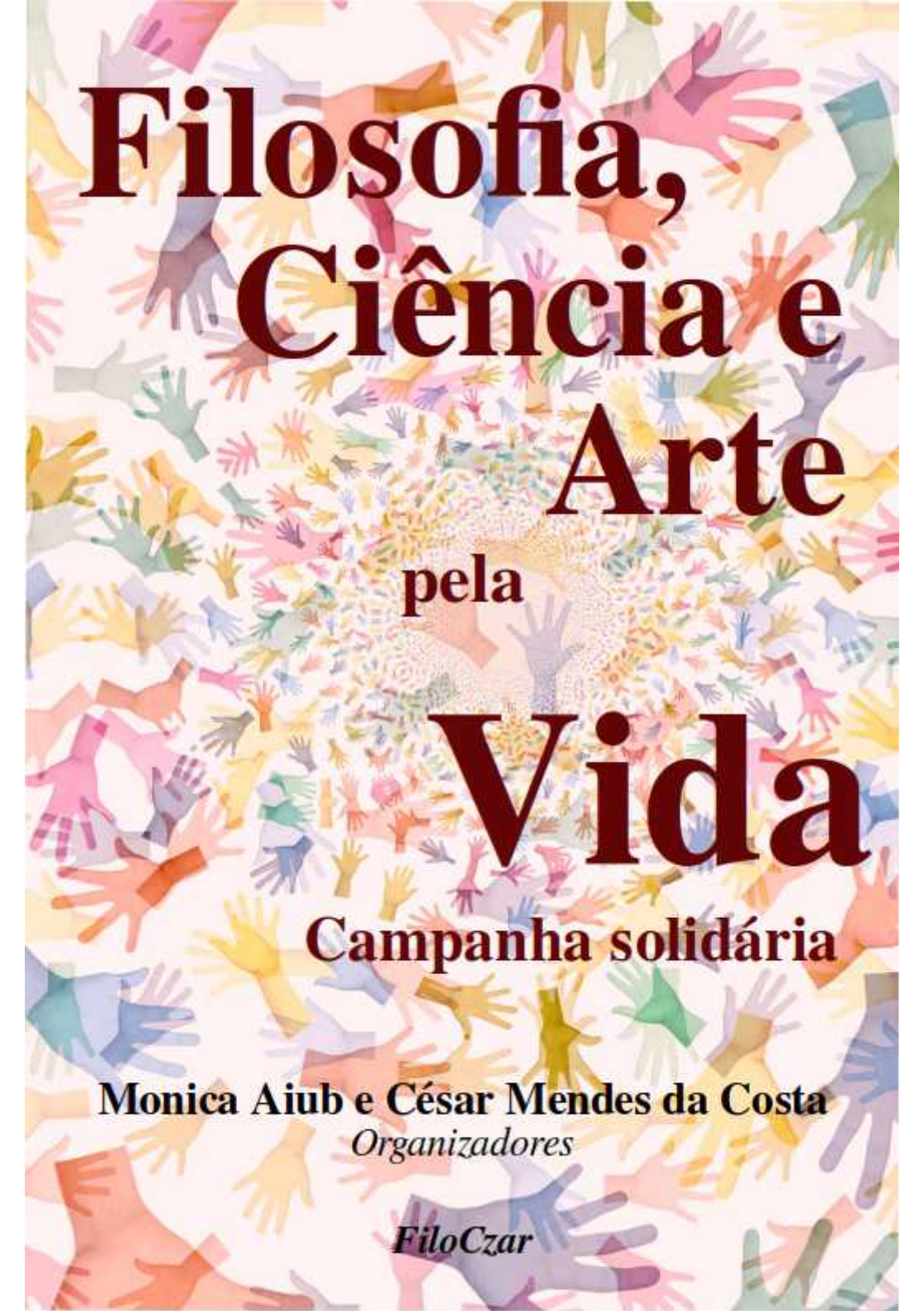




Filosofia, Ciência e Arte pela Vida

Campanha solidária

Monica Aiub e César Mendes da Costa
Organizadores

FiloCzar

Monica Aiub e César Mendes da Costa
Organizadores

Filosofia, Ciência e Arte pela Vida

São Paulo
FiloCzar
2020

Copyright © 2020 by *FiloCzar*

Editores: César Mendes da Costa e Monica Aiub da Costa

Revisão: Monica Aiub

Projeto Gráfico: Editora FiloCzar

Imagem de Capa: Pixabay/Domínio Público

Dados Internacionais de Catalogação na Publicação (CIP)

A258f Aiub, Monica; Costa, César Mendes da. (org).
Filosofia, ciência e arte pela vida. Cláudio Laureatti et all
-- São Paulo: FiloCzar, 2020.

