Modelos de democracia**: forma de governo e desempenho em 36 países. Rio de Janeiro: Zahar, 1997.

LOBAINA, Esther Marina Ruiz; SUÁREZ, C. P. Romero. Results obtained in a data mining process applied to a database containing bibliographic information concerning four segments of science. **Journal of Information Systems and Technology Management** (Online), v. 15, 2018. DOI: 10.4301/S1807-1775201815003. Disponível em: https://www.revistas.usp.br/ jistem/article/view/160839. Acesso em: 16 ago. 2023.

LUKOFF, B. Open Data Governance: Challenges and Opportunities. **International Journal of Public Administration in the Digital Age**, v. 3, n. 4, p. 1-14, 2016.

MACEDO, Daiane *et al*. Uma ferramenta para recomendação de visualização de dados governamentais abertos. *In*: WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO (WCGE). **Anais [...]**. Porto Alegre: SBC, 2020. Disponível em: https://sol.sbc.org.br/index.php/wcge/article/view/11261/11124. Acesso em: 16 jan. 2024.

MAIA, Tânia Sofia Vieira; CORREIA, Pedro Miguel Alves Ribeiro; COSTA, Cláudia S. Avaliação, accountability, transparência e governo aberto. **Lex Humana**, v. 14, n. 1, p. 164-185, 2022. Disponível em: https://bibliotecadigital.ipb.pt/bitstream/10198/26120/1/Maia.Correia.Costa_2022.pdf. Acesso em: 23 jan. 2024.

MANSON, Steven; MATSON, Laura. Maps, Society, and Technology. *In:* MANSON, Steven. **Mapping, Society, and Technology**. 1. ed. 2017. Disponível em: https://open.lib.umn.edu/mapping/chapter/1-maps-society-and-technology/. Acesso em: 24 jan. 2024.

MANYIKA, James *et al.* **Big data:** The next frontier for innovation, competition, and productivity. May 2011. Disponível em: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_exec_summary.pdf. Acesso em: 16 jan. 2024.

MARCOLIN, D. **A Lei 12.847/2013**: O que é a Lei de Dados Abertos e por que ela é importante? 2016. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2013/lei/l12847.htm. Acesso em: 3 jan. 2022.

MARTINEZ-PRIETO, M. A. *et al*. Open government: a review and outlook. **International Journal of Public Administration**, v. 38, n. 1, p. 13-23, 2015.

MATHEUS, Ricardo, RIBEIRO, Manuella Maia; VAZ, José Carlos. New perspectives for electronic government in Brazil: The adoption of open government data in national and subnational governments of Brazil. **ACM International Conference Proceeding Series**, p. 22–29, 2012.

MEIJ, Edgar, *et al*. Mapping queries to the Linking Open Data cloud: A case study using DBpedia. **Journal of Web Semantics**, v. 9, n. 4, p. 418–433, 2011. DOI https://doi.org/10.1016/j.websem.2011.04.001.

MEIJER, Albert. Transparency. *In*: BOVENS, Mark; GOODIN, Robert; SCHILLEMANS, Thomas (ed.). **The Oxford handbook public accountability**. Oxford: Oxford University Press, 2014. p. 507-524.

MELLO, Gilmar Ribeiro de. **Estudo das práticas de governança eletrônica**: instrumento de controladoria para a tomada de decisões na gestão dos estados brasileiros. Orientador: Valmor Slomski. 2009. 179 f. Tese (Doutorado) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2009.

MENDES, R. C. A importância da democratização das instituições públicas para o fortalecimento da democracia. *In:* SEMINÁRIO DE ADMINISTRAÇÃO, 8., 2010, Natal. **Anais [...].** Natal: UFRN, 2010.

MERTON, Robert K. Social Structure and Anomie. **American Sociological Review**, v. 3, n. 5, p. 672-682, Oct.1938. Disponível em: https://www.csun.edu/~snk1966/Robert%20K%20Merton%20-%20Social%20Structure%20and%20Anomie%20Original%201938%20Version.pdf. Acesso em: 16 jan. 2024.

