



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO ACADÊMICO EM CIÊNCIA DA INFORMAÇÃO

MAURÍCIO TORRES DE MATOS

**INTERSECÇÕES ENTRE A CIÊNCIA DA INFORMAÇÃO
E A CIÊNCIA DE DADOS:
CONTRIBUIÇÕES PARA OS SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS**

BELÉM
2021

MAURÍCIO TORRES DE MATOS

**INTERSECÇÕES ENTRE A CIÊNCIA DA INFORMAÇÃO
E A CIÊNCIA DE DADOS:
CONTRIBUIÇÕES PARA OS SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação do Instituto de Ciências Sociais Aplicadas, Universidade Federal do Pará, como requisito para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Gestão da Informação e Organização do Conhecimento

Linha de pesquisa: Organização do Conhecimento.

Orientadora: Profa. Dra. Marise Teles Condurú

BELÉM
2021

Dados Internacionais de Catalogação-na-Publicação (CIP)

- M425c Matos, Maurício Torres de
Intersecções entre a Ciência da Informação e a Ciência de Dados: contribuições para os sistemas de informação governamentais / Maurício Torres de Matos. — 2021.
79 f. : il. color.
Orientadora: Prof^a Dra. Marise Teles Condurú
Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação, Instituto de Ciências Sociais Aplicadas, Universidade Federal do Pará, Belém, 2021.
1. Sistemas de informação. 2. Ciência de dados. 3. Sistemas de informação governamentais. 4. Ciência da Informação. I. Condurú, Marise Teles, orient. II. Título.

CDD 020

MAURÍCIO TORRES DE MATOS

**INTERSECÇÕES ENTRE A CIÊNCIA DA INFORMAÇÃO
E A CIÊNCIA DE DADOS:
CONTRIBUIÇÕES PARA OS SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação do Instituto de Ciências Sociais Aplicadas, Universidade Federal do Pará, como requisito para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Gestão da Informação e Organização do Conhecimento

Linha de pesquisa: Organização do Conhecimento.

Data da aprovação: 29.01.2021

Banca examinadora:

Prof. Dra. Marise Teles Condurú - Orientadora
Doutora em Ciências do Desenvolvimento Socioambiental
Universidade Federal do Pará

Prof. Dr. Fernando de Assis Rodrigues - Membro interno
Doutor em Ciência da Informação
Universidade Federal do Pará

Prof. Dr. Ricardo Cesar Gonçalves Sant'Ana - Membro externo
Doutor em Ciência da Informação
Universidade Estadual Paulista Júlio de Mesquita Filho

À Jéssica, pelo apoio incondicional em todos os momentos.

AGRADECIMENTOS

À minha orientadora, Marise Teles Condurú, pelo incentivo, paciência, instrução e direcionamento durante o curso do mestrado.

Ao corpo docente e técnico do Programa de Pós-Graduação em Ciência da Informação da UFPA.

Aos amigos da turma 2019 de mestrado desse Programa, pelo convívio e compartilhamento de conhecimento.

Aos familiares e amigos pelo apoio em todas as fases, em especial à minha noiva, Jéssica, por toda a ajuda.

*“Para ganhar conhecimento, adicione coisas todos os dias.
Para ganhar sabedoria, elimine coisas todos os dias”.*

Lao-Tsé

RESUMO

Estudo sobre Sistemas de Informação Governamentais, levando-se em conta que seu uso crescente, em confluência com tecnologias, busca atender as demandas atuais por transparência e eficiência dos serviços públicos, e leva a discussões quanto a quais tecnologias podem ser utilizadas com esse propósito e como a Ciência da Informação (CI) pode contribuir nesse contexto. Desse modo, na pesquisa tem-se como objetivo explorar as relações entre Ciência da Informação e Ciência de Dados para aplicações em Sistemas de Informação Governamentais, viabilizando a transformação de dados em informação, e como objetivos específicos: a) entender os Sistemas de Informação e Sistemas de Informação Governamental, relacionando aos conceitos de Governo Aberto e Dados Abertos; b) identificar o conceito de Ciência de Dados e sua aplicação nos Sistemas de Informação Governamentais no contexto da Ciência da Informação; c) indicar as relações entre Ciência de Dados e Ciência de Informação aplicadas em processos de Sistemas de Informação Governamentais. A pesquisa é de cunho bibliográfico, de análise qualitativa, com estudo exploratório, a fim de analisar as relações multidisciplinares entre a Ciência da Informação e a Ciência de Dados, estudando aspectos do uso dessas intersecções em Sistemas de Informação Governamentais. Como resultados, a pesquisa evidencia que os Sistemas de Informação Governamentais tendem a incorporar o uso de tecnologias, incluindo as relacionadas à Ciência de Dados, de modo a atender demandas por transparência, responsabilização, acesso à informação e melhores serviços governamentais, havendo possibilidades de correlações entre Ciência da Informação e Ciência de Dados em áreas como Dados Abertos, interoperabilidade, Gestão do Conhecimento, *analytics* e Inteligência Artificial. Conclui-se que o estudo e aplicação dessas correlações é capaz de gerar benefícios à operação de Sistemas de Informação Governamentais, considerando as competências informacionais da Ciência da Informação como suporte ao uso de tecnologias, contribuindo para maior interação entre governo e sociedade e ao atendimento das necessidades informacionais dos envolvidos, havendo, assim, ampla possibilidade de estudos da temática e oportunidades de aplicações.

Palavras-chave: Sistemas de informação. Sistemas de informação governamentais. Ciência de Dados. Ciência da Informação. Interdisciplinaridade.

ABSTRACT

Studies Governmental Information Systems, considering that its growing use, in confluence with technologies, seeks to meet the current demands for transparency and efficiency of public services, and leads to discussions as to which technologies can be used for this purpose and how Library & Information Science (LIS) can contribute in this context. Thus, the research aims to explore the relationships between LIS and Data Science and their applications in Governmental Information Systems to enable the transformation of data into information, and as specific objectives: a) to understand Information Systems and Governmental Information Systems, relating to the concepts of Open Government and Open Data; b) identify the concept of Data Science and its application in Governmental Information Systems in the context of Library & Information Science; c) to indicate the relationships between Data Science and LIS applied in Governmental Information Systems processes. The research is of a bibliographic nature, of qualitative analysis, with an exploratory study, in order to analyze the multidisciplinary relations between LIS and Data Science, studying aspects of the use of these intersections in Governmental Information Systems. As a result, the research shows that Governmental Information Systems tend to incorporate the use of technologies, including those related to Data Science, in order to meet demands for transparency, accountability, access to information and better government services, with possibilities of correlations between LIS and Data Science in areas such as Open Data, interoperability, Knowledge Management, analytics and Artificial Intelligence. It is concluded that the study and application of these correlations is capable of generate benefits to the operation of Governmental Information Systems, considering the informational competences of LIS as support to the use of technologies, contributing to a greater interaction between government and society to meet the needs informational requirements, thus having the broad possibility of studies on the subject and application opportunities.

Keywords: Information Systems. Governmental Information Systems. Data Science. Libray & Information Science. Interdisciplinary.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação piramidal hierárquica do fenômeno conhecimento	20
Figura 2 – Modelo Genérico de Sistema.....	22
Figura 3 – Modelo de Sistema de Informação como Sistema de Trabalho.....	24
Figura 4 – Os “3Vs” do <i>Big Data</i>	33
Figura 5 - Principais Relações Interdisciplinares da Ciência de Dados.....	35
Figura 6 – Ciclo de Vida dos Dados (CVD)	42
Figura 7 – Ciclo de Vida dos Dados (CVD) para Ciência da Informação	43
Figura 8 – Características Principais dos Dados Abertos	50
Figura 9 – Proposta de Interoperabilidade Semântica	52
Figura 10 – Modelo de Interoperabilidade Semântica aplicada a Governo Eletrônico	53
Figura 11 – Modelo de Governo Eletrônico Centralizado.....	54
Figura 12 – Fases de Análise de Dados e o Papel do Profissional da Informação ...	59
Figura 13 – Áreas da Inteligência Artificial.....	60
Figura 14 – Intersecções entre Ciência de Dados e Ciência da Informação em aplicações de Sistemas de Informação Governamentais.....	62
Quadro 1 – Quadro-Síntese dos Grupos de Intersecções entre Ciência da Informação e Ciência de Dados com aplicações em Sistemas de Informação Governamentais.....	64

LISTA DE ABREVIATURAS

BI	<i>Business Intelligence</i>
CI	Ciência da Informação
COD	Ciência Orientada a Dados
CVD	Ciclo de Vida de Dados
e-Gov	<i>Eletronic Government</i>
e-Ping	Padrões de Interoperabilidade de Governo Eletrônico
GC	Gestão do Conhecimento
IA	Inteligência Artificial
OGD	Open Government Data
OGP	Open Government Partnership
ONU	Organização das Nações Unidas
RDF	<i>Resource Description Framework</i>
SOC	Sistemas de Organização do Conhecimento
TGS	Teoria Geral dos Sistemas
TIC	Tecnologia de Informação e Comunicação

SUMÁRIO

1	INTRODUÇÃO.....	13
2	SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS.....	18
2.1	COMPREENENDO OS DADOS, A INFORMAÇÃO E O CONHECIMENTO..	18
2.2	SISTEMAS DE INFORMAÇÃO.....	21
2.3	GOVERNO ABERTO E OS SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS	25
3	CIÊNCIA DE DADOS	31
3.1	SOBRE A CIÊNCIA DE DADOS.....	31
3.2	RELAÇÕES INTERDISCIPLINARES	34
3.2.1	Áreas de estudo relacionadas.....	37
3.3	IMPLICAÇÕES DA CIÊNCIA DE DADOS EM AMBIENTES ORGANIZACIONAIS	39
3.4	CICLO DE VIDA DOS DADOS	41
4	METODOLOGIA	45
4.1	OBJETO DE PESQUISA.....	45
4.2	PROCEDIMENTOS DE COLETA DE DADOS.....	45
4.3	FORMAS DE ANÁLISE DE DADOS.....	47
5	INTERSECÇÕES ENTRE CIÊNCIA DA INFORMAÇÃO E CIÊNCIA DE DADOS	49
5.1	DADOS ABERTOS.....	49
5.2	INTEROPERABILIDADE	52
5.3	GESTÃO DO CONHECIMENTO	56
5.4	ANALYTICS.....	58
5.5	INTELIGÊNCIA ARTIFICIAL.....	60
5.6	SOBRE AS INTERSECÇÕES ENCONTRADAS.....	62
6	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS.....	69

1 INTRODUÇÃO

A sociedade da informação, fenômeno observado a partir da década de 1970, vem gerando mudanças sociais e econômicas que alteram as relações entre as Tecnologias de Informação e Comunicação (TIC) e a sociedade (CASTELLS, 1999). Em uma era em que informação se torna o principal ativo socioeconômico, os Sistemas de Informação (SI) assumem grande importância para a sociedade, indivíduos e organizações em geral.

Essa importância é percebida também em Sistemas de Informação Governamentais, sistemas esses voltados para a cooperação e interação de governos com seu ambiente e cidadãos relacionando elementos organizacionais, humanos e tecnológicos, de forma a constituir um sistema com o objetivo de mudança das relações governamentais (PARYCEK; HOCHTL; GINNER, 2014), mudança esta que tem como fio condutor o uso de tecnologias de informação (CORDELLA; IANNACCI, 2010).

Na sociedade em rede, a busca por transparência e abertura da informação, inclusive no âmbito governamental, apresenta novos contornos e uma tendência à facilitação do acesso à informação (REGO; FREIRE, 2018). Moreira, Valentim e Sant'Ana (2018) destacam ainda o uso crescente de recursos tecnológicos para acesso a dados governamentais. Essa conjuntura pode gerar a necessidade de mecanismos e sistemas de informação mais eficientes no âmbito governamental, com a tecnologia podendo atuar como facilitadora desse processo.

Nesse contexto, a Ciência da Informação como área interdisciplinar preocupada em coletar, tratar, analisar, classificar, armazenar, recuperar e disseminar a informação (OLIVEIRA; ROSA; OLIVEIRA; LIMA, 2017), pode contribuir de maneira incisiva para o estudo da dinâmica de Sistemas de Informação Governamentais e no processo de transformação de dados em informação nesses sistemas de informação.

Nesta pesquisa, embora se considere a definição de SI como unidades de informação resultantes da inter-relação de elementos humanos, materiais e tecnológicos, dá-se destaque à necessidade cada vez maior do uso de computadores e tecnologias a fim de viabilizar a inovação e aprimorar a integração de informações.

Além de questões tecnológicas, a significação da informação para o

usuário de modo a permitir a geração de conhecimento é um aspecto importante a ser considerado na concepção e operação de SI. Para tanto, estratégias de integração, armazenamento e recuperação da informação devem ser pensadas de acordo com as necessidades de informação. Como é destacado em Rau (1987), à medida que a informação é cada vez mais armazenada eletronicamente, aumenta-se também a necessidade de métodos inteligentes para o acesso e recuperação da informação.

Outra consequência do desenvolvimento da sociedade da informação é verificada no surgimento e expansão de uma disciplina denominada “Ciência de Dados”, decorrente da geração de dados em grandes volumes e variedade (caracterizando o chamado *big data*) e a necessidade de interação entre eles em maiores velocidades. A ciência de dados surge e se baseia no desenvolvimento de equipamentos (*hardware*) e *softwares* voltados para a extração de informação útil a partir de bases de dados complexas, dinâmicas, heterogêneas e distribuídas (BUGNION; MANIVANNAN; NICOLAS, 2017 apud RAUTENBERG; CARMO, 2019) – permitindo seu armazenamento, recuperação e análise.

A Ciência de Dados apresenta intersecções com as funções de um Sistema de Informação, uma vez que a integração de informações de diversas fontes, visando a necessidade de usuários e geração de valor informacional, são objetivos tanto da Ciência de Dados – essa a partir de um forte componente tecnológico – como dos sistemas de informação. Entende-se que a tecnologia, cada vez mais presente no desenvolvimento de Sistemas de Informação pode se beneficiar de técnicas e métodos desenvolvidos pela Ciência de Dados, mas também a análise de dados por meio da conjunção de CI e Ciência de Dados pode trazer benefícios a esses sistemas.

Considera-se, assim, relevante o estudo de possíveis intersecções entre a Ciência da Informação e a disciplina de Ciência de Dados, de modo a identificar benefícios e práticas a partir dessa combinação, analisando, especificamente, aplicações em Sistemas de Informação Governamentais.

A pesquisa procurou responder à seguinte questão: quais as intersecções entre a Ciência da Informação e a Ciência de Dados em processos de Sistemas de Informação Governamentais e como elas podem ser classificadas?

Para responder a essa questão, teve-se como objetivo geral explorar as relações entre Ciência da Informação e Ciência de Dados para aplicações em

Sistemas de Informação Governamentais, viabilizando a transformação de dados em informação. Para tanto, foram tomados os seguintes objetivos específicos:

- a) entender os Sistemas de Informação e Sistemas de Informação Governamental, relacionando aos conceitos de Governo Aberto e Dados Abertos;
- b) identificar o conceito de Ciência de Dados e sua aplicação nos Sistemas de Informação Governamentais no contexto da Ciência da Informação;
- c) indicar as relações entre Ciência de Dados e Ciência de Informação aplicadas em processos de Sistemas de Informação Governamentais.

A pesquisa justifica-se pela importância crescente dos temas de transparência e *accountability* e da relevância dos Sistemas de Informação Governamentais para sua efetivação, bem como da necessidade de pesquisar por maneiras de aprimorar a operação desses SI e seu fluxo informacional. Adicionalmente, o estudo das tendências tecnológicas e como elas alteram as dinâmicas dos Sistemas de Informação e a interação com a informação e usuários é relevante à CI, bem como as mudanças verificadas com o crescimento da ciência com uso intensivo de dados na relação entre dados e informação em SI demanda investigação científica, principalmente considerando, ainda, a existência de poucos estudos correlacionando a CI e a Ciência de Dados.

Para compreender o significado de *accountability* e de transparência, tomou-se por base o que afirma Rocha (2012), de que são conceitos interligados e complementares, associando *accountability* à obrigação dos governantes de prestar contas de suas ações e de sua responsabilização por elas perante a sociedade, enquanto transparência relaciona-se à publicidade da administração e informações públicas.

Entende-se que o tema de dados na Ciência da Informação, bem como possíveis relações entre essa e a Ciência de Dados, ainda é pouco explorada, tornando pertinente a realização de pesquisa explorando essa relação. Ao mesmo tempo, o surgimento de movimentos sociais ao redor do mundo pleiteando maior transparência de informações governamentais e participação popular no desenvolvimento de políticas públicas faz com que novas soluções tecnológicas sejam estudadas de modo a satisfazer a essas demandas, levantando questões

sobre como a Ciência da Informação pode contribuir nessa conjunção.

Espera-se com a pesquisa atrair atenção e contribuir para a discussão das relações entre Ciência da Informação e Ciência de Dados e do tema da transparência governamental.

Pelo exposto, para a realização deste estudo, adotou-se como metodologia, a pesquisa exploratória e abordagem qualitativa, sendo caracterizada como bibliográfica e descritiva.

Inicialmente, realizou-se revisão bibliográfica voltada à busca de publicações científicas anteriores que abordam temáticas relacionadas a interações entre Ciência da Informação e Ciência de Dados com foco em Sistemas de Informação ou e-Gov. Foram pesquisados trabalhos nas bases de dados da Ciência da Informação: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) e *Library and Information Science Abstracts* (LISA), bem como nas plataformas referenciais de citações científicas *Web of Science* e *Scopus*, tendo-se dado preferência a publicações oriundas da Ciência da Informação.

Para a análise dos dados obtidos, utilizou-se a análise de conteúdo, o que permitiu contextualizar e classificar os resultados, resultando na criação de grupos temáticos categóricos de intersecção entre CI e Ciência de Dados. Em seguida, foram identificadas correlações possíveis entre a CI e Ciência de Dados e como essas correlações podem ser aplicadas para melhoria de Sistemas de Informação Governamentais e aprimoramento à disponibilidade de dados governamentais.

Dessa forma, a pesquisa está estruturada em seis capítulos. O primeiro capítulo, que contém esta introdução, sintetiza o tema, o problema de pesquisa, os objetivos e metodologia de pesquisa.

Nos capítulos 2 e 3 é apresentada a revisão bibliográfica acerca dos conceitos que compõem a pesquisa, sendo abordado no segundo o conceito de Sistemas de Informação e Sistemas de Informação Governamentais, relacionando-os a definições de Governo Aberto, Governo Eletrônico (e-Gov), transparência e Dados Abertos, enquanto no terceiro capítulo trata-se da conceituação de dados, informação e conhecimento e conceitos acerca da Ciência de Dados e suas relações com a CI.

No capítulo 4 tem-se a metodologia de como foi realizada a pesquisa, enquanto no capítulo 5 explora-se as relações entre CI e Ciência de Dados e suas aplicações para melhorias de Sistemas de Informação Governamentais, consistindo

no resultado da pesquisa. Por fim, são apresentadas as considerações finais, no capítulo 6, seguido das referências que fundamentaram a pesquisa.

Na próxima seção aborda-se os Sistemas de Informação Governamentais, no contexto do governo aberto e dados abertos.

2 SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS

Sistema de Informação é uma disciplina de interesse da Ciência da Informação, um conceito cuja compreensão passa pelo entendimento das noções individuais de sistema e informação.