ISBN 978-65-87117-30-0

Formato: Livro Digital – PDF

Veiculação: Digital

Filosofia. 2. Ciência. 3. Arte. I. Título. II. Laureatti, Cláudio; III. Cerqueira, Augusto; IV. Camargo, José Jorge Guedes de; V. Mendes, Zuleide; VI. Domezi, Geraldo; VII. Rótherdan, Tio Árlex de; VIII. Sweet, Mary; IX. Gottlieb, Liana; X. Vicente, Leandro Raphael; XI. Freitas, Bárbara; XII. Carvalho, José Maurício de; XIII. Denis, Leon; XIV. Santos, Cláudia Ferreira dos; XV. Rodrigues, Mariana Vitti; XVI. Cavassane, Ricardo Peraça; XVI. Rodrigues, Fernando de; XVII. Sant'Ana, Ricardo César Gonçalves; XVIII. Correia, Ricardo Lopes.

CDD 100

Índices para catálogo sistemático:

1. Filosofia CDD 100
2. Ciência CDD 500
3. Arte CDD 700

FiloCzar

Rua Durval Guerra de Azevedo, 511 – Parque Santo Antônio
São Paulo – SP

CEP: 05852-440

Tels.: (11) 5512-1110 - 96781-9707

E-mail: cesar@editorafiloczar.com.br

www.editorafiloczar.com.br

Sobre a campanha solidária

A Livraria & Editora FiloCzar está localizada na periferia da zona sul paulistana - Parque Santo Antônio. Nesta região, há muitas famílias em situação de vulnerabilidade, sofrendo neste momento de pandemia. Pensando em formas de auxiliar a comunidade a partir de nosso papel de editores, propusemos a produção deste e-book: **Filosofia, Ciência e Arte pela Vida**, distribuído gratuitamente para promover a leitura e a reflexão pela vida. Paralelamente, iniciamos uma campanha para arrecadar doações para auxiliar as famílias necessitadas neste período.

De imediato, vários autores responderam, contribuindo com textos poéticos, filosóficos, científicos, com histórias de vida... De imediato, também, várias pessoas se dispuseram a ajudar com doações para as famílias em situação de vulnerabilidade da região. Nossos sinceros agradecimentos a todos os que contribuíram.

Mas como a pandemia e as necessidades impostas por este contexto não cessaram, você pode participar da campanha enviando suas doações diretamente à Paróquia Santo Eugênio de Mazenod, que tem um importante papel social na região, acolhendo as famílias necessitadas e distribuindo alimentos. Se puder doar alimentos e estiver na região, o endereço para entrega é: Rua Deocleciano de Oliveira Filho, 249 - Parque Santo Antônio - São Paulo - SP. Se estiver longe e quiser contribuir financeiramente, a conta da Paróquia Santo Eugênio de Mazenod, para depósito/transferência, de qualquer valor, é:

Banco Itau - 341

Agência: 1667

Conta: 06705-0

CNPJ: 61.378.766/0059-15

Razão Social: Mitra Diocesana de Campo Limpo

Você também pode participar compartilhando este e-book com outras pessoas e promovendo a vida através da Filosofia, da Arte e da Ciência.

Nossos agradecimentos a todos aqueles que, de alguma maneira, contribuem para levar solidariedade, informação, auxílio, arte, reflexões, vida àqueles que sofrem mais duramente o impacto deste momento tão difícil.