MINICHIELLO, Alexandre Andrade. **Protótipo de aplicação web para dados abertos conectados**. Orientador: Herval Daminelli. 2017. 76 f. Trabalho de Conclusão de Curso (Bacharelado em Gestão da Tecnologia da Informação) – Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, Florianópolis, 2017.

MINN, Michael. **Geographic Correlation and Causation**. 2023. Disponível em: https://michaelminn.net/tutorials/correlation/. Acesso em: 5 fev. 2024.

MOREIRA, Diogo Luiz de Jesus; MALIN, Ana Maria Barcellos. Panorama sobre a utilização de dados governamentais abertos no brasil: um estudo a partir dos aplicativos desenvolvidos. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 16., 2015, João Pessoa. **Anais [...].** João Pessoa: ANCIB, 2015.

MOREIRA, Walter. Tesauros e ontologias como modelos de sistemas de organi zação do conhecimento. **Brazilian Journal of Information Science: Research Trends**, v. 13, n.1, p. 15–20, 2019. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/8277/5643. Acesso em: 16 jan. 2024.

MOURA, A. **Introdução aos Dados Abertos**. São Paulo: Editora Cactus, 2020.

MUKHERJEE, A., GHOSH, A., & ZHANG, T. **Data governance**: Managing corporate data assets, risks, and opportunities. Boca Raton, Fla: Auerbach Publications, 2015.

MUNTEAN, Mihaela *et al*. Models and patterns for achieving semantic interoperability. Sept. 2010. Disponível em: https://www.researchgate.net/publication/277248229. Acesso em: 16 jan. 2024.

NATIONAL INSTITUTE OF JUSTICE. U.S. Department of Justice. **Solicitation:** Predictive Policing Analytic and Evaluation Research Support. SL# 000879. 2009b. Disponível em: https://www.ojp.gov/sites/g/files/xyckuh171/files/media/document/NIJ-2009-2240.pdf. Acesso em: 16 jan. 2024.

NATIONAL INSTITUTE OF JUSTICE. U.S. Department of Justice. **Solicitation:** Predictive Policing Demonstration and Evaluation Program. SL# 000877. 2009a. Disponível em: https://nij.ojp.gov/sites/g/files/xyckuh171/files/media/document/NIJ-2009-2239.pdf. Acesso em: 16 jan. 2024.

NICHOLSON, Scott. O processo da bibliomineração: repositório de dados e mineração de dados para tomada de decisão em bibliotecas. **Transinformação** v.16, n. 3, p.253-261, set/dez. 2004. Disponível em: https://periodicos.puc-campinas.edu.br/transinfo/article/view/6382/4066. Acesso em: 16 jan. 2024.

OECD. **Promise and Problems of e-Democracy:** Challenges of Citizen Engagement. Paris: OECD, 2003. Disponível em: https://www.oecd-ilibrary.org/governance/promise-and-problems-of-e-democracy_g2gh481d-en. Acesso em: 16 jan. 2024.

OHM, Paul. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. **UCLA Law Review**, v. 57, p. 1701, 2010. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006. Acesso em: 19 jan. 2024.

OPEN GOVERNMENT PARTNERSHIP. **Right to Information Performance**. 2023. Disponível em: https://www.opengovpartnership.org/wp-content/uploads/2023/01/OGP_BL_PA_RighttoInfo_January2023.pdf. Acesso em: 5 fev. 2024.

OPEN KNOWLEDGE FOUNDATION. **What is open?** [2010]. Disponível em: https://okfn.org/en/library/what-is-open/. Acesso em: 23 jan. 2024.

PEARSALL, Beth. Predictive Policing: The Future of Law Enforcement? **NIJ Journal**, n. 266, p. 16-19, 2010. Disponível em: https://www.ojp.gov/pdffiles1/nij/230414.pdf. Acesso em: 16 jan. 2024.

PEIXOTO, Tiago. The Uncertain Relationship between Open Data and Accountability: A Response to Yu and Robinson's 'The New Ambiguity of Open Government. **UCLA Law Review Discourse**, n. 200, 2013.