O uso de sistemas de informação permite que se reúna informação que se encontra dispersa em alguma área do conhecimento ou setor, assim como se tenha a adequada sistematização, armazenamento, disseminação e aplicação da informação registrada (ALTER, 2008). Isso leva ao entendimento da importância desses sistemas em contextos organizacionais e especialmente na conjuntura governamental, o que será refletido neste capítulo, a fim de permitir análise de sistemas de informação governamentais, favorecendo o acesso e utilização da informação para a tomada de decisão em diversos níveis, individuais e globais.

Para isso, cabe, inicialmente, entender a relação entre dados e informação, que são objetos de estudo da Ciência de Dados e da Ciência da Informação, respectivamente, como será abordado no próximo item.

2.1 COMPREENDENDO OS DADOS, A INFORMAÇÃO E O CONHECIMENTO

A Ciência da Informação em suas fundações conceituais preocupa-se com os conceitos de dado, informação e conhecimento (ZINS, 2007). Para o autor, dado para a CI pode ser concebido como o material primário da informação, de modo que esta última é por sua vez a base do conhecimento. Nesse contexto, o dado é um símbolo quantificado ou qualificado e representa a menor unidade da informação não processada.

Cornella (2000, p. 1-2) corrobora essa conceituação, acrescentando que os dados refletem fatos da realidade, apresentando um elevado nível de estrutura e consistindo em uma representação precisa de tal fato.

A partir da definição apresentada para dado, Zins (2007) afirma que a informação consiste em um conjunto de dados processados como parte de um processo de comunicação. Por sua vez, ainda segundo o autor, o conhecimento significa a apropriação da informação pelo usuário.

Conhecimento para Kant (2001, p. 123) consiste em “um sistema, a abranger e determinar por uma ideia, sistema cuja perfeição e articulação possa

oferecer, ao mesmo tempo, uma pedra de toque da exatidão e genuinidade de todos os conhecimentos que nele se incluam”.

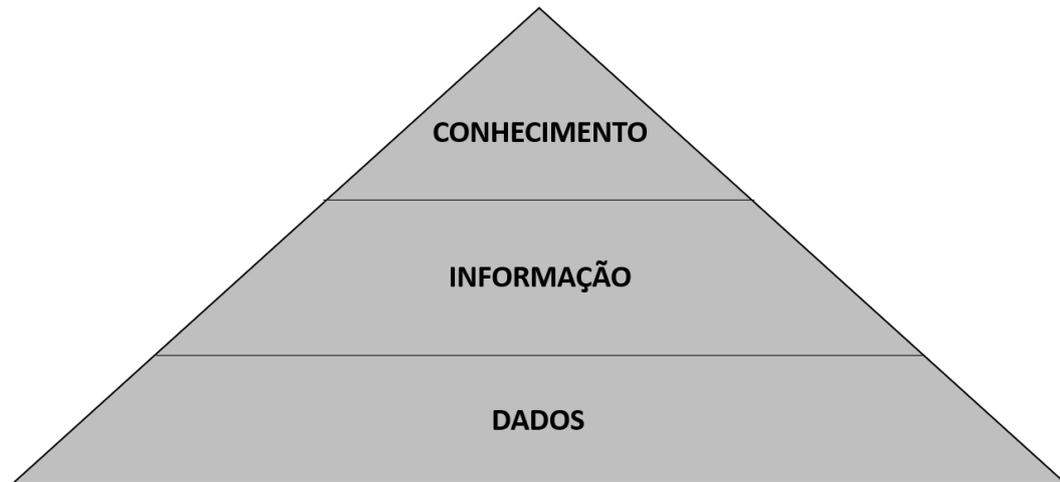
No âmbito da CI, Zins (2007) sustenta que se o conhecimento só puder ser considerado como um produto da mente, sua ordenação e estudo pela CI é impraticável, de modo que o autor elucida diversas definições de dados, informação e conhecimento que permitem seu estudo pela disciplina. Assim, pode-se definir conhecimento como “informação apropriada pelo usuário” (ZINS, 2007, p. 480), ou seja, informação assimilada de modo que, como explica Barreto (*apud* ZINS, 2007), a informação qualifica a si mesma como uma estrutura significativa capaz de gerar conhecimento ao indivíduo e seu grupo.

Para Cornella (2000), o contexto pessoal ou coletivo é o determinante que permite dar sentido a um conjunto de dados, sentido esse que será diferente para um mesmo conjunto de dados em contextos distintos. Ainda de acordo com o autor, a conversão de dados em informação só é possível a partir de sua interpretação, de modo que a geração de conhecimento se torna possível a partir da transformação pessoal da informação interpretada em estruturas mentais que permitam inferir conclusões e gerar novas ideias.

Borglund e Engvall (2014) alertam para o fato de que os termos “dados” e “informação” são utilizados comumente como sinônimos, uma situação que os autores afirmam ser errônea visto que associam informação à interpretação humana dos dados e a uma função e uso definidos, algo que não pode ser associado aos dados, que são conceituados por eles como unidade de representação da informação. Os autores ainda argumentam que o conhecimento pode ser descrito como a obtenção de informação por um indivíduo, e embora nem todo conhecimento seja informação a informação é o conhecimento em forma de comunicação.

Autores como Ackhoff (1989) e Rowley (2007) entendem que dados, informação e conhecimento são elementos distintos, mas relacionados a um mesmo fenômeno de maneira hierárquica, de modo que dados produzem informação que produz conhecimento. Essa hierarquia do fenômeno conhecimento é comumente representada na forma piramidal, conforme apresentado na Figura 1.

Figura 1 – Representação piramidal hierárquica do fenômeno conhecimento



Fonte: Rowley (2007, p. 167).

Observa-se na Figura 1 que os dados formam a base da pirâmide do fenômeno conhecimento, uma vez que constituem a unidade básica desse fenômeno, apresentando maior quantidade e não apresentando processamento (CORNELLA, 2000 e ZINS 2007). Em seguida a informação surge do processamento dos dados, como anteriormente mencionado, servindo como base para a produção e disseminação de conhecimento.

Ackhoff (1989) sugere ainda os conceitos de compreensão e sabedoria. Rowley (2007) ressalta que autores posteriores a Ackhoff, consideram a compreensão não como um elemento à parte, mas como uma habilidade necessária ao processamento de dados, informação e conhecimento. A autora afirma também que há pouca discussão em relação à natureza da sabedoria e aos processos a partir dos quais poderia ser obtida, bem como há quase nenhuma menção à sabedoria na literatura da Gestão do Conhecimento e Sistemas de Informação. A presente pesquisa, dessa forma, não considera os conceitos de compreensão e sabedoria no estudo desenvolvido.

Os conceitos de dado, informação e conhecimento são apropriados para o entendimento do objeto da Ciência e Dados, cujo foco consiste nos aspectos de armazenamento e análise de dados, sendo necessário ao profissional de dados o conhecimento da estrutura física de armazenamento e da codificação de algoritmos para análise de grandes quantidades de dados a partir de funções matemáticas.

Além disso, o aspecto humano deve ser considerado na execução de atividades relacionadas à Ciência de Dados, uma vez que o objetivo final dos

produtos por ela desenvolvidas é auxiliar a tomada de decisões e gerar valor para informação (CAO, 2016).

Santos e Sant'ana (2013) relacionam a CI à pesquisa de SI a partir da análise, projeto e evolução desses sistemas, que envolvem manipulação de dados, acesso à informação e apropriação do conhecimento, inferindo os dados como representações ou conjuntos mínimos de símbolos que tem significado decorrente de seu contexto.

Entende-se a partir de tal contextualização que a compreensão das necessidades de usuário, do contexto da informação e a interpretação de dados são requisitos importantes que dependem do entendimento dos conceitos básicos relacionados a dados, informação e conhecimento que podem ser aplicados à Ciência de Dados.

A partir do entendimento de dados e informação para a Ciência de Dados e Ciência da Informação, apresenta-se, a seguir, os Sistemas de Informação que são instrumentos relevantes para ter dados e informação reunidos e organizados para futuros usos.

2.2 SISTEMAS DE INFORMAÇÃO

Para compreensão dos Sistemas de Informação, parte-se inicialmente das definições de informação, já discutida anteriormente, e sistema.

Bertalanffy (2010) define sistema como “conjunto de elementos em inter-relação entre si e com o ambiente”. Esse conceito estabelece que um sistema é composto de elementos individuais que devem obrigatoriamente interligar-se formando um todo maior que as partes.

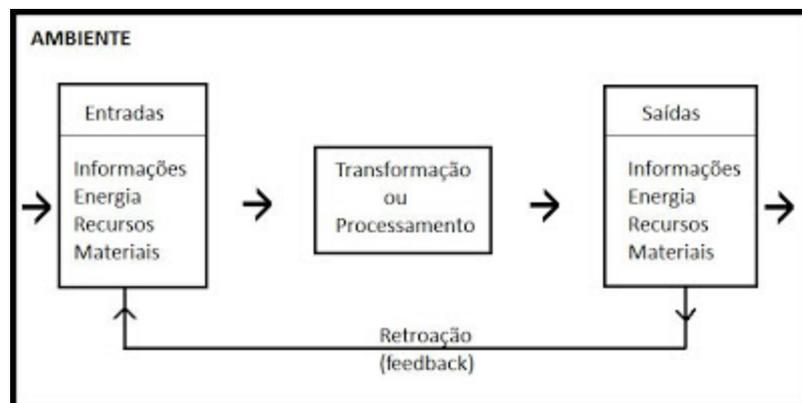
A Teoria Geral dos Sistemas (TGS), concebida por Bertalanffy, estuda os sistemas de forma global de modo a envolver a interdependência entre suas partes, destacando que sistemas possuem como características a presença de elementos individuais, a interação desses elementos a partir de atividades que visam o alcance de um objetivo central e a existência de um ambiente no qual o sistema opera. De acordo com essa teoria, os sistemas podem ser classificados como:

- Sistemas físicos e abstratos: sistemas cujas relações são mensuráveis fisicamente (físicos) ou formado por símbolos representativos de ideias (abstratos);

- Sistemas naturais e artificiais: sistemas originários de processos naturais ou de um ato consciente do ser humano;
- Sistemas fechados e abertos: sistemas isolados de seu ambiente e não sujeitos a variáveis externas (fechados) ou cujo comportamento é determinado por variáveis internas e interações dinâmicas com o ambiente (abertos).

Motta (2020) chama atenção às tentativas de aplicação da TGS em diversos campos de estudo, ressaltando que nas ciências sociais o modelo de sistema aberto revela enormes potencialidades devido à sua abrangência e flexibilidade. Como exemplifica o autor, uma organização constitui um sistema aberto, pois pode ser entendida como um complexo de elementos em interação e em intercâmbio contínuo com o ambiente. Essa mesma comparação pode ser feita com Sistemas de Informação, como será visto adiante. Um modelo geral de sistemas é apresentado na Figura 2.

Figura 2 – Modelo Genérico de Sistema



Fonte: Chiavenato (2014).

Considerando as diversas definições de informação no campo da CI, no presente estudo será utilizada a definição de Zins (2007) para se compreender um Sistema de Informação. Para o autor, informação é um conjunto de dados processados como parte de um processo de comunicação, sendo o dado, no contexto da CI, o material primário da informação, consistindo em um símbolo quantificado ou qualificado, representando a menor unidade da informação não processada.

Um Sistema de Informação, portanto, é um sistema cujo objetivo é a geração e transmissão de informações. Como explica Alter (2008, p. 451), “um sistema de informação é um sistema de trabalho cujos processos e atividades são dedicados ao processamento, captura, transmissão, armazenamento, recuperação, manipulação e exibição de informação”. Pode-se assim relacionar SI a sistemas abertos, devido à definição inferir a interação de elementos e intercâmbio com o ambiente.

Carvalho (2000 apud ALTER, 2008) entende que são características dos SI: lidar com informações; relacionar-se de algum modo a organizações ou às atividades desenvolvidas por elas; e a utilização de tecnologia da informação.

O interesse da CI em sistemas de informação justifica-se por estes constituírem:

uma ferramenta indispensável na produção, organização (metainformação ou metadados), armazenamento e recuperação da informação, incorporando na área da organização e representação/recuperação da informação as técnicas há muito inventadas, como sublinhou Lancaster, e assumindo estudos infométricos em ambiente digital ou *webmétricos*: é impossível trabalhar e estudar a informação, sem ter em conta o sofisticado meio ou suporte onde ela hoje e no futuro se encontra registada (SILVA, 2007, p. 42).

Outra definição de SI parte de Land (1985, p. 215), que o descreve como “[...] um construto social, embutido com tecnologia de informação”. Essa afirmação dá ênfase, além do aspecto social, ao aspecto tecnológico de sistemas de informação, pois ainda que esses sistemas sejam baseados na interação de pessoas, organizações, tecnologia e informações de modo interdependente, o ambiente digital e as tecnologias vêm tendo importância crescente nesses sistemas sendo cada vez mais presentes nas interações humanas. Marques (2017) menciona uma tendência de que parte significativa da gestão de SI seja orientada ao uso de tecnologia, ainda que a dimensão humana seja mais valiosa.

Headrick (2000) ressalta que na história humana ocorrem períodos em que o volume de informações a que as pessoas têm acesso apresenta elevada evolução, cabendo aos sistemas de informação apresentar soluções para lidar com essa necessidade. Na chamada “era da informação”, caracterizada pela emergência de informações complexas e interligadas, a utilização de tecnologias é uma tendência para SI (RAYWARD, 2014, p. 682).

O enfoque principal da CI é a dimensão humana e social da informação, de modo que a aplicação crescente de TIC como parte componente dos sistemas de informação passa a ser de interesse da CI transdisciplinar, com foco “no modo como o conjunto estruturado de representações mentais e emocionais codificadas (informação) é ajustado à tecnologia” (SILVA, 2007, p. 39).

Portanto, os sistemas de informação geram, coletam, armazenam, processam e disseminam dados, informações e conhecimento (SCHULZ, 2001). Além disso, a integração de dados entre diferentes sistemas, como citam Cruz e Nagano (2008), passaram a ser uma tendência em SI pela facilitação da transferência de informações e conhecimento, além de favorecer o processo de codificação deste último.

Considerando as diversas definições de vários autores, alguns aqui citados, Alter (2008) propôs um modelo de SI observando as características em comum dessas definições, entendendo SI como um sistema de trabalho no qual participantes – humanos e/ou máquinas – executam trabalho – processos ou atividades – utilizando informação, tecnologia e outros recursos para produção de produtos informacionais internos ou externos a uma organização. Seriam assim elementos do modelo: consumidores de informação; produtos e serviços informacionais; processos e atividades; participantes; informação; e tecnologia (Figura 3).

Figura 3 – Modelo de Sistema de Informação como Sistema de Trabalho



Fonte: Alter (2008, p. 461).

O modelo proposto apresenta uma visão voltada a ambientes organizacionais com foco na produção de produtos e serviços informacionais e na participação social dos integrantes da organização, apresentando uma visão sistêmica em oposição à visão instrumental muitas vezes comum na concepção de SI (ALTER, 2008).

E no governo aberto, como se observa a contribuição dos sistemas de informação? Na próxima seção apresenta-se reflexão sobre essa questão.

2.3 GOVERNO ABERTO E OS SISTEMAS DE INFORMAÇÃO GOVERNAMENTAIS

Uma organização, conforme descreve Maximiano (1992), “[...] é uma combinação de esforços individuais que tem por finalidade realizar propósitos coletivos”. Assim, são consideradas organizações órgãos do governo, empresas, associações, entidades públicas e privadas e diversas outras entidades.

Outra definição é dada por Motta (1971), que explana que “[...] organizações são uma classe de sistemas sociais, que por sua vez constituem uma classe de sistemas abertos”. Isso é fundamentado se considerarmos que organizações são compostas por diversos componentes humanos e materiais, recebendo entradas do ambiente (tais como mão-de-obra, matéria-prima etc.) e produzindo saídas (produtos), e funcionando em torno de um objetivo final.

Ainda segundo Motta (1971), “a organização se distingue, porém, dos demais sistemas sociais pelo seu alto nível de planejamento”, incluindo alto nível de controle em suas atividades, pressões ambientais e valores e expectativas controladas. É nesse contexto que os Sistemas de Informação são de grande valia às organizações como provedores de recursos informacionais e gestão de processos organizacionais (GRUZOVA, 2012).

Como expõem Silva, Santos e Konrad (2016), a necessidade dos Sistemas de Informação é cada vez maior devido ao “grande e crescente volume de informações que a organização possui”, vindo de encontro à valorização da informação e do conhecimento característica da era da informação.

A finalidade do Sistema de Informação Organizacional é, assim, relacionada principalmente ao processo decisório e à resolução de problemas, atuando a partir da “captura e/ou a recuperação de dados e sua análise em função

de um processo de decisão” (PEREIRA; FONSECA, 1997, p. 241 *apud* SILVA; SANTOS; KONRAD, 2016, p. 9).

Em organizações privadas, a utilização de sistemas de informação visa principalmente o aumento da produtividade e a adaptação às condições dinâmicas de mercado, podendo prover um diferencial mercadológico (SILVA; SANTOS; KONRAD, 2016). Nas organizações públicas, os sistemas de informação também objetivam o aumento da eficiência organizacional, nesse caso focando em prover melhores serviços ao público, transparência de informações e ações e melhoria na prestação de contas e responsabilização (*accountability*) (CORDELLA; IANNACCI, 2010).

O termo Governo Aberto refere-se a um modo de administração pública que promove projetos e ações pautadas na transparência, prestação de contas, participação cidadã, e tecnologia e inovação (BRASIL, 2020a). Esse conceito surge a partir do movimento *Open Government Data* (OGD), que visa governos como sistemas abertos capazes de cooperar e interagir com seus ambientes a partir da utilização de dados abertos (PARYCEK; HOCHTL; GINNER, 2014) – que conforme explicam Bazilian *et al.* (2012), são dados caracterizados por sua livre disponibilização e uso irrestrito por qualquer interessado.

Esse paradigma modifica, além do acesso a dados e informações, as barreiras entre o público e as instituições governamentais (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012). Huang, Wang, Zhang, Wu e Xie (2019) consideram o movimento OGD um fenômeno mundial, com legislações específicas sobre o tema sendo criadas na Europa e Estados Unidos, havendo também acordos internacionais como o assinado pelo fórum político do Grupo dos Oito (G8).

O Brasil, como citam Bertin, Fortaleza, Silva e Okawachi (2019), é um dos oito países cofundadores da iniciativa *Open Government Partnership* (OGP), cujo objetivo é difundir e incentivar, globalmente, práticas governamentais relacionadas à transparência dos governos, ao acesso à informação pública e à participação social (Open Government Partnership, 2011). O Brasil desenvolveu quatro Planos de Ação Nacional em Governo Aberto (BRASIL, 2020b), além de possuir legislações específicas para transparência de informações, como a Lei nº 12.527, de 18 de novembro de 2011 (lei de acesso à informação) e o Decreto nº 8.777, de 11 de maio de 2016 (que institui a Política de Dados Abertos do Governo Federal) (BRASIL, 2020c).

Para Barbalho (2018, p. 129), políticas de dados abertos no Brasil ainda consistem em um campo de ação estratégica emergente, no qual as iniciativas ainda não são suficientes para garantir a estabilidade, havendo “a necessidade de um arcabouço de governança que proporcione previsibilidade legal ao campo, rotinize a abertura de dados por parte dos órgãos de governo e viabilize a sustentabilidade de projetos”. O autor ainda identifica três grandes temas relacionados às políticas de dados abertos no Brasil: a ressignificação (utilização com outros propósitos) dos dados abertos; a influência do ambiente normativo internacional, representado principalmente pela OGP; e as dinâmicas associadas à posse e ao uso dos capitais observados em políticas públicas de dados abertos.