PELED, Alon. Re-Designing Open Data 2.0. *In*: CONFERENCE FOR E-DEMOCRACY AND OPEN GOVERNMENT - CeDEM14, May 2013, Danube University Krems, Austria. **Proceedings […].** 2013. p. 243-258.

PENG, Roger D. Reproducible research in computational science. **Science**, v. 334, n. 6060, p. 1226-1227, Dec. 2011.

PERRY, Walter L. *et al*. **Predictive Policing**: The Role of Crime Forecasting in Law Enforcement Operations. Santa Monica, CA: RAND Corporation, 2013. Disponível em: https://www.rand.org/pubs/research_reports/RR233.html. Acesso em: 19 jan. 2024.

PETERS, B. Guy. **The future of governing**: Four emerging models. Lawrence, KS: University Press of Kansas, 1996.

PICKLER, Maria Elisa Valentim. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**, v. 12, n. 1, 2007. Disponível em: https://www.scielo.br/j/pci/a/HHdw6KMPG45HxwShcwTmFSs/?lang=pt. Acesso em: 17 jan. 2024.

PIEDRA, Nelson *et al*. **Guidelines to producing structured interoperable data from Open Access Repositories**. 2016. Disponível em: https://www.researchgate.net/publication/311314181_Guidelines_to_producing_structured_interoperable_data_from_Open_Access_Repositories. Acesso em: 19 jan. 2024.

PIMENTEL, Carlos Alberto. **Polícia e política no Brasil**: a força pública paulista. São Paulo: Editora Unesp. 2006.

PIMENTEL, Rodrigo. **História da Polícia Militar do Estado de São Paulo**: Uma trajetória de conquistas. São Paulo: Editora Barcarolla, 2006.

PINTO, C. A. de M. **Operação Verão Permanente**: Proposta de Ferramenta de Planejamento. 2019.

PMESP. **Relatório Anual de 2019**. São Paulo: PMESP, 2020. Disponível em: https://www.policiamilitar.sp.gov.br/unidades/cged/relatorios-anuais/relatorio-anual-2019.html. Acesso em: 25 mar. 2023.

PRUD'HOMMEAUX, Eric; SEABORNE, Andy. **SPARQL Query Language for RDF**. W3C Recommendation 15 January 2008. Disponível em: https://www.w3.org/TR/rdf-sparql-query/. Acesso em: 23 jan. 2024.

RAHM, Erhard; HONG HAI DO. Data Cleaning: Problems and Current Approaches. **Data Engineering Bulletin**, v. 23, n. 4, p. 3-13, 2000.

RAUTENBERG, Sandro; BURDA, Alessandra Cassiana; SOUZA, Lucélia de. Um workflow para compartilhamento de dados científicos primários baseados em dados abertos conectados. **Encontros Bibli,** v. 23, n. 53, p. 110–123, 6 set. 2018. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p110/37293. Acesso em: 17 jan. 2024.

RATCLIFFE, Jerry H. **The spatial context of crime**: A conceptual model and empirical analysis. New York: Springer, 2015.

RAUTENBERG, Sandro *et al*. Dados abertos conectados e gestão do conhecimento: estudos de caso cientométricos em uma universidade brasileira. **Perspectivas em Ciência da Informação**, v. 22, n. 3, p. 116–142, jul./set. 2017. Disponível em: https://www.scielo.br/j/pci/a/KykXkxTPkz369RZjZCfmjnR/#. Acesso em: 17 jan. 2024.

REINKE, Herbert. The Politics of Police History in Germany since the 1990s. A Participant Observation. **Crime, Histoire & Sociétés**, v. 16, n. 2, p. 99-106, 2012. Disponível em: https://journals.openedition.org/chs/1363. Acesso em: 24 jan. 2024.