As políticas públicas de Governo Aberto, segundo Cordella e Iannacci (2010), utilizam TIC como facilitadores para o desenvolvimento de sistemas de informação governamentais. Esses sistemas e as TIC associadas são moldados a partir das metas e objetivos da organização, tencionando a realização de mudanças políticas e organizacionais dentro das entidades públicas e de suas relações com a sociedade, provendo maior transparência e eficiência.

Esses sistemas de informação cujo objetivo é a entrega de serviços estatais à sociedade e ampliação da participação cidadã na política associa-se diretamente ao modelo de gestão estatal nomeado como governo eletrônico, governança eletrônica, ou *e-government* (e-Gov), cujo objetivo é “melhorar a gestão de procedimentos e serviços públicos característicos da face governamental da Sociedade da Informação” (BARRETO JUNIOR; RODRIGUES, 2013, p. 93), a partir da utilização de TIC.

Os sistemas providos pela administração pública a partir dos portais de governo podem ter níveis de interação variados, existindo cinco níveis ou estágios diferenciados (FERNANDES, 2006): o primeiro estágio, denominado emergente, corresponde à simples criação de sítios de informação sobre atividades governamentais; no segundo estágio (destacado) há a atualização constante de conteúdo e informações; o estágio interativo permite a obtenção de formulários, contactar funcionários e encaminhar solicitações; no estágio transacional tornam-se viáveis transações financeiras e pagamentos; por fim, no estágio contínuo o cidadão possui integrado aos serviços e informações de governo eliminando fronteiras administrativas e departamentais.

A implementação de portais de governo eficientes, como destacam Araújo

e Laia (2004), envolve: a análise das necessidades de usuários, envolvendo processos e estratégias propícias de busca de informação e acessibilidade; cooperação entre poderes e esferas governamentais por meio da integração de informações; alinhamento da tecnologia de informação às metas e objetivos dos portais; desenvolvimento de culturas e comportamentos informacionais positivos pelos membros da organização pública; e definição adequada de conteúdo.

Sistemas de informação governamentais, conforme os objetivos de Governo Aberto, envolvem a aquisição, processamento, classificação, gerenciamento e disseminação de dados, sendo um dos pontos cruciais para sua operação a integração de dados de fontes diversas (HUANG; WANG; ZHANG; WU; XIE, 2019).

Para Rodrigues (2012), a utilização das TIC pelo Estado divide-se em uso interno – para o auxílio de processos administrativos internos – e uso externo – na relação entre a organização pública e entidades externas, tais como empresas, cidadãos, grupos de cidadãos e outras organizações públicas. A utilização de tecnologias voltadas ao Governo Aberto se enquadra dessa forma no uso externo.

Sant’ana (2008, p. 15) destaca que o uso externo das TIC “representam um novo desafio e podem apresentar um conjunto de características vantajosas no que diz respeito a interação entre a gestão pública e a sociedade”, ressaltando a possibilidade de alteração da relação entre os dois entes em um modelo mais eficiente e direto, envolvendo a sociedade na gestão pública a partir da transparência e responsabilidade nas ações e na tomada de decisões na gestão pública.

Rodrigues (2012, p. 4), uma década atrás, alertava quanto à utilização de TIC no âmbito governamental para o fato de que “alguns órgãos governamentais ainda as utilizam principalmente com a função de quadro de aviso eletrônico, deixando em segundo plano a sua capacidade de interação mais eficiente entre Estado e sociedade”.

Fachin e Rover (2015) apontam para uma mudança nesse cenário, ressaltando a incorporação crescente de tendências e tecnologias nas atividades organizacionais do governo eletrônico, com vista ao aumento da eficiência dos serviços públicos, destacando ainda a importância da inovação e manutenção para garantia da qualidade e usabilidade desses serviços.

Para Sant’ana e Rodrigues (2013), a relação entre Estado e sociedade

pode tornar-se mais eficiente a partir da exploração do potencial do uso externo das TIC, reduzindo a assimetria de informações entre as duas partes e proporcionando condições para a reflexão crítica, mobilização de interesses coletivos e aumentando a repercussão de ações governamentais.

O conceito de Governo Aberto relaciona-se ainda com os Objetivos de Desenvolvimento Sustentável (ODS), agenda de desenvolvimento estabelecida pela Organização das Nações Unidas (ONU) em 2015 em conjunto com seus países-membros e representantes da sociedade civil, que consistem de 17 objetivos para o desenvolvimento econômico sustentável (ONU, 2020). O Objetivo 16 – promover sociedades pacíficas e inclusivas para o desenvolvimento sustentável, proporcionar o acesso à justiça para todos e construir instituições eficazes, responsáveis e inclusivas em todos os níveis – prevê o desenvolvimento de instituições eficazes e transparentes, a tomada de decisão responsiva e participativa, bem como o acesso público à informação (ONU, 2020).

No Brasil, é possível perceber um avanço na proposição de mecanismos legais para a promoção da transparência governamental, tendo como marco a Lei Complementar nº 131 que determina a disponibilização de informações sobre todos os gastos e receitas públicas federais, estaduais e municipais, levando à criação de portais eletrônicos para divulgação dessas informações (MEDEIROS; NONATO; KATZ; SOUZA, 2013).

Também representa um marco a criação da Lei de Acesso à Informação (LAI), estabelecida pela Lei Federal nº 12.527 de 2012, cujo objetivo consistiu em “permitir acesso a informações públicas, buscando reverter o quadro de baixa transparência das informações, ao permitir seu acesso de forma mais direta que os meios até então existentes, favorecendo a transparência por meios digitais” (MEDEIROS; NONATO; KATZ; SOUZA, 2013, p. 455).

Para Gama (2015), há investimento crescente no desenvolvimento de SI no Governo Federal para melhoria da gestão e aumento do controle social, resultando no fortalecimento do Governo Eletrônico (e-Gov).

Nas democracias representativas, segundo Fernandes (2012), a participação social no controle e discussões sobre serviços do Estado, ainda que parcial, depende da criação de ambientes democráticos, o que pode ser alcançado principalmente a partir da ampliação da transparência e da cidadania, sendo um dos meios para alcançar esse objetivo a utilização de TIC para acesso a dados

governamentais. Como afirmam Medeiros, Nonato, Katz e Souza (2013, p. 454), “os atos dos agentes públicos devem ser devidamente informados à sociedade, para que esta seja capaz de exercer o controle social”.

Gama (2015) alerta, no entanto, que ainda existe uma cultura dentro da administração pública de limitar o acesso às informações governamentais devido a receios de críticas e obstrução de projetos. Também existem problemas na implementação dos mecanismos legais de transparência, por fatores como a falta de gestão eficiente de informações governamentais, baixa interoperabilidade de sistemas e bases de dados, baixa confiabilidade de dados existentes e falta de recursos humanos para atendimento às demandas informacionais (DAMATTA, 2011).

O estudo das relações entre as TIC e configurações sociais e organizacionais que compõem o sistema, da qualidade das informações disponibilizadas, da cultura de abertura de informações governamentais e das interações entre atores e dados do SI são questões de grande importância nesse cenário.

Considerando o uso cada vez mais comum de técnicas e ferramentas da Ciência de Dados nas TIC, torna-se pertinente analisar os conceitos dessa disciplina, conforme será abordado a seguir.

3 CIÊNCIA DE DADOS

A Ciência de Dados é uma disciplina relacionada à Ciência da Computação e Estatística que na última década vem recebendo destaque por parte dos meios científicos e profissionais, relacionada à prática de análise de dados digitais – representativos de fenômenos de interesse a pessoas ou grupos – em ampla quantidade e em altas velocidades, sustentada a partir de avanços tecnológicos e baixos custos de armazenamento de dados (PRIESTLEY; McGRATH, 2019).

Lin Wang (2018, p. 1246) acredita que a CI e a Ciência de Dados possuem amplas possibilidades de convergência, a partir do foco de ambas as disciplinas na extração de valor e na compreensão de dados, bem como na organização e utilização de conhecimento. Souza, Almeida e Baracho (2015, p. 160) também veem possibilidades de mudanças das fronteiras epistemológicas e paradigmáticas da CI a partir do “advento de novas tecnologias, construtos e formas de manuseio da informação”, citando em seu trabalho tecnologias relacionadas à Ciência de Dados.

Como versam Rautenberg e Carmo (2019), os avanços e uso massivo das TIC e as transformações sociais decorrentes instigam a discussão interdisciplinar pela CI da utilização de dados, informação e conhecimento gerados digitalmente.

Essa conjuntura convida à análise da relação entre as duas áreas, partindo dos conceitos epistemológicos e práticos da Ciência de Dados, conforme abordado a seguir.

3.1 SOBRE A CIÊNCIA DE DADOS

Cao (2016) explana que, em um contexto científico, a Ciência de Dados desenvolve metodologias, teorias e tecnologias para a captura, representação, armazenamento, busca, análise e apresentação de dados complexos, heterogêneos e de múltiplas fontes para suporte à tomada de decisão em tempo real.

Bufrem, Silva, Sobral e Correia (2016) destacam as mudanças no fazer da ciência ao longo dos anos e a influência mútua entre ciência e tecnologia, verificando-se em tempos recentes a implementação/evolução do compartilhamento

de dados por parte dos cientistas. Os autores destacam a Ciência Orientada a Dados (COD), como um dos avanços concretos nessa área.

A COD, ou *Data Science* (Ciência de Dados), tem origem em 2001, e “relaciona-se ao (campo) da Estatística e incorpora nessa relação, temas por ela abrangidos, como pesquisas multidisciplinares, modelos e métodos para dados, computação de dados, pedagogia, ferramenta de avaliação e teoria” (BUFREM; SILVA; SOBRAL; CORREIA, 2016, p. 41).

Conforme alerta Lin Wang (2018), ainda não há uma definição precisa sobre a Ciência de Dados. Dhar (2013) compara a Ciência de Dados à Estatística, ciência que estuda de forma sistemática a organização, propriedades e a análise de dados gerando inferências, diferenciando-se desta última a partir das características heterogênea e desestruturada dos dados analisados pela Ciência de Dados, além do foco na interpretação automática desses dados. Outros autores como Gelman *et al.* (2014 apud LIN WANG, 2018) a interpretam como o estudo do processamento, análise e interpretação de dados. Em linha similar, Provost e Fawcett (2013, p. 52) definem a Ciência de Dados como um “conjunto de princípios que apoiam e guiam a extração de informações e conhecimento a partir de dados”.

Lin Wang (2018) afirma haver algum consenso quanto às principais características da Ciência de Dados, que consistiria em um campo interdisciplinar emergente dedicado à identificação e extração de padrões a partir de dados computacionais, convertendo dados em informação e conhecimento através de análise ou mineração de dados.

Para Blei e Smyth (2017), a Ciência de Dados é a união da Estatística com a Ciência da Computação, pois além de herdar pensamentos e métodos das duas áreas, busca combiná-los e adaptá-los às necessidades atuais de análise de dados. Para os autores, pode-se abordar a Ciência de Dados a partir de três perspectivas distintas: a estatística, que forma a base técnica de análise de dados; a computacional, focada na construção de algoritmos e na implementação dos métodos estatísticos; e a humana, que depende das pessoas para o entendimento de contextos de dados e para a aplicação de ferramentas estatísticas e computacionais.

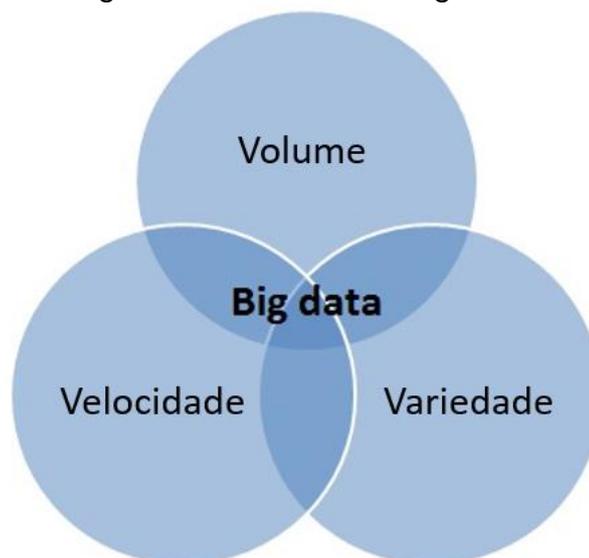
De forma similar, Cleveland (2014) destaca a relação da Ciência de Dados com a Estatística, a partir da utilização e expansão de métodos característicos da última, situando também a disciplina no contexto da Ciência da

Computação, pelo uso intensivo de algoritmos computacionais no seu desenvolvimento e atuação, assim como a associa às pesquisas em mineração de dados. Apesar disso, Cleveland destaca que a Ciência de Dados tem potencial para envolver mais áreas, pois a análise de dados envolve mais do que apenas tecnologia e o foco apenas no aspecto computacional limita os resultados da área, de forma que sugere que os trabalhos da área devem dedicar-se à investigação multidisciplinar, no desenvolvimento de modelos e métodos para dados e na construção de teorias.

Outro conceito normalmente associado à Ciência de Dados, por vezes confundido com a área (CAO, 2017), é o fenômeno de *big data*. Lnenicka e Komarkova (2019) afirmam ainda haver pouco consenso em relação ao significado do termo, mas segundo os autores há algum consenso quanto ao *big data* referir-se a conjuntos de dados que não podem ser coletados, manuseados e processados por TIC e *softwares* tradicionais em um limite de tempo tolerável, listando as seguintes características vinculadas, como pode ser observado na Figura 4:

- Volume: no *big data* há uma grande variedade de dados disponíveis
- Variedade: existem diferentes tipos de dados em um conjunto
- Velocidade: os dados são gerados em grandes velocidades

Figura 4 – Os “3Vs” do *Big Data*



Fonte: Big Data (2021, não paginado).

Os autores ainda ressaltam que os avanços tecnológicos e contextos

organizacionais, sociais e ambientais tornam difícil mensurar quais conjuntos de dados podem ser considerados *big data*.

Para Bertot *et al.* (2014), o *big data* precisa de três componentes de infraestrutura: uma plataforma para organizar, armazenar e tornar possível o acesso a dados; tecnologia e poder computacionais para processamento de conjuntos de dados de larga escala; e formato de dados estruturados e passíveis de uso.

Devido ao uso por governos e pela disponibilização de dados governamentais em grande escala, o *big data* vem sendo associado ao movimento de Dados Abertos, criando um novo termo denominado *big open data* (BERTOT *et al.*, 2014; LNENICKA; KOMARKOVA, 2019), que consiste na disponibilização de grandes conjuntos de dados governamentais.

Fayyad e Hamutcu (2020) situam o *big data* como uma das áreas de interesse da Ciência de Dados, assim como outras que serão abordadas adiante.

Considerando a mencionada interdisciplinaridade vinculada à Ciência de Dados, aborda-se a seguir esse conceito em conjunto com as disciplinas relacionadas à área.

3.2 RELAÇÕES INTERDISCIPLINARES

Se a disciplinaridade consiste na exploração científica especializada em um domínio homogêneo de estudo, a partir de fronteiras constituintes, Japiassu (1976) define a interdisciplinaridade como a interação entre disciplinas, havendo axiomática comum a um grupo de disciplinas conexas que é definida e tem sua finalidade determinada pelo nível hierárquico superior.

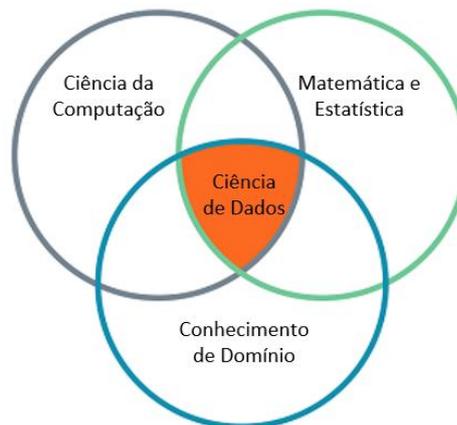
Ainda que pareça um conceito relativamente simples, o autor destaca que o mesmo apresenta variações tanto em termos de nomenclatura quanto em relação ao seu conteúdo, destacando assim modalidades de interdisciplinaridade: heterogênea (baseada na “soma” de informações de diversas disciplinas); pseudo-interdisciplinaridade (utilização de instrumentos epistemologicamente “neutros” para associação de disciplinas); auxiliar (“empréstimo” de métodos ou procedimentos de uma disciplina para aplicação em outra); compósita (conjugação de disciplinas por aglomeração para resolução de problemas complexos); e unificadora (coerência bastante estreita dos domínios de estudo das disciplinas). A interdisciplinaridade pode ocorrer a partir da simples permutação de informação entre disciplinas de

modo linear ou a partir de um processo de trocas recíprocas, uma combinação das disciplinas que levam ao surgimento de um novo campo.

A interdisciplinaridade da Ciência de Dados é destacada por autores como George, Osinga, Lavie e Scott (2016), que a caracterizam como tal por se relacionar aos campos da Estatística, mineração de dados, aprendizado de máquina e análise de dados, e Chao (2016 apud LIN WANG, 2018), que a definem como um campo preocupado com o processamento, gerenciamento, análise e desenvolvimento de produtos derivados do estudo de dados.

Como abordado anteriormente, a Ciência de Dados tem forte relação com a Estatística e a Ciência da Computação, representando as relações interdisciplinares que podem ser consideradas mais fortes (Figura 5). Para Asamoah, Doran e Schiller (2015), considerando que o objetivo da disciplina é o uso de dados para resolução de problemas práticos, a Ciência de Dados teria assim que combinar habilidades estatísticas e de programação de computador com conhecimento relevante do domínio para o qual pretende-se realizar a análise.

Figura 5 - Principais Relações Interdisciplinares da Ciência de Dados



Fonte: Bernabe (2017, não paginado).

Estudo realizado por Virkus e Garoufallou (2019) parece corroborar tal afirmação. A partir de análise bibliográfica de publicações científicas relacionadas à Ciência de Dados, os autores encontram artigos de áreas diversas, como Ciência da Computação, Medicina, Administração, Economia, Direito, Geografia, Astronomia, Física, Ciências Sociais, dentre outras, entre as quais a Ciência da Informação. De modo similar, Rautenberg e Carmo (2019) afirmam que há três domínios de

conhecimento essenciais à disciplina: programação de computadores; estatística e matemática; e domínio do conhecimento. Assim, a Ciência de Dados apresentaria relação interdisciplinar com três áreas principais, Ciência da Computação, Matemática e Estatística e Conhecimento do Domínio.

Nesses casos, considerando as definições do termo interdisciplinaridade apresentadas, entende-se que as relações entre a Ciência de Dados e as disciplinas relacionadas ao conhecimento pertinente à análise de dados é uma interdisciplinaridade auxiliar.

Em relação à Ciência da Informação, percebe-se a possibilidade de relações interdisciplinares mais completas considerando o objetivo principal da Ciência de Dados, a produção de informação e conhecimento a partir da análise de dados (PROVOST; FAWCETT, 2013), que envolve o principal objeto de estudo da CI, a informação em todas suas formas.

Virkus e Garoufallou (2020) reconhecem essa possibilidade a partir de análise de conteúdo de publicações da Ciência de Dados pela perspectiva da CI, entendendo que tópicos como a educação e treinamento em Ciência de Dados, curadoria digital, Gestão do Conhecimento, o processo de análise de dados e gestão da qualidade de dados são citadas por vários autores como pontos de intersecção entre as disciplinas.

Agarwal e Dhar (2014) relacionam ainda a Ciência de Dados à disciplina de Sistemas de Informação, destacando o papel histórico desta última na condução de pesquisas relacionando tecnologia e dados a atividades organizacionais e à sociedade, preocupando-se com aspectos tais como o armazenamento, gerenciamento e processamento de dados, assim como sua estruturação. Para os autores, a natureza interdisciplinar e transdisciplinar de SI em conjunto com as perspectivas, técnicas e ferramentas desenvolvidas pela área representam uma vantagem para a compreensão das questões relacionadas à Ciência de Dados.

Assim, pode-se perceber que a Ciência de Dados tem a interdisciplinaridade como uma de suas características principais, havendo, como citam Blei e Smyth (2017) e Cleveland (2014), potencial para a expansão de seus limites interdisciplinares, tendo relação com algumas áreas de estudo apresentadas no tópico seguinte.

Analisa-se a seguir as áreas de estudo relacionadas à Ciência de Dados.

3.2.1 Áreas de estudo relacionadas

Cao (2017) ressalta que vários termos, como *analytics*, *big data* e Ciência de Dados, estão extremamente interconectados e são muitas vezes confundidos entre si. Para o autor, a Ciência de Dados é a denominação de um campo de estudo, o termo *analytics* refere-se à aplicação de teorias e ferramentas a fim de permitir a análise de dados complexos e obter conclusões a partir desses dados, enquanto *big data* é uma massa de dados vasta e/ou complexa de tal forma a não ser possível interagir com ela a partir de teorias, tecnologias ou métodos tradicionais.

É possível ainda fazer paralelos entre Ciência de Dados e *e-Science*, ou e-Ciência. Medeiros e Caregnato (2012, p. 312) definem *e-Science* como uma “infraestrutura que visa permitir que cientistas e pesquisadores possam ter acesso a dados científicos primários distribuídos”, mas capaz de prover algo além da simples infraestrutura, de modo a facilitar o gerenciamento e compartilhamento de dados científicos estimulando a colaboração científica, e permitindo ainda o tratamento e análise de dados “já existentes e compartilhados, reduzindo custos e esforços e possibilitando avanços efetivamente mais rápidos” (MEDEIROS; CAREGNATO, 2012, p. 316).

Como explicam Monteiro, Sant’ana e Santarém Segundo (2016, p. 15), a *e-Science* foi desenvolvida a partir da necessidade de interagir com um grande volume de dados de fontes diferentes, sendo a “ciência intensiva em dados” – caracterizada pela integração, fusão e mineração de dados, permitindo sua captura e processamento em torno de um objetivo – um de seus componentes.

A partir das definições e características apresentadas de *e-Science*, percebe-se que a mesma pode ser qualificada como uma aplicação voltada a fins científicos da Ciência de Dados.

Misnevs e Yatiskv (2016), em estudo analisando os papéis profissionais dos cientistas de dados (praticantes da disciplina da Ciência de Dados), percebem que as atividades e habilidade relacionadas a esse campo envolve o conhecimento de áreas como o aprendizado de máquina (*machine learning*), mineração de dados, *big data*, programação de computador e infraestrutura de dados, dentre outras.

Para Fayyad e Hamutcu (2020), ainda há alguma confusão em relação às ferramentas, métodos e habilidades associadas à Ciência de Dados. No entanto, os

autores expõem que áreas como aprendizado de máquina, matemática, programação de computador, *analytics*, *business intelligence*, bancos de dados, mineração de dados, estatística e *big data* são comumente associadas à Ciência de Dados.

Abordando alguns dos conceitos principais relacionados à Ciência de Dados, temos que o aprendizado de máquina é uma disciplina da Inteligência Artificial (IA), que busca o desenvolvimento de sistemas de computador capazes de autoaperfeiçoamento a partir de experiência e de teorias relacionadas à estatística, computação e informação capazes de explicar o aprendizado em computadores, seres humanos e organizações (JORDAN; MITCHELL, 2015).

Bernabe (2017) afirma que o aprendizado de máquina envolve o uso de computação baseada em estatística para inferir tendências, padrões, características e atributos a conjuntos de dados, permitindo observações e previsões a partir deles. Para Jordan e Mitchell (2015), o aprendizado de máquina vem sendo usado para o desenvolvimento de *softwares* em diversas áreas da tecnologia e ciência, com aplicações na robótica, reconhecimento de linguagem, processamento de linguagem natural, entre outras.

Também associadas à Ciência de Dados são os termos *Business Intelligence* (BI) e *analytics*. Para Wang (2016), o BI é uma ferramenta tecnológica composta de arquiteturas e técnicas capazes de transformar dados brutos em informação útil para suporte à decisão em organizações. O conceito é similar ao de *analytics* proposto por Cao (2017) e anteriormente mencionado, no qual utilizam-se tecnologias e teorias para análise de dados complexos e obtenção de conclusões. Chen, Chiang e Storey (2012) correlacionam as duas áreas, definindo *Business Intelligence & Analytics* como um processo de coleta, extração, análise e visualização de dados baseado no uso de tecnologias, estas comumente concebidas a partir de métodos estatísticos e técnicas de mineração de dados.

Já a mineração de dados é um subcampo da Ciência da Computação que envolve o processamento, exploração e análise de conjuntos de dados visando a descoberta de padrões (JAIN; SRIVASTAVA, 2013), sendo parte de um processo mais amplo denominado *Knowledge Discovery in Databases* (KDD), que tem o mesmo objetivo (BARTSCHAT; REISCHL; MIKUT, 2019).

A utilização de tecnologias de mineração de dados busca a criação de modelos a partir dos dados, que podem ser descritivos – que apresenta de maneira

sucinta as características de um conjunto de dados – ou preditivos – que permitem prever o valor futuro de uma variável de um conjunto de dados –, buscando a resolução de problemas a partir de dados pré-existentes em bases de dados (JAIN; SRIVASTAVA, 2013).

Devido à inexistência de um currículo padrão para a Ciência de Dados (MISNEVS; YATISKV, 2016), não é possível precisar todas as áreas vinculadas à disciplina. Entretanto, entende-se que as aqui citadas são as mais comumente associadas a esse campo.

As áreas aqui citadas são aplicadas em processos a atividades desenvolvidas em ambientes organizacionais, como será apresentado a seguir.

3.3 IMPLICAÇÕES DA CIÊNCIA DE DADOS EM AMBIENTES ORGANIZACIONAIS

Dhar (2013) destaca que o crescimento exponencial do volume de dados disponíveis na *internet* e nos próprios ambientes organizacionais, assim como a heterogeneidade desses dados, levou à adoção de ferramentas de mineração de dados e aprendizado de máquina por parte das organizações de modo a tornar viável o processo de tomada de decisões a partir de dados.

Provost e Fawcett (2013) relacionam a Ciência de Dados a uma série de processos organizacionais relacionados a dados, como a tomada de decisões relacionada a dados (o embasamento de decisões organizacionais a partir de dados), processamento de dados voltado a extração de conhecimento e suporte à decisão (pois como os autores destacam, nem todo processamento de dados envolve Ciência de Dados) e o suporte ao pensamento analítico direcionado a dados (desenvolvimento de habilidades no ambiente organizacional que permitam a compreensão de aspectos básicos relacionados a dados).

No setor público, a crescente demanda por transparência e provimento de serviços governamentais resulta no aumento do uso de tecnologias. No entanto, o uso de tecnologias de Ciência de Dados ainda se encontra em estágios iniciais (ELEZAJ; TOLE; BACI, 2018).

Países como Canadá, China, Estados Unidos, Estônia e Holanda (FERNEDA *et al.*, 2016; WAGNER, 2013) possuem Sistemas de Informação Governamentais com altos graus de maturidade, provendo serviços e informações a

seus cidadãos a partir de aplicações de tecnologias para Governo Eletrônico com interoperabilidade, padrões de dados e transparência.

A interoperabilidade, como explicam Sayão e Marcondes (2008), se refere ao compartilhamento de conteúdo de fontes distintas. Ferneda *et al.* (2016) entendem que a interoperabilidade é a capacidade mútua de duas entidades em realizar trocas e usar informação, ao invés de apenas dados.

Elezaj, Tole e Baci (2018), assim como Ferneda *et al.* (2016), argumentam que a interoperabilidade é um elemento imprescindível para as operações de e-Gov, não só em termos tecnológicos, como também políticos e sociais integrando pessoas, processos e dados. Para Wagner (2013), a colaboração entre servidores públicos, o fornecimento de informações aos cidadãos e a Gestão do Conhecimento são igualmente importantes.

Essas necessidades, que se encontram com os objetivos de transparência, efetividade e satisfação do público que permeiam o conceito de Governo Eletrônico, podem ser atingidas a partir de tecnologias da Ciência de Dados, como argumentam Elezaj, Tole e Baci (2018). Para os autores, tais tecnologias podem ajudar a prover abertura de dados governamentais, no suporte ao engajamento digital, na tomada de decisões para desenvolvimento de políticas e no provimento de serviços digitais voltados à inovação.

Al-Mushayt (2019) sugere que sistemas de e-Gov podem tornar-se mais eficientes e criar maiores interações com o público a partir do uso de técnicas de IA, especificamente de aprendizado de máquina (a partir de algoritmos de deep learning, uma subárea do aprendizado de máquina). Tal abordagem, segundo o autor, poderia permitir o desenvolvimento de funcionalidades tais como a centralização dos recursos informacionais de e-Gov, a conversão de documentos analógicos para digitais, a análise de sentimentos dos cidadãos relacionados a determinado serviço governamental, a indexação automática de arquivos governamentais, a personalização de serviços para cidadãos e a integração de informações.

Apesar desses possíveis benefícios, poucos países adotam serviços baseados em aprendizado de máquina, porém havendo projetos para inclusão dessas tecnologias em serviços de e-Gov em países do Golfo Pérsico, ainda de acordo com Al-Mushayt (2019). O autor destaca ainda que questões éticas e de privacidade como compartilhamento de informações pessoais e reconhecimento

facial devam ser consideradas na adoção de tecnologias.

Outra aplicação possível seria a utilização de tecnologias de mineração de dados. Verma e Shrivastava (2015) listam possíveis usos para a mineração de dados em bases de dados governamentais, tais como análises demográficas para suporte a políticas públicas, descoberta de necessidades públicas locais, suporte à tomada de decisões para políticas de saúde, ambientais, urbanas e de segurança públicas, entre outras. Para os autores, há ainda a possibilidade do uso da mineração de dados por parte do público, aplicada a dados governamentais abertos, que podem ter utilidade para controle social e usos particulares.

Por sua vez, Matheus, Janssen e Maheshwari (2018) evidenciam a possibilidade da Ciência de Dados, através da utilização de painéis (*dashboards*) como instrumento de apresentação de informações – ferramenta tecnológica geralmente associada a *analytics* –, atuar não só no processo de tomada de decisões como também com o objetivo de prover transparência de informações. Considerando a conjuntura governamental, os autores argumentam que a partir desses painéis é possível influenciar o processo de construção de políticas públicas e gerar impactos positivos na interação com o público.

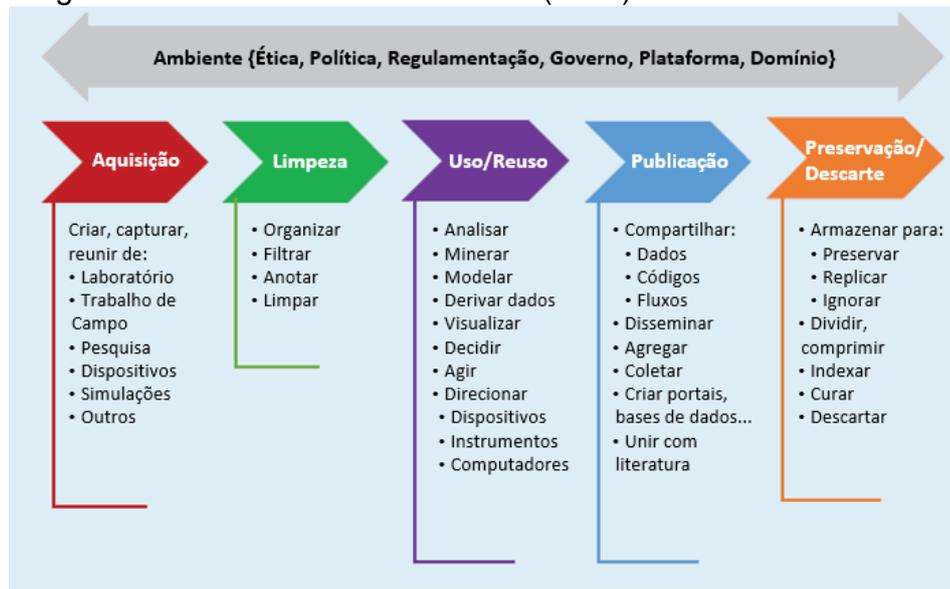
Percebem-se assim várias implicações possíveis do uso de tecnologias da Ciência de Dados em ambientes organizacionais, especialmente relacionados ao setor público.

No entanto, é relevante a compreensão de como os dados podem ser utilizados, analisados e compreendidos, como será abordado na próxima seção, a partir do Ciclo de Vida de Dados.

3.4 CICLO DE VIDA DOS DADOS

Para Berman *et al.* (2018), a utilização de dados depende da existência de um suporte físico, contexto social e funções para seu uso. Para os autores, o Ciclo de Vida dos Dados (CVD) é crítico para o entendimento de conjuntos de dados e sua utilização, compreendendo etapas de aquisição, limpeza (tratamento e adequação de dados), uso/reuso, publicação e preservação/descarte, como mostrado na Figura 6.

Figura 6 – Ciclo de Vida dos Dados (CVD)



Fonte: Berman *et al.* (2018).

Ainda de acordo com Berman *et al.* (2018), o uso de um ciclo de vida permite a criação de um ecossistema de dados, permitindo sua organização, gerenciamento e a padronização de ações para transformação de dados em conhecimento. Embora concebidas de forma sistemática, cada etapa do ciclo costuma ser executada de forma isolada por áreas de conhecimento com maior interesse em uma ou mais etapas.

A Ciência da Computação, Ciência da Informação, Estatística e áreas do domínio do conhecimento dos dados atuam em etapas particulares do CVD, havendo possibilidades pouco exploradas de integração entre elas que podem permitir trocas benéficas para utilização de dados (BERMAN *et al.*, 2018).

Sant'ana (2016) afirma que a CI pode contribuir para o desenvolvimento do cenário de acesso e uso intensivo de dados, propondo a estruturação da análise a partir do uso do Ciclo de Vida dos Dados por ele proposto em sua pesquisa, evidenciando os fatores e etapas envolvidas no processo. As fases do ciclo de vida proposto pelo autor são: coleta; armazenamento; recuperação; e descarte. Essas fases relacionam-se a fatores do CVD, presentes em todas elas, que são: privacidade; integração; qualidade; direitos autorais; disseminação; e preservação.

A fase de coleta de dados diz respeito à obtenção dos dados para uma necessidade informacional específica a partir do desenvolvimento de um plano de ação, análise de viabilidade e da execução da coleta, sendo necessárias

competências específicas relacionadas às necessidades informacionais para sua realização.

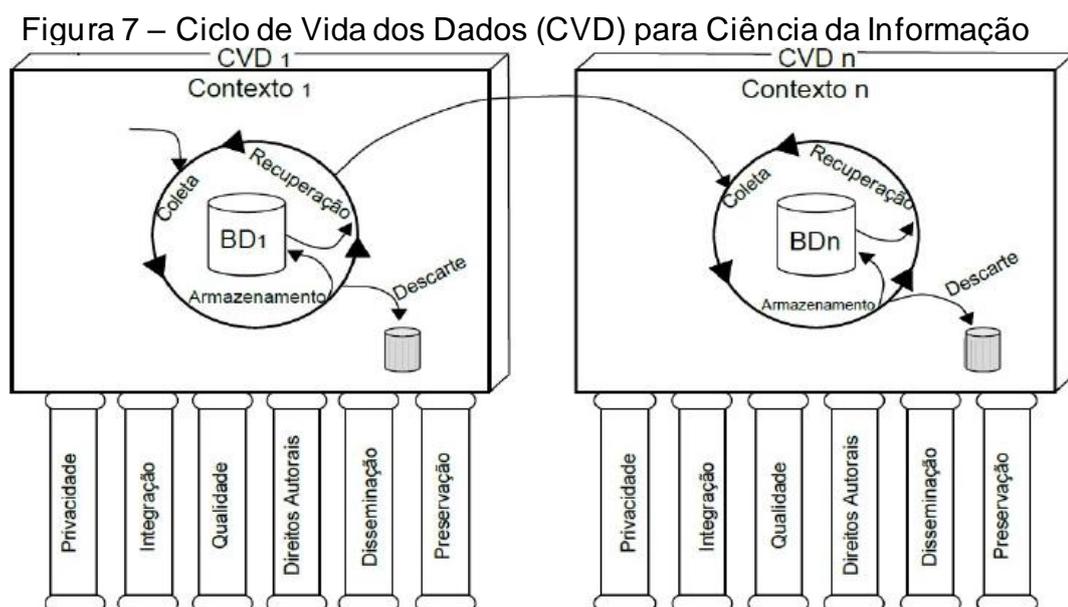
A fase seguinte de armazenamento trata dos procedimentos para preservar os dados obtidos na fase de coleta para acesso futuro, de modo a manter seus conteúdos em um determinado suporte.

O objetivo da fase de recuperação é a viabilização de alternativas para a encontrabilidade, acesso e interpretação dos dados

Em alguns casos, após a coleta, os dados obtidos podem ser considerados descartáveis (não há necessidade futura de acesso), ou mesmo após seu armazenamento pode-se perceber que não há mais necessidade de manter determinados dados. Essas situações justificam a fase de descarte, na qual há a eliminação de parte dos dados, podendo ocorrer em bloco (exclusão de subconjuntos inteiros), horizontalmente (exclusão de elementos da estrutura de dados por meio de critérios específicos) ou verticalmente (eliminação de elementos estruturais de dados).

Segundo Sant'ana (2016), nas fases de armazenamento, recuperação e descarte são necessários conhecimentos mais profundos da Ciência da Computação a fim de planejar e executar ações para cada fase, envolvendo ainda forte potencial de participação para a CI e pouca participação dos usuários.

Na Figura 7 é esquematizado o CVD para a Ciência da Informação.



Fonte: Sant'ana (2016).

No contexto do Governo Aberto, o CVD pode identificar características e estratégias adotadas por governos que restringem ou expandem as possibilidades do uso de dados governamentais pela sociedade, de modo a avaliar o grau de maturidade da transparência de dados (MATHEUS; RIBEIRO; VAZ, 2018).

Similar ao ciclo proposto por Berman *et al.* (2018) e por Sant'ana (2016), Matheus, Ribeiro e Vaz (2018) propõe um ciclo de vida para dados governamentais, que compreende:

- criação: incluindo os processos de coleta, armazenamento e seleção de dados a serem compartilhados;
- publicação: tratamento e limpeza de dados, criação de metadados e publicação em portal;
- uso: inclui conexão, descoberta, acesso, exploração e uso dos dados;
- curadoria: ações relacionadas à manutenção e governança de dados.

Entende-se, dessa forma, que a padronização de ações voltadas à utilização e disponibilização de dados é essencial para sua utilização para produção de conhecimento, havendo possibilidades diversas de estudo e aplicações em cada fase do ciclo.

Na próxima seção aborda-se a metodologia utilizada para desenvolvimento desta pesquisa, a fim de atender aos objetivos estabelecidos e responder à questão de pesquisa elaborada.