REIS, Luiz Claudio Rezende; SÁ, Maria Irene da Fonseca e. Big data: um novo campo de atuação para bibliotecários. **Prisma.Com,** Portugal, n. 41, p. 231–250, 2020. Disponível em: https://ojs.letras.up.pt/index.php/prismacom/article/view/6752/6243. Acesso em: 23 jan. 2024.

RIBEIRO, Ludmila. A produção decisória do sistema de justiça criminal para o crime de homicídio. **DADOS – Revista de Ciências Sociais,** Rio de Janeiro, v. 53, n. 1, p. 159-193, 2010. Disponível em: https://www.scielo.br/j/dados/a/qbPCJwRRZLCdjWSxJRvRPDJ/?format=pdf&lang=pt. Acesso em: 17 jan. 2024.

ROCHA, Alexandre Pereira da. **Análises Criminal e de Inteligência**: Definições Teóricas e Desafios Práticos para as Polícias do Brasil. 2020. Disponível em: https://editora.pucrs.br/edipucrs/acessolivre/anais/congresso-internacional-de-ciencias-criminais/assets/edicoes/2020/arquivos/78.pdf. Acesso em: 5 ago. 2023.

ROCHA, A. C. A.; SILVA, L. A. C.; MELO, L. A. B. A Lei de Acesso à Informação e sua aplicação em municípios brasileiros. Revista de Gestão e Secretariado, v. 10, n. 2, p. 1-19, 2019.

RODRIGUES, Fabio A.; MACIEL, Cristiano. Um método para captura e compartilhamento de dados abertos educacionais via um processo ETL. *In:* WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO (WCGE), 10., 2022, Niterói. **Anais [...].** Porto Alegre: Sociedade Brasileira de Computação, 2022. p. 133-144.

RODRIGUES, Fernando Assis; SANT'ANA, Ricardo César Gonçalves; FERNEDA, Edberto. Análise do processo de recuperação de conjuntos de dados em repositórios governamentais. **InCID: Revista de Ciência da Informação e Documentação**, v. 6, n. 1, p. 38-56, 2015. DOI: 10.11606/issn.2178-2075.v6i1p38-56. Disponível em: https://www.revistas.usp.br/incid/article/view/73496/96247. Acesso em: 23 jan. 2024.

RODRIGUES, Karina Furtado. Desvelando o conceito de transparência: seus limites, suas variedades e a criação de uma tipologia. **Cadernos EBAPE. BR**, v. 18, n. 2, p. 237-253, 2020. Disponível em: https://www.scielo.br/j/cebape/a/x7BckSpN4dvNMqQmkM5QHcq/?format=pdf&lang=pt. Acesso em: 5 fev. 2024.

RONCHI, Carlos Cesar; TODARO, Mauro Enrique Carozzo; SERRA, Antônio Roberto Coelho. Cidades inteligentes, pessoas inteligentes e desinformação. **Revista de Ciências da Administração**, v. 1. n. esp., p. 1–13, 2023. Disponível em: https://periodicos.ufsc.br/index.php/adm/article/view/96300/55014. Acesso em: 23 jan. 2024.

ROWE, Michael. **Introduction to Policing**. 2nd ed. California: Sage Publishing, 2013.

ROZSA, Vitor; DUTRA, Moisés Lima; NHACUONGUE, Januário Albino. Linked Open Data no contexto acadêmico. **Brazilian Journal of Information Science: research trends**, v. 11, n. 3, p. 34-52, 9 out. 2017. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/6780/4651. Acesso em: 17 jan. 2024.

SAEED, Khalid; CHAKI, Rituparna; JANEV, Valentina. (ed.). **Computer Information Systems and Industrial Management**. Cham: Springer International Publishing, 2019. 536 p.

SANTARÉM SEGUNDO, José Eduardo. Web semântica: fluxo para publicação de dados abertos e ligados. **Informação em Pauta**, v. 3 n. especial, nov. 2018. Disponível em: http://www.periodicos.ufc.br/informacaoempauta/article/view/39721/pdf. Acesso em: 17 jan. 2024.