4 METODOLOGIA

A pesquisa baseou-se em uma estratégia qualitativa, com caráter exploratório. Quanto aos procedimentos, utilizar-se-á a observação e a pesquisa bibliográfica.

4.1 OBJETO DE PESQUISA

Trata da análise das interrelações entre Ciência da Informação e Ciência de Dados buscando aplicações em Sistemas de Informação Governamentais de modo a prover benefícios à geração de informação e conhecimento.

Tal objeto justifica a pesquisa qualitativa, pois como explicado por Creswel (2007, p. 186 apud AUGUSTO *et al.*, 2013, p. 748), na pesquisa qualitativa “o ambiente natural é a fonte direta de dados e o pesquisador, o principal instrumento, sendo que os dados coletados são predominantemente descritivos”. A estratégia permitirá explorar aspectos da realidade que não são sujeitos à quantificação, visando à produção de informações aprofundadas e à compreensão das relações entre conceitos de SI e Ciência de Dados.

A investigação exploratória foi considerada a mais adequada para a pesquisa. Entende-se que esse tipo de investigação é justificado considerando que a disciplina da Ciência de Dados é uma área ainda em desenvolvimento, de modo que seu estudo a partir da ótica de um campo científico já sedimentado que é a Ciência da Informação, necessita de maior familiaridade com suas possíveis relações interdisciplinares, de modo que essa metodologia permitirá, ao final do trabalho, obter uma nova percepção acerca dos temas estudados.

Na pesquisa exploratória, busca-se conceber a teoria, prática e instrumentos relacionados ao contexto da pesquisa, utilizando como instrumento a pesquisa bibliográfica.

4.2 PROCEDIMENTOS DE COLETA DE DADOS

Inicialmente, realizou-se revisão bibliográfica voltada à busca de publicações científicas anteriores que abordam temáticas relacionadas a interações entre Ciência da Informação e Ciência de Dados com foco em Sistemas de

Informação ou e-Gov.

Foram pesquisados trabalhos nas bases de dados da Ciência da Informação: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI) e *Library and Information Science Abstracts* (LISA), bem como nas plataformas referenciais de citações científicas *Web of Science* e *Scopus*, tendo-se dado preferência a publicações oriundas da Ciência da Informação. A busca nessas bases utilizou os seguintes descritores: “information system” ou “e-government” ou “government”; “data science” ou “big data” ou “analytics” ou “e-science” ou “data mining” ou “machine learning”.

Na pesquisa inicial, foram utilizados os descritores em conjunto, em seguida foram utilizadas combinações dos descritores, a fim de procurar temas relacionados a Sistemas de Informação Governamentais e Ciência de Dados, de modo a analisar as práticas utilizadas e possíveis relações entre os temas. Limitou-se a pesquisa de publicações a artigos científicos cuja área de publicação fosse a Ciência da Informação, não tendo sido adotados parâmetros temporais para a pesquisa.

A pesquisa inicial encontrou resultados em todas as bases pesquisadas, sendo em seguida feita análise e interpretação dos dados, para verificação da adequação dos artigos encontrados ao tema de interesse.

Após seleção de artigos, a base da *Web of Science* apresentou os resultados mais relevantes à pesquisa, com artigos abordando temas como treinamento digital, análise de dados, dados abertos, mineração de dados, *big data*, *linked data* e interoperabilidade de dados.

A base científica *Scopus* apresentou artigos com foco em abordagens de cidades inteligentes, mineração de dados, dados abertos, Gestão do Conhecimento, *internet* das coisas e aprendizado de máquina – com aplicações voltadas a questões como planejamento urbano e tratamento de solicitações de cidadãos.

Na plataforma LISA foram encontrados artigos científicos com foco em mineração de dados, dados abertos, *big data*, análise de dados e *business intelligence*.

Por fim, o portal Brapci – no qual foram utilizadas as combinações de descritores anteriormente apresentados, mas traduzidos à língua portuguesa – apresentou menos resultados relacionados à temática da pesquisa, no entanto apresentando artigos relevantes para o tema de interoperabilidade e dados abertos.

A pesquisa bibliográfica permitiu a identificação de temas nos quais a Ciência de Dados pode ser utilizada para beneficiar a operação de Sistemas de Informação Governamentais, servindo como base para o aprofundamento nesses temas e posterior relação a teorias, métodos e práticas da CI.

Também permitiu a adoção de autores como Lin Wang (2018), Ferneda *et al.* (2016), Berman *et al.* (2018) e Wagner (2013) como referência para a pesquisa, uma vez que esses autores abordam aplicações da Ciência de Dados em ambientes organizacionais ou governamentais, fazendo em alguns casos relações com a CI.

A análise de dados, detalhada a seguir, permitiu criar inferências decorrentes dos dados de pesquisa e sua classificação a partir de eixos temáticos. Realizou-se pesquisa bibliográfica complementar para o aprofundamento nesses temas para construção da base teórica da pesquisa.

4.3 FORMAS DE ANÁLISE DE DADOS

A análise de dados, como explica Teixeira (2003), é a formação de sentido e significado a partir dos dados, constituindo a constatação de um estudo.

Essa etapa da pesquisa foi feita agrupando os resultados da pesquisa em grupos de significado relacionados à temática abordada, organizando e interpretando os dados. A técnica de análise utilizada foi a análise de conteúdo, que para Teixeira (2003) representa um conjunto de técnicas para interpretação de dados a partir de um modelo conceitual, contando a frequência de um fenômeno e identificando relações entre fenômenos.

Bardin (1977, p. 38) descreve a análise de conteúdo como “a inferência de conhecimentos relativos às condições de produção (ou, eventualmente de recepção), inferência esta que recorre a indicadores (quantitativos ou não)”.

A técnica foi aplicada aos dados de pesquisa encontrados a fim de identificar correlações possíveis entre a CI e a Ciência de Dados e como essas correlações podem ser utilizadas para melhoria de Sistemas de Informação Governamentais permitindo aprimorar a disponibilidade de dados governamentais.

Para tal, seguiram-se as etapas descritas por Bardin (1997):

- pré-análise do material: realização de leitura flutuante no material encontrado, para contextualização e criação de indícios iniciais para aprofundamento nas fases seguintes;

- exploração do material: seleção e criação de unidades temáticas para posterior classificação;
- tratamento e interpretação dos resultados: classificação dos dados em categorias integrando temas próximos.

Os resultados foram analisados a partir de seus títulos, resumos e palavras-chaves dos documentos, de modo a selecionar aqueles que poderiam contribuir para o desenvolvimento do estudo, em seguida sendo realizada a leitura do texto completo.

Assim, na fase de tratamento e interpretação dos resultados, foi realizada classificação dos dados de publicações encontradas, de modo a correlacionar as publicações a partir de temas comuns.

A criação de categorias originárias dos dados de pesquisa baseou-se, como mencionado, em temas em comum entre as publicações levantadas na pesquisa bibliográfica, resultando na criação dos grupos apresentados, que representam áreas de estudo correlacionando as disciplinas de Ciência da Informação e Ciência de Dados em aplicações em Sistemas de Informação Governamentais, baseando-se, assim, não em áreas de uma ou outra disciplina, mas possíveis áreas de estudo que relacionam uma ou mais subáreas de interesse de cada disciplina.

Considerando os temas encontrados com maior relevância e as correlações entre eles, decidiu-se por explorar os resultados da pesquisa a partir dos seguintes eixos principais decorrentes da análise de conteúdo: Dados Abertos, interoperabilidade, Gestão do Conhecimento, *analytics* e Inteligência Artificial, como visto a seguir.

5 INTERSECÇÕES ENTRE CIÊNCIA DA INFORMAÇÃO E CIÊNCIA DE DADOS

Objetiva-se neste capítulo abordar relações entre a Ciência da Informação e a Ciência de Dados com os Sistemas de Informação Governamentais, explorando os conceitos apresentados nos capítulos anteriores.

Foram encontradas, a partir de levantamento bibliográfico em bases de dados científicas, um número significativo de interações entre a CI e Ciência de Dados em artigos científicos relacionados à CI, com possíveis aplicações em Sistemas de Informação Governamentais.

Apresenta-se a seguir os resultados de intersecções entre as duas áreas obtidos pela pesquisa, dividindo-os em grupos categóricos: dados abertos, interoperabilidade, Gestão do Conhecimento, *analytics* e Inteligência Artificial.

5.1 DADOS ABERTOS

À medida que cresce o volume de informações digitais produzidas, aumenta a possibilidade de reuso dessas informações, argumentam Borglund e Engvall (2014). Para os autores, essa ocorrência é geralmente manifestada sob o termo “Dados Abertos” (ou *Open Data*), que seria assim um conceito representando a ideia de que a informação pública deve estar disponível para o uso de todos.

São características dos dados abertos: a disponibilidade e o reuso irrestrito (JANSEN; CHARALABIDIS; ZUIDERWIJK, 2012), bem como a interoperabilidade pela adoção de padrões (LNENICKA; KOMARKOVA, 2019), visando a participação universal pela abertura de dados sem discriminação e incentivando o engajamento (Open Government Partnership, 2011) (Figura 8).

Figura 8 – Características Principais dos Dados Abertos



Fonte: Página ... (2021, não paginado).

Kassen (2020, p. 766) afirma que o conceito de Dados Abertos governamentais, que o autor relaciona ao fenômeno dos Dados Abertos, é amplamente aceito nas comunidades acadêmicas e profissionais como significando “um maciço estoque digital de informação pública de várias formas e tipos de informação, documentos governamentais e conjuntos de dados”. Nesse sentido, é possível relacionar os Dados Abertos governamentais ao *big data*, como fazem BERTOT *et al.* (2014) e Lnenicka e Komarkova (2019) ao descrever o *big open data*, que seria a disponibilização de grandes conjuntos de dados governamentais.

O acesso aberto a dados governamentais possui grande potencial de colaboração entre entes governamentais e a sociedade, dependendo da criação de um ambiente participativo de aprendizado e cooperação (KASSEN, 2020).

Apesar das vantagens apresentadas para o uso de Dados Abertos no contexto governamental, Correa, Melo Jr. e Silva (2020) alertam que nem todos os portais de governo publicam Dados Abertos, apontando estudos que indicam que grande parte dos portais no Brasil em infraestruturas que não têm as funcionalidades essenciais para o provimento de Dados Abertos. Autores como Layne e Lee (2001) e Thomas, Cipolla, Lambert e Carter (2019) chamam atenção para a necessidade de se avaliar os níveis de maturidade na utilização de Dados Abertos em governos, a fim de mitigar problemas, reduzir custos, gerenciar questões autorais e de privacidade dos dados e obter melhores serviços de informação.

Como explicam Matheus, Janssen e Maheshwari (2020), a Ciência de Dados no âmbito governamental relaciona-se com a extração, interpretação e

apresentação de observações (ou *insights*, na expressão anglicista) de dados estruturados ou não estruturados e que podem tanto ser abertos quanto proprietários, de modo que a partilha, uso e interpretação de tais dados atuam para aproximar governo e público e potencialmente influenciar no desenvolvimento de políticas públicas. As ferramentas da Ciência de Dados também seriam adequadas aos processos de disponibilização de Dados Abertos relacionados ao *big open data* (ELEZAJ; TOLE; BACI, 2018).

Hongyan e Mingdi (2012) ponderam que um dos objetivos mais importantes dos Dados Abertos é a satisfação das necessidades informacionais do usuário ao invés das necessidades do criador da informação.

Em relação aos Sistemas de Informação Governamentais, a intersecção entre CI e Ciência de Dados pode se dar a partir da concepção do regime de informação, que Gómez (2012, p. 43) define como:

[...] o modo informacional dominante em uma formação social, o qual define quem são os sujeitos, as organizações, as regras e as autoridades informacionais e quais os meios e os recursos preferenciais de informação, os padrões de excelência e os modelos de sua organização, interação e distribuição, enquanto vigentes em certo tempo, lugar e circunstância.

Conforme elabora Gómez (2012), o regime de informação abrange a gestão da informação e as políticas de informação – estas visando a satisfação das necessidades dos usuários. Assim, o regime de informação deve influenciar nas metodologias e tecnologias utilizadas pela Ciência de Dados.

Além do regime de informação, a intersecção entre as duas disciplinas em relação a dados abertos pode ocorrer por meio do Ciclo de Vida de Dados, que determina os processos a serem executados para coleta, armazenamento, recuperação e descarte dos dados de Sistemas de Informação Governamentais.

A organização, gerenciamento e a padronização de ações para transformação de dados em conhecimento são vantagens associadas ao CVD, lembrando Berman *et al.* (2018), de forma que a padronização provida por esse ciclo pode complementar o regime de informação associado a dados abertos, nos quais várias etapas podem ser associadas a práticas da CI (SANT'ANA, 2016).

Dados governamentais abertos normalmente são heterogêneos (possuem formatos distintos), passíveis de uso, reuso e modificações e normalmente

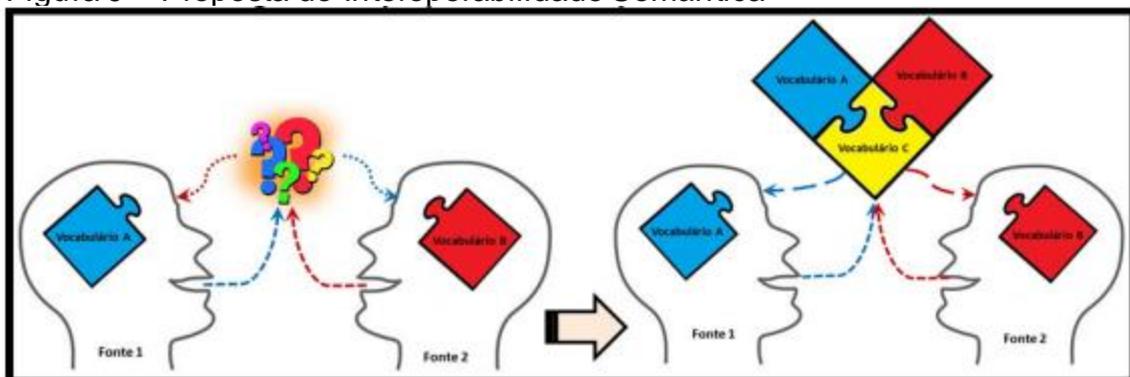
distribuídos em várias fontes (KASSEN, 2020; LNENICKA; KOMARKOV, 2019), de modo que a interoperabilidade entre dados é uma questão relevante em Sistemas de Informação Governamentais, tópico discutido a seguir.

5.2 INTEROPERABILIDADE

A interoperabilidade entre sistemas consiste na sua capacidade de interagir com outros sistemas sem esforço especial por parte de seu usuário, só sendo possível alcançá-la a partir da utilização de padrões e normas (IEEE apud FARINELLI; ALMEIDA, 2014).

Como explicam Farinelli e Almeida (2014), na comunicação entre duas fontes computacionais com vocabulários (padrões) distintos, existe um problema de identificação dos termos entre as fontes que é potencializada pela ausência de capacidade interpretativa das máquinas, problema este que pode ser resolvido a partir da utilização de interoperabilidade semântica com a criação de um padrão para tradução dos dois vocabulários distintos permitindo a compreensão (Figura 9).

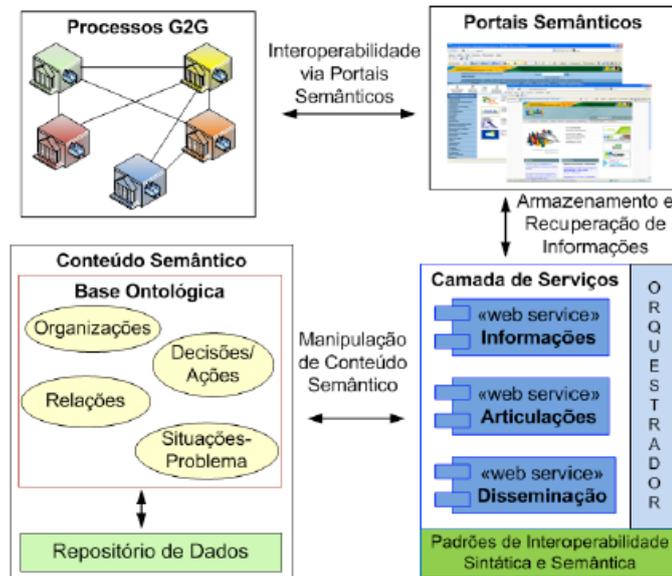
Figura 9 – Proposta de Interoperabilidade Semântica



Fonte: Farinelli e Almeida (2014, p. 10).

Para Fereda *et al.* (2016), a interoperabilidade é um dos requisitos essenciais na viabilização do Governo Eletrônico. Silva, Cunha e Oliveira (2011) destacam o crescimento de iniciativas de e-Gov na divulgação de informações e interações relacionadas a governos, destacando o uso de TIC para o provimento de interoperabilidade entre serviços através de padrões semânticos (Figura 10).

Figura 10 – Modelo de Interoperabilidade Semântica aplicada a Governo Eletrônico



Fonte: Silva, Cunha e Oliveira (2011, p. 5).

Considerando que a Ciência de Dados busca lidar com fontes de dados provenientes de origens diversas, suas técnicas e ferramentas tornam-se uma possível solução tecnológica para a interoperabilidade. Mahmoud, Omar e Ouksel (2019) destacam a possibilidade de integração e interoperabilidade de várias bases de dados públicas a partir de soluções da Ciência de Dados com o intuito de oferecer melhores serviços a cidadãos e empresas, para isso devendo coletar dados de fontes heterogêneas. Silva, Cunha e Oliveira (2011, p. 2) definem que a colaboração entre entes governamentais para o provimento de melhores serviços de e-Gov consiste na “[...] capacidade de um sistema, informatizado ou não, de se comunicar de forma transparente com outro sistema de mesma natureza ou natureza distinta”.

Para Santos (2008), o cenário ideal para o Governo Eletrônico seria a existência de um único ponto de acesso a serviços e informações, integrando dados, processos e transações, similar ao modelo da Figura 11.

Figura 11 – Modelo de Governo Eletrônico Centralizado



Fonte: E-Ping (2006 apud SANTOS, 2008, p. 108).

Como afirmam Ferneda *et al.* (2016), a complexidade crescente dos Sistemas de Informação construídos para lidar com dados e informações heterogêneas dificulta o provimento de serviços governamentais eficientes, principalmente considerando que esses sistemas passam a ter a preocupação de compartilhar não apenas dados, mas também informação, havendo, assim, a necessidade de interoperabilidade semântica em sistemas e-Gov a partir do uso de padrões. Nesse sentido, a CI pode integrar-se às soluções tecnológicas provendo abordagens, modelos e padrões referentes aos processos informacionais para atendimento da interoperabilidade semântica.

Silva, Cunha e Oliveira (2011) destacam o potencial para uso da Web Semântica em soluções de e-Gov, devido a sua capacidade de prover o compartilhamento e reuso de dados além dos limites das aplicações e organizações (W3C, 2021). A Web Semântica utiliza formatos universais, baseados no padrão *Resource Description Framework* (RDF), que permitem a integração e combinação de dados de fontes diversas, além de utilizar linguagem para registrar como os dados se relacionam a objetos do mundo real de modo a possibilitar a navegação por bases de dados – tanto pessoas quanto máquinas – a partir de assuntos (W3C, 2021).

Autores como Bernstein, Hendler e Noy (2016) e Jainl e Kumar (2018) também destacam as possibilidades da Web Semântica para interoperabilidade, integração de dados e recuperação de informação em ambientes *web*. No entanto, Jainl e Kumar (2018) chamam atenção para o fato de que a tecnologia *web* atual não faz uso apropriado dos conceitos de Web Semântica, ainda que os autores

acreditem que ela possa ser usada para extração e seleção de dados heterogêneos, distribuídos e em grande volume, em conjunto com ontologias.

No Brasil existe a iniciativa Padrões de Interoperabilidade de Governo Eletrônico (e-Ping), que define um conjunto premissas, políticas e especificações técnicas regulamentando o uso de TIC na interoperabilidade de serviços de Governo Eletrônico, estabelecendo as condições de interação com os demais Poderes e esferas de governo e com a sociedade em geral (BRASIL, 2018). O documento de referência do e-Ping estabelece diretrizes para interoperabilidade de soluções tecnológicas do Governo Brasileiro, baseando-se no uso de padrões como o XML – padrão da W3C (*World Wide Web Consortium*, a principal organização de padronização para *internet*) geração de linguagens de marcação – e *web services* (SANTOS, 2018).

Ferneda *et al.* (2016) abordam a possibilidade do uso de ontologias – instrumentos de representação do conhecimento que permitem o entendimento da linguagem natural, integração de fontes de informação e interoperabilidade semântica a partir da especificação formal de domínios (LIMA; MACULAN, 2017) – para provimento de interoperabilidade semântica em Sistemas de Informação Governamentais, citando iniciativas dos governos do Brasil, Palestina, União Europeia, Estônia e Países Baixos nessa direção. Os autores reconhecem que o uso de ontologias para interoperabilidade em governos eletrônicos ainda não é uma alternativa consolidada, no entanto, trata-se de ferramenta de grande potencial para esse objetivo. Em linha similar, Silva, Cunha e Oliveira (2011) propõem um modelo de interoperabilidade semântica com a utilização de ontologias baseado na Web Semântica.

Já Maali, Cyganiak e Peristeras (2010) abordam a possibilidade do uso do *Linked Data* para a interoperabilidade semântica. O *Linked Data* proporciona a criação de ligações entre dados da *web* pela criação de documentos compartilhados na rede, como explica Berners-Lee (2006), diferenciando-se de *links* comuns pelo uso de descrições através do padrão RDF, fornecendo, assim, dados estruturados e semântica formal relacionados a um domínio específico (CONEGLIAN; DIEGER; SANTARÉM SEGUNDO; CAPTREZ, 2017).

No estudo de Maali, Cyganiak e Peristeras (2010), é proposto o uso desse formato para padronização e representação de catálogos de dados governamentais – que consistem basicamente em listas dos conjuntos de dados

governamentais em conjunto com metadados –, de modo a facilitar a busca de informação, especialmente a automatizada, a preservação digital dos dados, apresentar uma descrição formal do catálogo de dados e garantir a interoperabilidade a partir da padronização. Os autores identificam ainda um crescimento no uso de tecnologias de *linked data* na publicação de dados governamentais.

Verifica-se, assim, que a interoperabilidade entre dados e sistemas governamentais apresenta diversas possibilidades de interação entre a CI e a Ciência de Dados, buscando sobretudo a transformação de simples dados em informação, por meio da criação de padrões informacionais a serem consumidos pelas soluções tecnológicas.

A disponibilização e interoperabilidades de dados governamentais envolve ainda questões relacionadas ao compartilhamento de informações e conhecimento derivados desses dados, conforme abordado a seguir.

5.3 GESTÃO DO CONHECIMENTO

O conhecimento, como lembram Cardoso e Machado (2008), é um dos recursos mais importantes de uma organização, propiciando ações e incentivando a inovação, justificando a necessidade de gerir a maneira como o conhecimento é gerado, armazenado, distribuído e utilizado, sendo essa a função da Gestão do Conhecimento (GC).

A GC “é uma disciplina relativamente nova, que vem sendo objeto de estudo em outras áreas além da Ciência da Informação” (BETTENCOURT; CIANCONI, 2012, p. 2), “apresentando abordagens que se destacam na literatura da Ciência da Informação que estão, em grande medida, voltadas para a gestão do conhecimento explícito” (CÂMARA; FERREIRA, 2017), conhecimento esse que é facilmente transmitido entre indivíduos e grupos, pois é representado de maneira formal (CHOO, 1996).

Wagner (2003) lista aplicações da GC em Sistemas de Informação Governamentais de Estados Unidos, Canadá e Hong Kong, destacando que sua utilização busca o compartilhamento interno (entre setores governamentais) e/ou externo (visando ao público) de conhecimento a partir de sua formalização, utilizando princípios da categorização e indexação, em alguns dos exemplos

abordados pelo autor, para realizar a formalização de conhecimentos em dados governamentais. O foco das aplicações de GC em e-Gov a partir dos exemplos estudados por Wagner (2003) é a padronização de infraestruturas, treinamento de usuários e criação de um ambiente colaborativo.

A padronização de infraestruturas é percebida em SI dos governos dos Estados Unidos e Canadá, estudados por Wagner (2003), na qual essa padronização é utilizada para promover a troca de informação e conhecimento por meio de métodos formalizados e para a disponibilização de dados de múltiplas fontes acessíveis através de portal eletrônico. O autor destaca ainda que os portais canadense e de Hong Kong disponibilizam serviços de treinamento na utilização do portal e de fontes de dados, assim como disponibilizam “redes virtuais” (no caso canadense) e serviços de consulta e de recomendações (no caso de Hong Kong) promovendo a interação entre usuários e governo com foco na disponibilização e compartilhamento de conhecimento, assim como na colaboração.

Outra tendência que pode ser percebida é o uso, cada vez mais habitual, de TIC na Gestão do Conhecimento devido ao grande volume de dados que passaram a ser comuns nas atividades de organizações, percebendo-se que a “velocidade de coleta de informações era muito maior do que a velocidade de processamento ou análise das mesmas” (CARDOSO; MACHADO, 2008, p. 496).

Misra, Hariharan e Khaneja (2003) entendem que os SI baseados em computadores, que segundo os autores tornaram-se comuns nas esferas governamentais, são potenciais repositórios de informações relevantes que podem ser disponibilizadas a servidores do governo para promover o desenvolvimento do conhecimento. As organizações atuais, inclusive as governamentais, atuam em um ambiente de mudanças constantes no qual “torna-se necessária a aplicação de técnicas e ferramentas automáticas que agilizem o processo de extração de informações relevantes de grandes volumes de dados”, de modo que ferramentas tecnológicas como as de mineração de dados podem ser usadas com esse fim apoiando a Gestão do Conhecimento (CARDOSO; MACHADO, 2008, p. 497).

Misra, Hariharan e Khaneja (2003) destacam que a GC é importante em ambientes governamentais devido à sua atuação efetiva depender da disseminação de conhecimento, além de se tratar de uma estrutura distribuída em vários órgãos e departamentos. Os autores alertam que a Gestão do Conhecimento em entes governamentais ainda é, em grande parte, feita de maneira desorganizada, a

despeito do uso de TIC e de Sistemas de Informação computacionais. No entanto, os autores veem grande potencial na adoção de princípios de GC em conjunto com tecnologias de *business intelligence* (BI) e mineração de dados, permitindo a geração de conhecimento a partir da qualificação da informação.

Ferramentas de BI e mineração de dados são comumente associadas à Ciência de Dados (LIN WANG, 2018; BERMAN *et al.*, 2018), contribuindo para a apresentação de informações e a descoberta de padrões em dados em processos dessa disciplina. As ferramentas de BI são usadas na geração, no tratamento e na comunicação da informação (REGINATO; NASCIMENTO, 2007) e permitem o cruzamento de dados, visualização de informação em várias dimensões e análise de indicadores (BATISTA, 2004 apud REGINATO; NASCIMENTO, 2007). Já a mineração de dados consiste em um dos passos do processo *Knowledge Discovery in Databases* (KDD), processo esse que utiliza algoritmos e análise de dados para descoberta de padrões em bases de dados, havendo grande número de ferramentas que podem ser usadas para esse propósito, sendo o termo mineração de dados muitas vezes usado como sinônimo do processo KDD (BARTSCHAT; REISCHL; MIKUT, 2019).

Percebe-se, assim, que abordagens da CI relacionada à Gestão do Conhecimento podem ser relacionadas à Ciência de Dados no governo eletrônico para geração de valor informacional à organização e aos usuários.

O valor informacional também pode relacionar-se à análise de dados governamentais, tópico discutido a seguir.

5.4 ANALYTICS

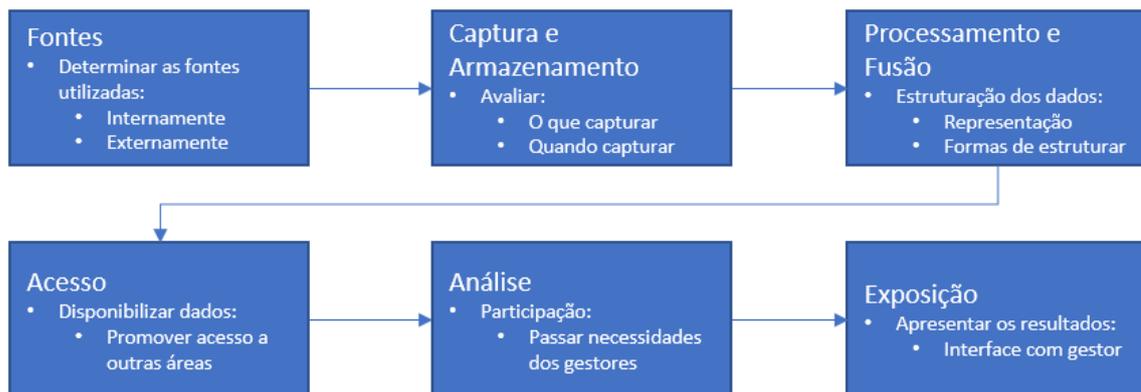
Conforme já mencionado, o conceito de *analytics* (ou *data analytics*) refere-se à aplicação de teorias e ferramentas para análise de dados complexos e obtenção de conclusões. Como Gandomi e Haider (2015) explicam, seu valor depende da capacidade de processar e converter dados em informações capazes de nortear a tomada de decisões.

No setor público, como citam Lnenicka e Komarkova (2019), o uso de *analytics* vem se tornando cada vez mais comum a partir da evolução das tecnologias e técnicas de análise de dados e da maior disponibilidade de dados. Para Chen, Mao e Liu (2014), isso pode contribuir na exploração científica de dados

governamentais, na conformidade regulatória e no uso de dados como base de serviços governamentais de informação.

Nesse cenário, a análise de dados deve obedecer a etapas padronizadas para geração de valor, preocupando-se com aspectos como confiabilidade de fontes, coleta, formatação, integração e análise de dados (CONEGLIAN; GONÇALVEZ; SANTARÉM SEGUNDO, 2017; TAURION, 2013). Bugembe (2016 apud CONEGLIAN; GONÇALVEZ; SANTARÉM SEGUNDO, 2017) sugere, nesse sentido, as seguintes fases para um processo de análise de dados: fonte; captura e armazenamento; processamento e fusão; acesso; análise; e exposição. Na Figura 12 são apresentadas essas fases, relacionando-as a papéis que podem ser assumidos por profissionais da informação em cada uma delas, conforme será discutido adiante.

Figura 12 – Fases de Análise de Dados e o Papel do Profissional da Informação



Fonte: Coneglian, Gonçalves e Santarém Segundo (2017, p. 137).

A CI pode ser inserida nesse processo, com diferentes graus de protagonismo, em todas suas fases, como citado por Coneglian, Gonçalves e Santarém Segundo (2017), devido à necessidade de compreensão de necessidades informacionais internas e externas à organização, atuando na seleção de fontes, na avaliação de dados a serem capturados e armazenados e quando, na aplicação de conceitos de representações de dados, na definição de dados para tratamento, no direcionamento da análise e na apresentação de resultados, atuando em consonância com as tecnologias e técnicas computacionais utilizadas.

Discute-se a seguir a possibilidade da melhoria de serviços

governamentais a partir de sua automação e da formalização de domínios de conhecimento.

5.5 INTELIGÊNCIA ARTIFICIAL

A inteligência artificial (IA) é um campo multidisciplinar cujo objetivo é automatizar atividades replicando aspectos da inteligência humana, atividades essas que incluem reconhecimento de fala, aprendizado, resolução de problemas, dentre outras (HABEED, 2017) (Figura 13).

Figura 13 – Áreas da Inteligência Artificial



Fonte: Blog (2021, não paginado).

Tecuci (2011) lembra que o campo interdisciplinar da IA é tão antigo quanto a Ciência da Computação, obtendo sucesso inicial na resolução de problemas de domínios simplificados, mas não conseguindo o mesmo sucesso na resolução de problemas mais complexos relacionados ao mundo real devido à dificuldade de compreensão de processos cognitivos humanos e a consequente impossibilidade em automatizá-los. No entanto, o autor destaca um crescimento de pesquisas relacionados a IA em anos recentes graças à disponibilidade de grandes conjuntos de dados na internet e à crescente capacidade de processamento de

computadores modernos, levando à criação de oportunidades de aplicação de métodos de IA dependentes mais de dados do que de algoritmos.

Autores como Ayachi *et al.* (2016), Al-Mushayt (2019) e Kaya (2019) preveem a possibilidade da utilização de algoritmos de IA para melhoria dos serviços oferecidos por Sistemas de Informação Governamentais. Possíveis aplicações envolvem o uso de técnicas inteligentes para prover serviços personalizados e melhorar a interação dos cidadãos com os sistemas, automação de serviços e criação de um ambiente digital com maior participação e colaboração cidadã.

Como contribuição para soluções de IA mais eficiente, pode-se considerar o uso de ontologias. Lima e Maculan (2017) afirmam que ontologias, na CI e na ciência da computação, são estudadas como instrumento de representação do conhecimento, em geral, como um artefato computacional, tendo ganhado mais destaque em função do desenvolvimento da *web* semântica, devido a possibilitar o “entendimento da linguagem natural, integração de fontes de informação, interoperabilidade semântica e por servir como base de conhecimento em várias aplicações” (LIMA; MACULAN, 2017, p. 65). Smith e Welty (2001) afirmam que o uso de ontologias se entrelaçou ao campo da Inteligência Artificial e dos Sistemas de Informação, a partir de seu uso para criação de bases de conhecimento e modelagem de domínios, permitindo correlação de dados e provendo base semântica para a resolução de problemas. Sánchez-Cuadrado *et al.* (2007) afirmam que ontologias são certamente um recurso útil para sistemas que contém relações entre seus elementos e dos quais se espera certo grau de inferência ou raciocínio.

Outros Sistemas de Organização do Conhecimento (SOC) além das ontologias, como taxonomias – o agrupamento ordenado de itens baseado em características (MAZZOCHI, 2020) – e tesouros – vocabulários controlados e estruturados que contém relações hierárquicas, associativas e equivalentes entre termos e conceitos (MAZZOCHI, 2020) –, contam com histórico de uso na área de IA. Gilchrist (2003) destaca a criação de “redes semânticas” na construção de sistemas especialistas, que consistem em estruturas conceitualmente similares a tesouros para relação de conceitos, assim como o uso de taxonomias na classificação, indexação e recuperação de informação em sistemas de IA.

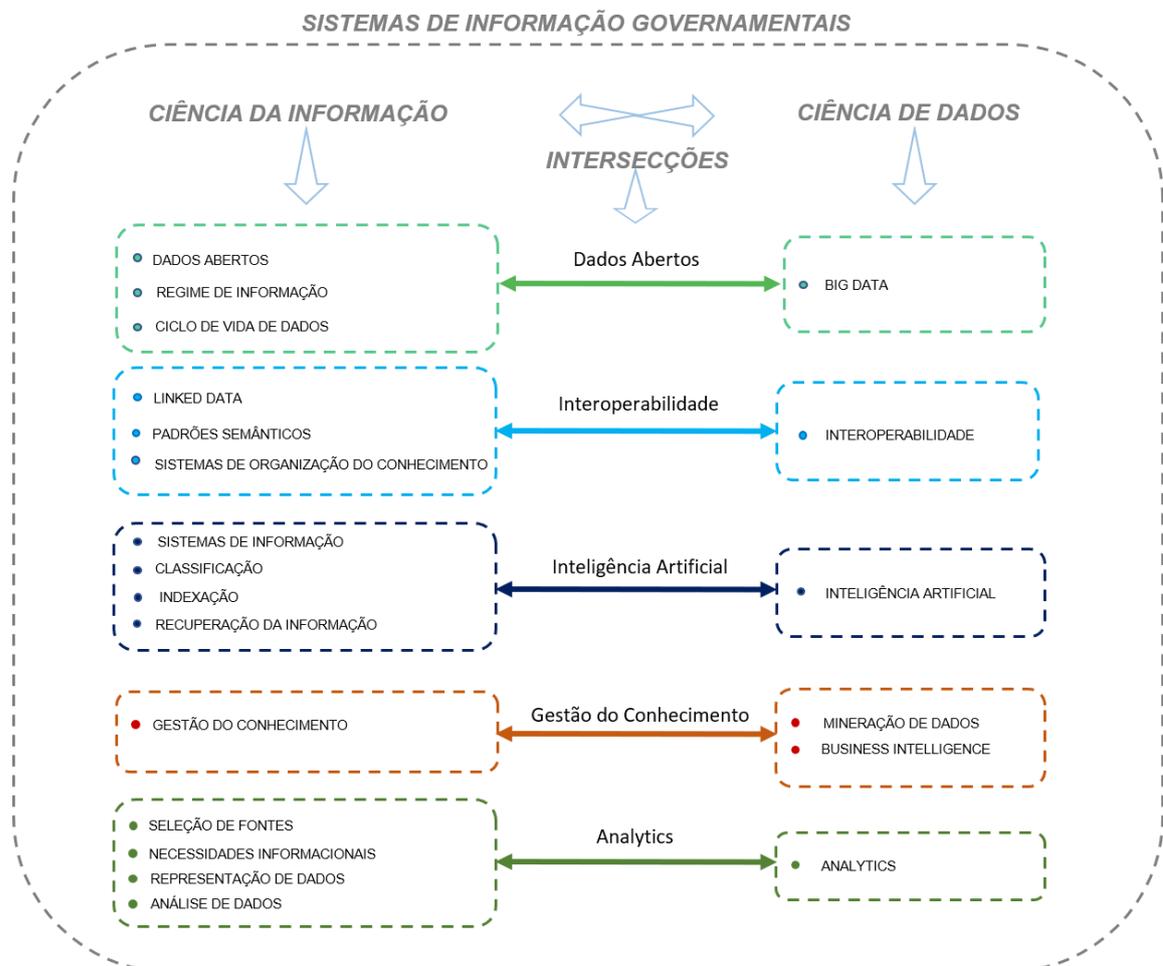
Percebe-se, assim, que a CI pode assumir papel ativo no projeto de ferramentas auxiliares para IA como os SOC, especialmente ontologias, favorecendo a construção de sistemas e-Gov mais eficientes, sobretudo considerando o pouco

uso da IA no âmbito governamental, como aponta Al-Mushayt (2019).

5.6 SOBRE AS INTERSECÇÕES ENCONTRADAS

Os resultados da pesquisa permitiram a análise de possíveis aplicações da Ciência de Dados em Sistemas de Informação Governamentais e das intersecções entre a Ciência da Informação e a primeira neste contexto. Partindo das categorias apresentadas, na Figura 14 apresenta-se modelo esquemático das relações encontradas.