SANTOS, Elisângela Oliveira dos. A luta de um comando e o uso dos dados como instrumento para a elaboração de Estratégias de atuação de um batalhão da polícia militar do Estado do Rio de Janeiro. **Mediações - Revista de Ciências Sociais**, v. 26, n. 2, p, 292-310, 2021. Disponível em: https://ojs.uel.br/revistas/uel/index.php/mediacoes/article/view/42816/32455. Acesso em: 17 jan. 2024.

SANTOS, Maria Teresa Silva; GASPARINI, Isabela; FRIGO, Luciana Bolan; DALLE MULLE, Laís de Oliveira. Ferramenta de visualização de Dados Abertos do Portal de Transparência da Câmara Municipal da Cidade de Florianópolis. *In*: WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO (WCGE), 9., 2021. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 71-82. DOI: https://doi.org/10.5753/wcge.2021.15978. Disponível em: https://sol.sbc.org.br/index.php/wcge/article/view/15978/15819. Acesso em: 17 jan. 2024.

SANTOS, Paloma Maria. **Framework de apoio à democracia eletrônica em portais de governo com base nas práticas de gestão do conhecimento**. Orientador: Aires José Rover. 2014. 430 f. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento, Florianópolis, 2014.

SAUERWEIN, Clemens; PEKARIC, Irdin; FELDERER, Michael; BREU, Ruth. An analysis and classification of public information security data sources used in research and practice. **Computers and Security**, v. 82, p. 140–155, 2019.

SILVEIRA, Lúcia da; RIBEIRO, Nivaldo Calixto; SANTOS, Sarah Rúbia de Oliveira; SILVA, Fernanda Meirelle de Almeida; SILVA, Fabiano Couto Corrêa da; CAREGNATO, Sônia Elisa; OLIVEIRA, Adriana Carla Silva de; OLIVEIRA, Dalgiza Oliveira; GARCIA, Joana Coeli Ribeiro; ARAÚJO, Ronaldo Ferreira. Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. Encontros Bibli: revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 26, n. 1, p. 1-27, 2021. DOI 10.5007/1518-2924.2021.e79646. Disponível em: https://lume.ufrgs.br/handle/10183/231138 Acesso em: 10 fev. 2024.

REINKE, Herbert. The Politics of Police History in Germany since the 1990s. A Participant Observation. **Crime, Histoire & Sociétés**, v. 16, n. 2, p. 99-106, 2012. Disponível em: https://journals.openedition.org/chs/1363. Acesso em: 24 jan. 2024.

SELTZER, Ethan; MAHMOUDI, Dillon. Citizen Participation, Open Innovation, and Crowd- sourcing: Challenges and Opportunities for Planning. **Journal of Planning Literature**, v. 28, n. 1, p. 3–18, Feb. 2013.

SENTO-SÉ, João Trajano. Prevenção ao Crime e Teoria Social. **Lua Nova: Revista de Cultura e Política**, n. 83, 2011. Disponível em: https://www.scielo.br/j/ln/a/mHYMCDRWhLWDDztLBTqhbxn/#. Acesso em: 19 jan. 2024.

SHADBOLT, Nigel; BERNERS-LEE, Tim; HALL, Wendy. The Semantic Web Revisited. **IEEE Intelligent Systems**, v. 21, n. 3, p. 96-101, Jan./Feb. 2006.

SHAFRANOVICH, Yakov. **Common Format and MIME Type for Comma-Separated Values (CSV) Files**. RFC 4180. Oct. 2005. Disponível em: https://www.ietf.org/rfc/rfc4180.txt. Acesso em: 17 jan. 2024.

SHAW, Clifford R.; MCKAY, Henry D. Delinquency Rates and Community Characteristics. *In*: SHAW, Clifford R.; MCKAY, Henry D. **Juvenile delinquency and urban areas**. Chicago: University of Chicago Press,1942. Chap. 6, p. 140-169. Disponível em: https://faculty.washington.edu/matsueda/courses/517/Readings/Shaw%20and%20McKay%206-7.pdf. Acesso em: 17 jan. 2024.