Figura 14 – Intersecções entre Ciência de Dados e Ciência da Informação em aplicações de Sistemas de Informação Governamentais



Fonte: Elaboração própria, 2021.

Na Figura 14 é apresentada, à esquerda as áreas de estudo relacionadas à CI, à direita as áreas de estudo relacionadas à Ciência de Dados e entre as duas

temáticas as setas as interligando representam os grupos resultantes da análise de dados da pesquisa, representando categorias de intersecção entre a CI e Ciência de Dados. Cabe destacar que dessa intersecção tem-se as aplicações em Sistemas de Informação Governamentais.

As categorias (ou grupos) de intersecção apresentados não representam áreas de estudo específicas da Ciência da Informação ou da Ciência de Dados, mas temas a partir dos quais as intersecções identificadas se relacionam.

Apresentando individualmente os grupos de intersecção, o primeiro grupo é o de dados abertos, que apresenta relações da CI com a Ciência de Dados a partir de aplicações em Sistemas de Informação Governamentais para a garantia de acesso aberto a dados. A seguir, o grupo de interoperabilidade é caracterizado por aplicações para prover interação entre dados e sistemas governamentais para divulgação de informações. O grupo de Gestão do Conhecimento relaciona-se ao compartilhamento de conhecimento dentro do próprio ambiente governamental e também o compartilhamento externo para usuários dos Sistemas de Informação Governamentais, provendo a formalização de conhecimento. O grupo seguinte de *analytics* diz respeito à exploração e uso de dados governamentais. Por fim, o grupo de Inteligência Artificial apresenta intersecções entre as mencionadas disciplinas a partir da automação de serviços governamentais.

Essas aplicações, identificadas a partir da análise de dados dos resultados da pesquisa bibliográfica realizada, apresentam potencial para estudo de áreas tanto da CI quanto da Ciência de Dados, sendo possível a integração entre métodos, processos e teorias das duas áreas de modo a permitir o desenvolvimento de padrões e melhorias para processos de Sistemas de Informação Governamentais.

As intersecções apresentadas no grupo de dados abertos remetem a autores como Kassen (2020), Bertot *et al.* (2014), Lnenicka e Komarkova (2019), Gómez (2012) e Berman *et al.* (2018), que abordam relações entre *big data* e áreas de interesse da CI como dados abertos, sistemas de informação e ciclo de vida de dados.

A interoperabilidade semântica de dados é destacada em pesquisas realizadas por Ferneda *et al.* (2016), Mahmoud, Omar e Ouksel (2019), Silva, Cunha e Oliveira (2011), Lima e Maculan (2017) e Maali, Cyganiak e Peristeras (2010).

A intersecção relacionada à Gestão do Conhecimento é verificada por

Wagner (2003), Cardo e Machado (2008) e Misra, Hariharan e Khaneja (2003).

Lnenicka e Komarkova (2019), Chen, Mao e Liu (2014), Coneglian, Gonçalves e Santarém Segundo (2017) e Taurion (2013) apresentam possibilidades de intersecções relacionadas a *analytics*.

Por fim, a automação de serviços governamentais a partir de Inteligência Artificial envolvendo inter-relações entre CI e Ciência de Dados é abordada por autores como Lima e Maculan (2017), Smith e Welty (2001), Smith e Welty (2001) e Al-Mushayt (2019).

Esquematisando os grupos de intersecção definidos na pesquisa, é apresentado quadro-síntese explorando-os e as relacionando a áreas de estudo da CI e Ciência de Dados, conforme descrito no Quadro 1.

Quadro 1 – Quadro-Síntese dos Grupos de Intersecções entre Ciência da Informação e Ciência de Dados com aplicações em Sistemas de Informação Governamentais

GRUPOS DE INTERSECÇÃO	DESCRIÇÃO	ÁREAS RELACIONADAS DA CIÊNCIA DA INFORMAÇÃO	ÁREAS RELACIONADAS DA CIÊNCIA DE DADOS
Dados Abertos	Relaciona-se a processos para garantia de acesso aberto a dados governamentais	- Dados Abertos - Regime de Informação - Ciclo de Vida de Dados	- <i>Big Data</i>
Interoperabilidade	Relaciona-se à interação de dados e sistemas para divulgação de informações e interações relacionadas a governos	- <i>Linked Data</i> - Padrões Semânticos - Sistemas de Organização do Conhecimento	- Interoperabilidade
Gestão do Conhecimento	Refere-se ao compartilhamento de conhecimento, visando a formalização de conhecimentos em dados governamentais	- Gestão do Conhecimento	- Mineração de Dados - <i>Business Intelligence</i>
Analytics	Relaciona-se à exploração e uso de dados governamentais	- Seleção de Fontes - Necessidades	- <i>Analytics</i>

	a partir de teorias e ferramentas para análise de dados complexos e obtenção de conclusões	Informacionais - Representação de Dados - Análise de Dados	
Inteligência Artificial	Refere-se à automação de serviços de Sistemas de Informação Governamentais a partir do uso de algoritmos de Inteligência Artificial	- Sistemas de Informação - Classificação - Indexação - Recuperação da Informação	- Inteligência Artificial

Fonte: Elaboração própria, 2021.

Entende-se assim que as áreas de *big data*, aprendizado de máquina, mineração de dados, *business intelligence* e *analytics* da Ciência de Dados, assim como tecnologias para interoperabilidade relacionadas à disciplina, representam aplicações possíveis em Sistemas de Informação Governamentais.

Do mesmo modo, essas aplicações da Ciência de Dados possuem relações com áreas, técnicas e conhecimentos da Ciência da Informação, sendo elas: Dados Abertos, regime de informação, Ciclo de Vida dos Dados, *linked data*, padrões semânticos, Sistemas de Organização do Conhecimento, Gestão do Conhecimento, seleção de fontes de dados, necessidade informacionais, representação de dados, análise de dados, sistemas de informação, classificação, indexação e recuperação da informação.

Como abordado nas seções anteriores, as intersecções encontradas apresentam possibilidades de estudo e aplicação capazes de beneficiar serviços e processos relacionados aos Sistemas de Informação Governamentais, contribuindo para a geração de valor informacional, satisfação de usuários, transparência e maior eficiência.

Portanto, foi possível identificar intersecções relevantes para o estudo da Ciência de Dados pela CI, ainda que possibilidades não identificadas devam existir, assim como novos avanços nas disciplinas levarão ao surgimento de novas intersecções.

6 CONSIDERAÇÕES FINAIS

A investigação realizada buscou responder quanto às possíveis intersecções entre a Ciência da Informação e a Ciência de dados tendo em vista aplicações para o desenvolvimento de Sistemas de Informação Governamentais. A pesquisa evidenciou possibilidades de correlações em áreas como Dados Abertos, interoperabilidade, Gestão do Conhecimento, *analytics* e Inteligência Artificial, correlações essas que podem ser estudadas e aplicadas para melhoria de serviços governamentais.

Apreendeu-se ainda que os Sistemas de Informação Governamentais caminham cada vez mais para o uso de tecnologias de modo a atender às demandas sociais por maior transparência do poder público, maior acesso à informação e melhores serviços, associando assim também a conceitos como Governo Aberto, Dados Abertos e Governo Eletrônico.

A transparência das ações públicas é um dos requisitos mais importantes para o alcance dos objetivos de Governo Aberto e Governo Eletrônico, permitindo maior controle social e maior participação cidadã no governo, daí percebe-se a importância do movimento de Dados Abertos aplicado à área pública, cuja eficiência depende da utilização de tecnologias adequadas em conjunto com práticas de gestão de dados e informação.

Para garantia de melhores serviços e satisfação do público no desenvolvimento do Governo Eletrônico, a interoperabilidade entre dados e sistemas governamentais é imprescindível. A utilização de padrões e ferramentas semânticas são facilitadores nesse processo, assim como a tecnologia, entendendo-se que a Ciência de Dados incorpora ferramentas indicadas como suporte a esse objetivo.

Ao considerar os dados governamentais como matéria-prima para serviços em Sistemas de Informação Governamentais, o processo de transformação de dados em informação e conhecimento e o compartilhamento desse conhecimento tanto interna quanto externamente depende de procedimentos definidos e gerenciáveis, que podem ser providos a partir da Gestão do Conhecimento. Também se considera que o uso de ferramentas tecnológicas com objetivos similares, como BI e técnicas de mineração de dados, pode atuar como facilitador à GC.

A disseminação da informação é outra questão importante para os objetivos de Governo Aberto, associando-se não só à transparência como também à

geração de valor informacional. Nesse sentido, o uso de *analytics* como elemento norteador em conjunto a *softwares* de BI para visualização de informação é capaz de aumentar a eficiência de serviços em Sistemas de Informação Governamentais, sendo indicada a utilização de processos definidos e orientados a necessidades informacionais.

A eficiência de serviços governamentais eletrônicos pode também ser relacionada à automação de processos e atividades, objetivo que pode ser relacionado à utilização de IA, especialmente algoritmos de aprendizado de máquina que são baseados em dados históricos. A automação é facilitada pela formalização de domínios de conhecimento, o que pode ser efetivado pelo uso de SOC.

Com a presença crescente das TIC nos serviços governamentais, a Ciência de Dados apresenta potencial para uso em Sistemas de Informação Governamentais para aumento da eficiência de serviços e maior interação entre governo e sociedade, tendo sido identificadas na pesquisa as possíveis áreas de aplicação e como elas podem ser beneficiadas com técnicas, modelos e teorias da Ciência da Informação.

A CI, a partir de conceitos como SOC, regime de informação e *linked data*, da disciplina da Gestão do Conhecimento e da *expertise* em processos informacionais, análise de dados e estudos de usuários, é capaz de exercer papel de importância para a garantia de qualidade de informações e consequentemente de melhores serviços de informação governamentais.

A pesquisa inicialmente apresentou as bases teóricas a partir das quais foi desenvolvida, explorando os conceitos de Sistemas de Informação, Sistemas de Informação Governamental, Governo Aberto e Ciência de Dados, permitindo o aprofundamento do tema de Sistemas de Informação Governamental e a abordagem de aplicações correlacionando a Ciência da Informação e Ciência de Dados.

Os serviços oferecidos pelo governo a partir de SI ainda parecem em um estágio inicial de utilização de tecnologias associadas à Ciência de Dados, bem como desta com a CI, no entanto, o potencial de conceitos como Dados Abertos e interoperabilidade de bases de dados parece ter grande potencial no provimento de transparência e de responsabilização governamental, para isso sendo necessário o planejamento, padronização de dados e criação de modelos, bem como a compreensão das necessidades informacionais envolvidas.

De forma similar, ferramentas como *analytics* e IA, bem como a Gestão do

Conhecimento, podem ser incorporadas a serviços governamentais, gerando valor informacional, aumentando a participação cidadã e contribuindo para a criação de um ambiente digital colaborativo, indo de encontro aos objetivos do movimento OGD.

Entende-se que as possibilidades de intersecções entre a CI e a Ciência de Dados são amplas, devido a existirem correlações nos objetos de estudo das duas disciplinas e da natureza interdisciplinar comum a ambas. A pesquisa buscou explorar as principais a partir de pesquisa exploratória e revisão bibliográfica, no entanto, acredita-se que existem outras possibilidades de colaboração entre as áreas de conhecimento e que novas possibilidades surgirão a partir do desenvolvimento de novas tecnologias e técnicas.

Quanto aos Sistemas de Informação Governamentais, conclui-se que as relações entre sociedade e governo caminham para mudanças que devem ser refletidas nos serviços ofertados por esses Sistemas de Informação, tornando-se cada vez mais necessário a adoção de práticas e tecnologias que caminhem para a facilitação da participação popular e da qualidade informacional.

REFERÊNCIAS

- AGARWAL, R.; DHAR, V. *Editorial – Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research*. **Information Systems Research**, v. 25, n. 3, p. 443-448, 2014.
- AL-MUSHAYT, O. S. *Automating e-government services with artificial intelligence*. **IEEE Access**, v. 7, p. 146821-146829, 2019.
- ALTER, S. *Defining information systems as work systems: implications for the IS field*. **European Journal of Information Systems**, v. 17, n. 5, p. 448-469, dez., 2008.
- ARAÚJO, W. F. G., LAIA, M. M. Governança eletrônica e gestão da informação em portais de governo: uma análise da reestruturação do Portal Minas. In: **Encontro Anual da Associação Nacional dos Programas de Pós-Graduação Em Administração**, 28., 2004, Curitiba. Anais... Rio de Janeiro: Anpad, 2004. p. 1-16.
- ASAMOAHA, D. A.; DORAN, D.; SCHILLER, S. *Teaching the foundations of data science: an interdisciplinary approach*. In: PRE-ICIS BUSINESS ANALYTICS CONGRESS CONFERENCE. Fort Worth, Texas, EUA. 2015.
- AUGUSTO, C. A. et al. Pesquisa qualitativa: rigor metodológico no tratamento da teoria dos custos de transação em artigos apresentados nos congressos da Sober (2007-2011). **RESR**, v. 51, n. 4, p. 745-764, out.-dez., 2013. Disponível em: <https://www.scielo.br/pdf/resr/v51n4/a07v51n4.pdf>. Acesso em: 18 jul. 2020.
- AYACHI, R. et al. *Proactive and reactive e-government services recommendation*. **Universal Access in the Information Society**, v. 15, n. 4, p. 681-697, nov., 2016.
- BARBALHO, F. A. A emergência do campo de políticas públicas de dados abertos governamentais no Brasil. **Conhecer: debate entre o público e o privado**, v. 8, n. 20, 2018.
- BARDIN, L. **Análise de conteúdo**. Edições 70: Lisboa, 1977.
- BARRETO JUNIOR, I. F.; RODRIGUES, C. B. Governo eletrônico, cidadania e inclusão digital. **Revista Direito e Justiça**, v. 12, n. 19, p. 91-112, mar., 2013.
- BARTSCHAT, A.; REISCHL, M.; MIKUT, R. Data mining tools. **WIREs Data Mining and Knowledge Discovery**, v. 9, n. 4, jul./ago., 2019.
- BAZILIAN, M. et al. *Open source software and crowdsourcing for energy analysis*. **Energy Policy**, v. 49, p. 149-153, out., 2012.
- BERMAN, F. et al. Realizing the potential of data science. **Communications of the ACM**, v. 61, n. 4, p. 67-72, abr., 2018.

BERNERS-LEE, T. **Linked Data Principles**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 07 jan. 2021.

BERTALANFFY, L. V. **Teoria geral dos sistemas: fundamentos, desenvolvimento e aplicações**. Petrópolis: Vozes, 2010.

BERTIN, P. R. B.; FORTALEZA, J. M.; SILVA, A. C.; OKAWACHI, M. F. A parceria para Governo Aberto como plataforma para o avanço da Ciência Aberta no Brasil. **Transformação**, v. 31, 2019. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862019000100301#B11. Acesso em: 16 jun. 2020.

BERTOT, J. C. *Big data, open government and e-government: Issues, policies and recommendations*. **Information Polity**, v. 19, n. 1, p. 5-16, jun., 2014.

BERNABE, M. **Defining the data science landscap**. 2017. Disponível em: <https://insidebigdata.com/2017/07/27/defining-data-science-landscape/>. Acesso em: 14 jan. 2021.

BERNSTEIN, A.; HENDLER, J.; NOY, N. *A new look at the semantic web: seeking to make web data "smarter" by utilizing a new kind of semantics*. **Communications of the ACM**, v. 59, n. 9, set., 2016.

BETTENCOURT, M. P. L.; CIANCONI, R. B. Gestão do conhecimento: um olhar sob a perspectiva da ciência da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 5, n. 1, 2012.

BIG data. Disponível em: : <https://bigdataldn.com/intelligence/big-data-the-3-vs-explained/>. Acesso em: 16 jan. 2021.

BLEI, D. M.; SMYTH, P. *Science and data science*. **PNAS**, v. 114, n. 33, ago., 2017.

BLOG FLAI: Comunidade de Ciência de Dados. Disponível em: <https://www.flai.com.br/juscudilio/como-a-inteligencia-artificial-esta-transformando-os-negocios/>. Acesso em: 16 jan. 2021.

BORGLUND, E.; ENGVALL, T. *Open data?: Data, information, document or record?*. **Records Management Journal**, v. 24, n. 2, p. 163-180, 2014.

BRASIL. **Padrões de Interoperabilidade de Governo Eletrônico**. 2018. Disponível em: <http://eping.governoeletronico.gov.br/>. Acesso em: 07 jan. 2021.

BRASIL. Controladoria Geral da União. **O que é Governo Aberto**. Disponível em: <https://governoaberto.cgu.gov.br/governo-aberto-no-brasil/o-que-e-governo-aberto>. Acesso em: 15 jun. 2020.

BRASIL. Controladoria Geral da União. **Plano de Ação: Governo Aberto**. Disponível em: <https://governoaberto.cgu.gov.br/a-ogp/planos-de-acao>. Acesso em: 16 jun. 2020.

BRASIL. Controladoria Geral da União. **Legislação: portal da transparência**.

Disponível em: <http://www.portaltransparencia.gov.br/sobre/legislacao>. Acesso em: 16 jun. 2020.

BUFREM, L. S.; SILVA, F. M.; SOBRAL, N. V.; CORREIA, A. E. G. C. Produção internacional sobre ciência orientada a dados: análise dos termos Data Science e E-science na Scopus e na Web of Science. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 40-67, maio/ago., 2016.

CÂMARA, M. A.; FERREIRA, M. A. T. Gestão do conhecimento: existe apenas uma?. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18. Marília, SP, 2017. Disponível em: http://enancib.marilia.unesp.br/index.php/XVIII_ENANCIB/ENANCIB/paper/viewFile/352/788. Acesso em: 08 jan. 2021.

CAO, L. *Data science and analytics: a new era*. **International Journal of Data Science and Analytics**, v. 1, n. 1, p. 1-2, abr. 2016.

CAO, L. Data science: a comprehensive overview. **ACM Computing Survey**, v. 50, n. 3, jun. 2017

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Rev. Adm. Pública**, v.42, n. 3, mai./jun., p. 495-528, 2008.

CASTELLS, M. **A Sociedade em rede**. 14^o reimpressão, São Paulo; Paz e Terra, I-XXX, 1999.

CERVO, A. L.; BERVIAN, P. A.; DA SILVA, R. **Metodologia Científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. *Business intelligence and analytics: from big data to big impact*. **MIS Quarterly**, v. 36, n. 4, p. 1165-1188, dez., 2012.

CHEN, M.; MAO, S.; LIU, Y. *Big data: a survey*. **Mobile Netw Appl**, v. 19, n. 2, p. 171-209, 2014.

CHOO, C. W. The knowing organization: how organizations use information to construct meaning, create knowledge and make decisions. **International Journal of Information Management**, v. 16, n. 5, p. 329-340, out., 1996.

CLEVELAND, W. S. *Data Science: an action plan for expanding the technical areas of the field of statistics*. **International Statistical Review / Revue Internationale de Statistique**, v. 69, n. 1, p. 21-26, abr., 2001.

CONEGLIAN, C. S.; DIEGER, R.; SANTARÉM SEGUNDO, J. E.; CAPTREZ, M. O papel estratégico da web semântica no contexto do *big data*. In: **I Workshop de Informação, Dados e Tecnologia**, Santa Catarina, 2017. *Anais...* Santa Catarina: Universidade Federal de Santa Catarina, 2017.

CONEGLIAN, C. S.; GONÇALVES, P. R. V. A.; SANTARÉM SEGUNDO, J. E. O profissional da informação na era do big data. **Encontros Bibli: revista eletrônica de**

biblioteconomia e ciência da informação, v. 22, n. 50, p. 128-143, set./dez., 2017.

CORDELLA, A.; IANNACCI, F. *Information systems in the public sector: the e-Government enactment framework*. **Journal of Strategic Information Systems**, v. 19, n. 1, p. 52-66, mar., 2010.

CORREA, A. S.; MELO JR., A.; SILVA, F. S. C. *A deep search method to survey data portals in the whole web: toward a machine learning classification model*. **Government Information Quarterly**, v. 37, n. 4, p. 1-15, 2020.

CORNELLA, A. *La información no es necesariamente conocimiento*. In: _____. (org.). **Infonomía: la empresa es información**. Barcelona: Deusto, 2000. p. 2-8. Disponível em:

<http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/3/autocapacitaciones/infomacion-conocimiento/capta.pdf>. Acesso em: 18 jul. 2020.

CRUZ, C. A.; NAGANO, M. S. Gestão do conhecimento e sistemas de informação: uma análise sob a ótica da teoria de criação do conhecimento. **Perspectivas em Ciência da Informação**, v. 13, n. 2, p. 88-106, mai./ago. 2008.

DAMATTA, R. (Coord.). **Sumário Executivo**: Pesquisa diagnóstico sobre valores, conhecimento e cultura de Acesso à Informação Pública no Poder Executivo Federal brasileiro. Brasília, 2011. Disponível em: <http://pfdc.pgr.mpf.mp.br/atuacao-e-conteudos-de-apoio/publicacoes/aceso-a-informacao/sumario-executivo-pesquisa-diagnostico-sobre-valores-conhecimento-e-cultura-de-aceso-a-informacao-publica-no-poder-executivo-federal-brasileiro-2013-coordenacao-professor-phd-roberto-damatta/view>. Acesso em: 15 jan. 2021.

DHAR, V. *Data science and prediction*. **Communications of the ACM**, v. 56, n. 12, p. 64-73, 2013.

ELEZAJ, O.; TOLE, D.; BACI, N. Big data in e-government environments: albania as a case study. **Academic Journal of Interdisciplinary Studies**, v. 7, n. 2, jul., 2018.

FARINELLI, F.; ALMEIDA, M. B. Interoperabilidade semântica em sistemas de informação de saúde por meio de ontologias formais e informais: um estudo da norma OpenEHR. In: Anais da IV Conferência BIREDIAL-ISTEC 2014. Acesso Aberto, Preservação Digital, Interoperabilidade, Visibilidade e Dados Científicos; Porto Alegre, RS, 2014. Disponível em http://mba.eci.ufmg.br/downloads/Biredial2014_144_web.pdf. Acesso em: 07 jan. 2021.

FAYYAD, U.; HAMUTCU, H. (2020). *Toward foundations for data science and analytics: a knowledge framework for professional standards*. **Harvard Data Science Review**, v.2, n. 2, 2020.

FERNANDES, C. C. C.. **Maturidade do governo eletrônico: análise de experiências estaduais**. Projeto de Pesquisa (Doutorado Gestão Estratégica da Informação) - Escola de Administração de Empresas de São Paulo da FGV, São Paulo, 2006. Disponível em:

<https://repositorio.enap.gov.br/bitstream/1/1694/1/Maturidade%20do%20governo%20eletr%C3%B4nico%20nos%20estados.pdf>. Acesso em: 16 jun. 2020.

FERNEDA, E. et al. Potential of ontology for interoperability in e-government: discussing international initiatives and the brazilian case. **Brazilian Journal of Information Science: Research Trends**, v. 10, n. 2, 2016.

GAMA, J. R. **Instrumentos de transparência e acesso às informações públicas: um estudo das demandas por informações contábeis nas universidades federais**. 2015. 189 f. Tese (Doutorado em Ciência da Informação) - Universidade de Brasília, Brasília, 2015. Disponível em: <http://repositorio.unb.br/handle/10482/19201>. Acesso em: 15 jan. 2021.

GANDOMI, A.; HAIDER, M. *Beyond the hype: Big data concepts, methods, and analytics*. **International Journal of Information Management**, v. 35, n. 2, p. 137-144, 2015.

GEORGE, G.; OSINGA, E. C.; LAVIE, D.; SCOTT, B. A. *Big data and data science methods for management research: From the Editors*. **Academy of Management Journal**, v. 59, n. 5, p. 1493-1507, 2016.

GILCHRIST, A. *Thesauri, taxonomies and ontologies: an etymological note*. **Journal of Documentation**, v. 59, n. 1, p. 7-18, 2003.

GÓMEZ, M. N. G. Regime de Informação: construção de um conceito. **Inf. & Soc.**, v. 22, n. 3, p. 43-60, set./dez., 2012.

GRUZOVA, A. A. *The role of the information service of an organization in introducing a corporate information system*. **Scientific and Technical Information Processing**, v. 39, n. 4, p. 199-203, 2012. Disponível em: <https://link.springer.com/article/10.3103/S0147688212040028>. Acesso em: 14 jun. 2020.

HABEEB, A. **Introduction to Artificial Intelligence**. 2017. Disponível em: https://www.researchgate.net/profile/Ahmed_Habeeb2/publication/325581483_Introduction_to_Artificial_Intelligence/links/5b16fdadaca272d24cc3a6c0/Introduction-to-Artificial-Intelligence.pdf. Acesso em: 08 jan. 2021.

HEADRICK, D. R. **When information came of age: Technologies of knowledge in the age of reason and revolution, 1700–1850**. Oxford: Oxford University Press, 2000.

HONGYAN, Y.; MINGDI, C. *Reuse of public sectors information in US: challenges and strategies to information resource management*. **Canadian Social Science**, v. 8, n. 5, p. 83-90, 2012. Disponível em: <http://www.cscanada.net/index.php/css/article/view/j.css.1923669720120805.ZT0608/3354>. Acesso em: 06 jan. 2021.

HUANG, R.; WANG, C.; ZHANG, X.; WU, D.; XIE, Q. *Design, develop and evaluate an open government data platform: a user-centred approach*. **The Electronic**

Library, v. 37, n. 3, p. 550-562, 2019.

JAIN, N.; SRIVASTAVA, V. *Data mining techniques: a survey paper*. **IJRET: International Journal of Research in Engineering and Technology**, v. 2, n. 11, nov., 2013.

JAINL, S.; KUMAR, P. *Semantic Web, Ontologies and E-Government: a review*. **Mody University International Journal of Computing and Engineering Research**, v. 2, n. 1, p. 40-44, 2018.

JANSSEN M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. *Benefits, adoption barriers and myths of open data and open government*. **Information Systems Management**, v. 29, n. 4, p. 258-268, 2012.

JAPIASSU, H. **Interdisciplinaridade e patologia do saber**. Rio de Janeiro: Imago, 1976.

JORDAN, M. I.; MITCHELL, T. M. *Machine learning: trends, perspectives, and prospects*. **Science**, v. 349, n. 6245, p. 255-260, jul., 2015.

KANT, I. **Crítica da Razão Pura**. 5. ed. Lisboa: Fundação Calouste Gulbenkian, 2001.

KASSEN, M. *Open data and its peers: understanding promising harbingers from Nordic Europe*. **Aslib Journal of Information Management**, v. 72, n. 5, p. 765-785, 2020.

KAYA, T. *Artificial intelligence driven e-government: the engage model to improve e-decision making*. In: EUROPEAN CONFERENCE ON DIGITAL GOVERNMENT, 19. Nicósia, Chipre, 2019. Disponível em: https://www.researchgate.net/publication/336831763_Artificial_Intelligence_driven_E-Government_The_Engage_Model_to_Improve_E-Decision_Making. Acesso em: 09 jan. 2021.

LAND, F. Is an information theory enough? **The Computer Journal**, v. 28, n. 3, p. 211-215, (1985).

LAYNE, K.; LEE, J. *Developing fully functional e-government: a four stage model*. **Government Information Quarterly**, v. 18, n. 2, p. 122-136, 2001.

LIMA, G. A.; MACULAN, B. C. M. S. Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. **Ci.Inf.**, Brasília, DF, v.46, n.1, p.60-60, jan./abr. 2017.

LIN WANG. Twinning data science with information science in schools of library and information science. **Journal of Documentation**, v. 74, n. 6, p.1243-1257, 2018.

LNENICKA, M.; KOMARKOVA, J. *Big and open linked data analytics ecosystem: Theoretical background and essential elements*. **Government Information Quarterly**, v. 36, n. 1, p. 129-144, 2019.

MAALI, F.; CYGANIAK, R.; PERISTERAS V. Enabling Interoperability of Government Data Catalogues. In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT, 9. Suíça, 2010. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-14799-9_29. Acesso em: 07 jan. 2021.

MAHMOUD, E. B. M.; OMAR, E. B.; OUKSEL, A. M. Big data interoperability for E-Governance. **Journal of Computer Science**, v. 15, n.10, p. 1430-1438, 2019.

MARQUES, M. B. Gestão da informação em sistemas de informação complexos. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 12, n. 2, p. 060-076, 2017.

MATHEUS, R.; JANSSEN, M.; MAHESHWARI, D. *Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities*. **Government Information Quarterly**, v. 37, n. 3, jul., 2020.

MATHEUS, R.; RIBEIRO, M. M.; VAZ, J. C. *Strategies and instruments for the dissemination and promotion of open government data use in Brazil: case study of Rio de Janeiro city hall*. **R. Technol. Soc.**, v. 14, n. 33, p. 172-189, jul./set., 2018.

MAXIMIANO, A. C. A. **Introdução a administração**. 3. ed. São Paulo: Editora Atlas, 1992.

MAZZOCHI, F. **Knowledge organization system (KOS)**. Disponível em: <https://www.isko.org/cyclo/kos>. Acesso em: 09 jan. 2021.

MEDEIROS, J. S.; CAREGNATO, S. E. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, v. 8, n. 2, p. 311-322, set. 2012.

MEDEIROS, A. K.; NONATO, R. S.; ZATZ, I. B.; SOUZA, Z. B. Transparência de compras públicas sustentáveis: um caminho a ser traçado. **Liinc em Revista**, v. 9, n. 2, p. 452-468, nov. 2013.

MISNEVS, B.; YATSKIV, I. *Data science: professional requirements and competence evaluation*. **Baltic J. Modern Computing**, v. 4, n. 3, p. 441-453, 2016.

MISRA, D. C.; HARIHARAN, R.; KHANEJA, M. E-Knowledge management framework for government organizations. **Information Systems Management**, v. 20, n. 2, p. 38-48, 2003.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G.; SANTARÉM SEGUNDO, J. E. e-Science semântica: integração dos dados na comunicação científica. **Inf. Pauta**, Fortaleza, CE, v. 1, n. 1, jan./jun. 2016.

MOREIRA, F. M.; VALENTIM, M. L. P.; SANT'ANA, R. C. G. A interdisciplinaridade da Ciência da Informação e suas contribuições no estudo do compartilhamento de dados governamentais na internet. **Em Questão**, Porto Alegre, v. 24, n. 2, p. 300-

329, maio/ago., 2018.

MOTTA, F. C. P. A teoria geral dos sistemas na teoria das organizações. **Rev. Adm. Empres.**, v. 11, n. 1, jan./mar., 1971. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75901971000100003. Acesso em: 14 jun. 2020.

OLIVEIRA, A. L. T.; ROSA, M. N. B.; OLIVEIRA, B. M. J. F.; LIMA, I. F. Patrimônio cultural imaterial na perspectiva da ciência da informação: análise das produções científicas apresentadas no grupo de trabalho informação e memória no encontro nacional de pesquisa em ciência da informação. **Revista Analisando em Ciência da Informação**, v. 5, n. 2, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/80762>. Acesso em: 16 mar. 2021.

ONU – Organização das Nações Unidas. **17 objetivos para transformar nosso mundo**. Disponível em: <https://nacoesunidas.org/pos2015/>. Acesso em: 23 mai. 2020.

Open Government Partnership. **Open Government Declaration**. 2011. Disponível em: <https://www.opengovpartnership.org/process/joining-ogp/open-government-declaration/>. Acesso em: 16 jun. 2020.

PÁGINA OpenData.info. Disponível em: <https://opendata.info/>. Acesso em: 15 jan. 2021.

PARYCEK, P.; HOCHTL, J.; GINNER, M. *Open government data implementation evaluation*. **Journal of theoretical and applied electronic commerce research**, v. 9, n. 2, mai., 2014. Disponível em: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0718-18762014000200007. Acesso em: 15 jun. 2020.

PRIESTLEY, J.; McGRATH, R. J. The evolution of data science: a new mode of knowledge production. **International Journal of Knowledge Management**, v. 15, n. 2, abr./jun., 2019.

PROVOST, F.; FAWCETT, T. *Data Science and its relationship to big data and data-driven decision making*. **Big Data**, v. 1, n. 1, p. 51-59, 2013.

RAU, L. F. *Knowledge organization and access in a conceptual information system*. **Information Processing & Management**, v. 23, n. 4, p. 269-283. 1987. Disponível em: <https://www-sciencedirect.ez3.periodicos.capes.gov.br/science/article/pii/0306457387900185?via%3Dihub>. Acesso em: 23 mai. 2020.

RAUTENBERG, S.; CARMO, P. R. V. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Studies: Research Trends**, v. 13, n.1, p.56-67, 2019.

RAYWARD, W. B. *Information revolutions, the information society, and the future of the history of information science*. **Library Trends**, v. 62, n. 3, 2014.

REGINATO, L.; NASCIMENTO, A. M. Um estudo de caso envolvendo Business Intelligence como instrumento de apoio à controladoria. **Rev. contab. finanç.**, v.18, n. spe, jun., 2007.

REGO, H. O.; FREIRE, I. M. *Accountability*: novo conceito para a ciência da informação? **Ci. Inf. Rev.**, Maceió, v. 5, n. 1, p. 29-40, jan./abr., 2018.

ROCHA, H. H. N. Transparência e *accountability* no Estado Democrático de Direito: reflexões à luz da Lei de Acesso à Informação. **Revista TCEMG**, Edição Especial 2012, p. 90. Disponível em: <https://revista1.tce.mg.gov.br/Content/Upload/Materia/1682.pdf>. Acesso em: 14 mar. 2021.

RODRIGUES, F. A. Restrições tecnológicas e de acesso a dados disponíveis sobre destinos de repasses financeiros federais para a saúde pública em ambientes informacionais digitais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13. Rio de Janeiro, 2012. Disponível em: <http://www.eventosecongressos.com.br/metodo/enancib2012/arearestrita/pdfs/19435.pdf>. Acesso em: 20 jul. 2020.

SÁNCHEZ-CUADRADO, S. et al. *De repente, ¿ todos hablamos de ontologías?*. **El Profesional de la Información**, v.16, n. 6, nov./dez., 2007

SANT'ANA, R. C. G. **Mensuração da disponibilização de informações e do nível de interação dos ambientes informacionais digitais da administração municipal com a sociedade**. Tese de Doutorado em Ciência da Informação – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília, São Paulo, 2008. 153 p.

SANT'ANA, R. C. G.; RODRIGUES, F. A. Uso de modelos de dados multidimensionais para ampliação da transparência ativa. **Liinc em Revista**, v. 9, n. 2, p. 469-487, nov., 2013.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 116-142, maio/ago., 2016.

SANTOS, E. M. Implementing interoperability standards for electronic government: an exploratory case study of the e-PING Brazilian framework. **International Journal of Electronic Government Research**, v.4, n. 3, p. jul./set., 2008.

SANTOS, P. L. V. A. C.; SANTANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela ciência da informação. **Ciência da Informação**, v. 42, n. 2, 2013.

SCHULZ, M. *The uncertain relevance of newness: organizational learning and knowledge flows*. **Academy of management journal**, v. 44, n. 4, p. 661-681, ago., 2001.

SILVA, A. M. Ciência da informação e sistemas de informação: (re)exame de uma

- relação disciplinar. **Prisma.com** (Portugal), n. 5, p. 2-46, 2007. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/70366>. Acesso em: 13 jun. 2020.
- SILVA, A. P.; SANTOS, J. C.; KONRAD, M. R. Teoria geral dos sistemas: diferencial organizacional que viabiliza o pleno entendimento da empresa. **Educação, Gestão e Sociedade**, v. 6, n. 22, jun. 2016. Disponível em: <http://uniesp.edu.br/sites/biblioteca/revistas/20170509162834.pdf>. Acesso em: 14 jun. 2020.
- SILVA, G. B.; CUNHA, A. M.; OLIVEIRA, J. M. P. Um modelo para interoperabilidade semântica no e-gov brasileiro. In: SIMPÓSIO DE APLICAÇÕES OPERACIONAIS EM ÁREAS DE DEFESA, 12. São José dos Campos, SP, 2011. Disponível em: https://www.researchgate.net/publication/267333446_Um_Modelo_para_Interoperabilidade_Semantica_no_e-Gov_Brasileiro. Acesso em: 07 jan. 2021.
- SMITH, B.; WELTY, C. *Ontology: towards a new synthesis*. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 2. Ogunquit, Maine, EUA, 2001. Disponível em: <https://philpapers.org/archive/SMIOTA-9.pdf>. Acesso em: 09 jan. 2021.
- SOUZA, R. R.; ALMEIDA, M. B.; BARACHOS, R. M. A. Ciência da informação em transformação: *big data*, nuvens, redes sociais e *web* semântica. **Ci. Inf.**, v. 42, n. 2, p. 159-173, mai./ago., 2013.
- TAURION, C. **Big data**. 1. ed. Rio de Janeiro: Brasport, 2013.
- TECUCI, G. *Artificial intelligence*. **WIREs Computational Statistics**, v. 4, n. 2, 2011.
- TEIXEIRA, E. B. A análise de dados na pesquisa científica: importância e desafios em estudos organizacionais. **Desenvolvimento em Questão**, v. 1, n. 2, jul./dez., 2003.
- THOMAS, M. A.; CIPOLLA, J.; LAMBERT, B.; CARTER, L. *Data management maturity assessment of public sector agencies*. **Government Information Quarterly**, v. 36, n. 4, 2019.
- VERGARA, S. M. C. **Projetos e relatórios de pesquisa em administração**. 5.ed. São Paulo: Atlas, 2004.
- VERMA, K. K.; SHRIVASTAVA, N. *Exploring Role and associated challenges of data mining technique and its implementation in e-governance*. **International Journal of Advanced Research in Computer Engineering & Technology**, v. 4, n. 8, ago., 2015.
- VIRKUS, S.; GAROUFALLOU, E. *Data science from a library and information science perspective*. **Data Technologies and Applications**, v. 53, n. 4, p. 422-441, 2019.
- VIRKUS, S.; GAROUFALLOU, E. *Data science and its relationship to library and information science: a content analysis*. **Data Technologies and Applications**, v. 54, n. 5, p. 643-663, 2020.
- W3C - *World Wide Web Consortium*. **W3C semantic web activity**. Disponível em:

<https://www.w3.org/2001/sw/>. Acesso em: 07 jan. 2021.

WAGNER, C. Knowledge management in e-government. In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS, 9. Tampa, Flórida, EUA, 2003. Disponível em: <https://aisel.aisnet.org/amcis2003/105/>. Acesso em: 08 jan. 2021.

WANG, C. *A novel approach to conduct the importance-satisfaction analysis for acquiring typical user groups in business-intelligence systems.* **Computers in human behavior**, v.54, p.673-681, jan., 2016.

ZINS, C. *Conceptual approaches for defining data, information, and knowledge.* **Journal of the American Society for Information Science and Technology**, v.58, n.4, p.479-493, 2007.