SHETH, Amit; HENSON, Cory; SAHOO, Sathia S. Semantic sensor web. **IEEE Internet Computing**, v. 12, n. 4, p. 78–83, July 2008.

SILVA, André Henrique de Oliveira; SENA, Luis Cosme Marinho de. **Análise criminal na polícia militar do Estado do Rio de Janeiro**: experiências bem-sucedidas. Rio de Janeiro: Escola Superior de Polícia Militar: Polícia Militar do Estado do Rio de Janeiro, 2015.

SILVA, D. G. D. *et al.* A formação em direitos humanos na Polícia Militar de Santa Catarina: uma análise crítica. **Revista Brasileira de Segurança Pública**, v. 13, n. 2, p. 2-22, 2019.

SILVA, Edson Emanuel Nonato; ROLIM, Vanderlan Hudson. A importância da atividade de inteligência de segurança pública na prevenção criminal. **O Alferes**, Belo Horizonte, v. 70, n. 27, p. 139-168, jan./jun. 2017.

SILVA, M. C. C. da. Polícia Militar e direitos humanos: Da resistência à construção de uma nova cultura organizacional. **Revista de Direitos e Garantias Fundamentais**, v. 21, n. 1, p. 331-354, 2020.

SINGHAL, Bikramaditya; DHAMEJA, Gautam; PANDA, Priyansu Sekhar. **Beginning Blockchain**: A Beginner's guide to building Blockchain solutions. New York: Apress, 2018.

SØRENSEN, Estrid; KOCKSCH, Laura. Data Durabilities: Towards Conceptualizations of Scientific Long-Term Data Storage. **Engaging Science, Technology, and Society**, v. 7, n. 1, p. 12-21, 2021.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, p. 132-141, 2004. Disponível em: https://revista.ibict.br/ciinf/article/view/1077/1177. Acesso em: 17 jan. 2024.

STAAB, S., & STUDER, R. (2004). Handbook on Ontologies. Springer Science & Business Media.

TANDOC, E. Data openness and government transparency: A systematic review of the literature. **Government Information Quarterly**, v. 35, n. 3, p. 301-315, 2018.

TEKLI, Joe, CHBEIR, Richard & YETONGNON, Kokou. An overview on XML similarity: Background, current trends and future directions. **Computer Science Review**, v. 3, n. 3, p. 151–173, 2009.

TRIVIÑOS, Augusto N. S. **Introdução à pesquisa em ciências sociais**: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987. 175p.

UCHIDA, Craig D. **A National Discussion on Predictive Policing**: Defining our Terms and Mapping Successful Implementation Strategies. Los Angeles, Calif: NIJ, May 2009. Disponível em: https://www.ncjrs.gov/PDFFILES1/NIJ/GRANTS/230404.PDF. Acesso em: 17 jan. 2024.

UNESCO. **Open Data:** Challenges and Opportunities. 2018. Disponível em: https://www.unesco.org/en/open-solutions/open-data. Acesso em: 17 jan. 2024.

UNITED NATIONS. **Declaração Universal dos Direitos Humanos**. [1948]. Artigo 12. (Tradução oficial da High Commission for Human Rights). Disponível em: https://www.ohchr.org/en/human-rights/universal-declaration/translations/portuguese?LangID=por. Acesso em: 13 jan. 2023.

UNITED NATIONS. **E-Government Survey 2014.** Disponível em: https://publicadministration.un.org/egovkb/en-us/reports/un-e-government-survey-2014. Acesso em: 22 mar. 2023.

USA.Gov. **O que são dados abertos?** 2018. Disponível em: https://resources.data.gov/resources/data-gov-open-data-howto/. Acesso em: 12 dez. 2022.

USCHOLD, Michael; GRUNINGER, Michael. Ontologies: Principles, Methods and Applications. **Knowledge Engineering Review,** v. 11, n. 2, 1996.

VASCONCELLOS, M. A. Análise de dados abertos para prevenção da violência. **Revista Brasileira de Segurança Pública**, v. 10, n. 2, p. 157-175, 2016.

VENTURA, M. A. **História da democracia**. São Paulo: Contexto, 2016.

VICTORINO, Marcio de Carvalho *et al*. Uma proposta de ecossistema de big data para a análise de dados abertos governamentais conectados. **Informação & Sociedade: Estudos**, v.27, n.1, p. 225-242, jan./abr. 2017. Disponível em: https://periodicos.ufpb.br/ojs/index.php/ies/article/view/29299/17505. Acesso em: 17 jan. 2024.

VU THUY HUONG LE *et al*. The Effects of Daily Temperature on crime events in Urban Hanoi, Vietnam using seven years of data (2013–2019). **International Journal of Environmental Research and Public Health**, v. 19, n. 21, 2022. DOI https://doi.org/10.3390/IJERPH192113906.

WEST, Johnny. **The next target user group for the open data movement is governments**. Sept, 2018. Disponível em: https://blog.okfn.org/2018/09/18/the-next-target-user-group-for-the-open-data-movement-is-governments/. Acesso em: 23 jan. 2024.

WILSON, Ronald E. *et al*. **Geospatial Technology Working Group Meeting Report on Predictive Policing**. Scottsdale, Arizona: U.S. Department of Justice, 2009. Disponível em: https://www.ojp.gov/pdffiles1/nij/237409.pdf. Acesso em: 8 abr. 2023.

WISHART, David S. *et al*. **DrugBank 5.0**: A major update to the DrugBank database for 2018. Nucleic Acids Research, v. 4, n. 46, p. D1074–D1082, Jan. 2018.

WOELFLE, Michael; OLLIARO, Piero; TODD, Matthew H. Open science is a research accelerator. **Nature Chemistry**, v, 3, p. 745–748, 2011.

WOOD, E. La policía mexicana y las reformas policiacas. **Nueva Sociedad**, Caracas, n. 282, p. 31-44, 2019.

YU, Harlan; ROBINSON, David G. The new ambiguity of 'Open Government'. **UCLA Law Review Discourse,** v. 178, 2012.

ZAPF, Stefan; KRAUSHAAR, Christopher. **A new visualization to beautifully explore correlations**: Introducing the solar correlation map, and how to easily create your own. Jan 2017. Disponível em: https://www.oreilly.com/content/a-new-visualization-to-beautifully-explore-correlations/. Acesso em: 17 jan. 2024.

ZAVERI, A. *et al*. Quality assessment for Linked Data: A Survey. **Semantic Web**, v. 1 2012. Disponível em: https://www.semantic-web-journal.net/system/files/swj773.pdf. Acesso em: 17 jan. 2024.

ZHIYUAN ZHANG *et al.* Visual correlation analysis of numerical and categorical data on the correlation map. **IEEE Transactions on Visualization and Computer Graphics**, v. 21, n. 2, p. 289–303, Feb. 2015.

ZUIDERWIJK, Anneke; JANSSEN, Marijn. Open data policies, their implementation and impact: A framework for comparison. **Government Information Quarterly**, v. 31, n. 1, p. 17–29, 2014.

APÊNDICE A
CÓDIGO RDF DA PROPOSTA ONTOLÓGICA

```
This XML file does not appear to have any style information associated with
it. The document tree is shown below.
1 of 7

2/6/24, 12:38 PM
-<rdf:RDF
xml:base="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predit
iva/"> <owl:Ontology
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva"/>
-<!--
--> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#acusadoCrime
   -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#acusadoCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
<rdfs:range rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#historicoCri
me
   -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#historicoCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
<rdfs:range rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Historico"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#latlngLocal
   -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#latlngLocal">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/ >
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Latlng"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#localAcusado
   -->
```

```
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#localAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#localClima
   -->
///////////////////////////////////////////////////////////////////////////////
////
//
// Object Properties
//
///////////////////////////////////////////////////////////////////////////////
////
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
2 of 7
2/6/24, 12:38 PM
policia_preditiva#localClima">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Clima"/ >
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#localCrime
 -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#localCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#localVitima
 -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#localVitima">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Vitima"/>
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#municipioLoc
al
```

```
 -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#municipioLocal">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/ >
<rdfs:range rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Municipio"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#naturezaCrim
e
 -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#naturezaCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
<rdfs:range rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#NaturezaCrime"/>
</owl:ObjectProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#vitimaCrime
 -->
-<owl:ObjectProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#vitimaCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
<rdfs:range
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Vitima"/>
</owl:ObjectProperty> -<!--

//////////////////////////////////////////////////////////////////////////////
////
      //
Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
3 of 7
2/6/24, 12:38 PM
      // Data properties
      //

//////////////////////////////////////////////////////////////////////////////
////
--> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#bairro
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#bairro">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--
```

```
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#cep
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#cep">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Local"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#cpfAcusado
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#cpfAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#cpfVitima
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#cpfVitima">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Vitima"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#datahoraClim
a
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#datahoraClima">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Clima"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#datahoraCrim
e
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#datahoraCrime">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Crime"/ >
Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
4 of 7
2/6/24, 12:38 PM
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
</owl:DatatypeProperty>
```

```
-<!--
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#descricaoNat
ureza
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#descricaoNatureza">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#NaturezaCrime"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#indicePluvio
metrico
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#indicePluviometrico">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_pre
ditiva#Clima"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#integer"/>
</owl:DatatypeProperty> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#latlng
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#latlng">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Latlng"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#modusoperand
isAcusado
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#modusoperandisAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#nomeAcusado
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#nomeAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#nomeMunicipi
o
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#nomeMunicipio">
```

<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Municipio"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
5 of 7
2/6/24, 12:38 PM
-<!--
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#nomeVitima
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#nomeVitima">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Vitima"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#numeroResidencia
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#numeroResidencia">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Local"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rgAcusado
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rgAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Acusado"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rgVitima
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rgVitima">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Vitima"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rua
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#rua">
<rdfs:domain
rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Local"/ >
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

```
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#telefoneAcus
ado
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#telefoneAcusado">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Acusado"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#telefoneViti
ma
Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
6 of 7
2/6/24, 12:38 PM
-->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#telefoneVitima">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Vitima"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#textoHistori
co
 -->
-<owl:DatatypeProperty
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#textoHistorico">
<rdfs:domain rdf:resource="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#Historico"/>
<rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty> -<!--

//////////////////////////////////////////////////////////////////////////////
////
       //
       // Classes
       //

//////////////////////////////////////////////////////////////////////////////
////
--> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Acusado
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Acusado"/> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Clima
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Clima"/> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Crime
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Crime"/> -<!--
```

```
http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Historico
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Historico"/> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Latlng
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Latlng"/> -<!--
    http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Local
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Local"/> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Municipio
 -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Municipio"/> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#NaturezaCrim
e
 -->
<owl:Class rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#NaturezaCrime"/> -<!--
Firefox
file:///Users/macbook/Downloads/policia_preditiva_v8%20(1).rdf
7 of 7
2/6/24, 12:38 PM
      http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#Vitima
   -->
<owl:Class
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/policia_predi
tiva#Vitima"/> -<!--

//////////////////////////////////////////////////////////////////////////////
////
         //
         // Individuals
         //

//////////////////////////////////////////////////////////////////////////////
////
--> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#incidenciaFu
rto
   -->
<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#incidenciaFurto"/> -<!--

http://www.semanticweb.org/macbook/ontologies/2024/0/policia_preditiva#2024_001_09_
1
   -->
-<owl:NamedIndividual
rdf:about="http://www.semanticweb.org/macbook/ontologies/2024/0/
policia_preditiva#2024_001_09_1">
```

```
<policia_preditiva:indicePluviometrico
rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">50</
policia_preditiva:indicePluviometrico> </owl:NamedIndividual>
</rdf:RDF> -<!--
    Generated by the OWL API (version 4.5.26.2023-07-17T20:34:13Z)
https://github.com/owlcs/owlapi
 -->
```