*André Cavalcante era professor
Amigo de todos e pai do Pedrinho.
O Bruno Campelo seguiu seu caminho
Tornou-se enfermeiro por puro amor.
Já Carlos Antônio, era cobrador
Estava ansioso pra se aposentar.
A Diva Thereza amava tocar
Seu belo piano de forma eloquente
Se números frios não tocam a gente
Espero que nomes consigam tocar...*

Inumeráveis (Braulio Bessa/Chico César)

A todos que se foram
Aos que perderam seus amores
Aos que não ficam indiferentes diante da dor
Diante de um mundo que se esvai
Aos que estendem as mãos
Dispõem-se ao outro
Entoam seus cantos
Distribuem afetos
Compartilham saberes
Semeiam vida



Fotografia de Wagner Brito

Sumário

Apresentação

Monica Aiub e César Mendes da Costa

#Pneumotórax#DeNovo

Cláudio Laureatti

Corneto ao 'novo normal'

Augusto Cerqueira

Quatro reflexões em um Brasil pandêmico

José Jorge Guedes de Camargo

Labirinto

Zuleide Mendes

Nonormal e Nonada

Geraldo Domezi

As máscaras e o mundo - Uma necessidade do ano de 2020

Tio Árlax de Rótherdan

Ah! Minha paixão por livros... (e como achei minha Alma de volta)

Mary Sweet

A arte do tempo e a arte de viver em meio à pandemia da COVID-19 - Vivendo, Amando e Aprendendo

Liana Gottlieb

A paciência, o Hulk e os estoicos

Leandro Raphael Vicente

Como aprendi a lidar com o medo da morte ao sair das Testemunhas de Jeová

Bárbara Freitas

O pseudodilema: vida ou economia?

Monica Aiub

O mundo na e pós pandemia

José Mauricio de Carvalho

Pensar a vida: da Grécia Antiga à periferia de São Paulo

Leon Denis

Política

César Mendes da Costa

Filosofia, Ciência e Arte pela Vida: uma descoberta sobre nós mesmos

Cláudia Ferreira dos Santos

A arte do hábito

Mariana Vitti Rodrigues

Lave bem as mãos, mantenha a distância e não leia os comentários

Ricardo Peraça Cavassane

Traduzindo o termo DATA para o cotidiano

Ricardo César Gonçalves Sant'Ana

Fernando de Assis Rodrigues

‘Inclusão injusta’ e o ‘direito à cidade’ – “me inclua fora dessa!”: reflexões sobre os discursos de uso e de troca do bem público de direito social

Ricardo Lopes Correia

Traduzindo o termo DATA para o cotidiano

*Ricardo César Gonçalves Sant'Ana*²⁸

*Fernando de Assis Rodrigues*²⁹

Introdução

Dados, dados e mais dados. Dados por todos os lados, notícias, opiniões, decisões e até quando tentamos construir uma percepção sobre as mais diferentes questões os dados são utilizados para sustentar os principais conceitos envolvidos.

A participação dos dados no nosso cotidiano não chega a ser uma novidade. Desde sempre nos baseamos neles para estabelecer nossa percepção de mundo e na construção de opiniões e na tomada de decisões. No entanto, cada vez mais disponíveis, os dados passaram a fazer parte do cotidiano de todos e em todas as esferas de ação. Todos os dias somos colocados em contato com informações baseadas em dados e, com eles ganhando o protagonismo crescente, somos impelidos a tentar entender do que se trata. Mas nem sempre é tarefa fácil, são muitos 'DATA-alguma-coisa' definindo conceitos que encerram em si outros conceitos relacionados ao que podemos fazer com os dados.

Esta inundação de citações ao termo 'DATA' vem crescendo de forma exponencial: são postagens, notícias, artigos, livros e todos trazendo os 'DATA-alguma-coisa' como se todos tivessem a obrigação de conhecer o significado destes conceitos.

Muitos destes conceitos são oriundos de situações reais e resultantes de fatores que os avanços tecnológicos proporcionam, enquanto outros são definidos por empresas (as tão famosas *buzzwords*). Uma questão que emerge deste cenário é que muitas vezes utilizados como argumento de venda, estes conceitos acabam por receber conotações que se sobrepõem aos sentidos denotativos.

Nesse sentido, apresenta-se uma seleção de conceitos que permeiam a questão dos dados, especialmente aqueles que têm

²⁸ Universidade Estadual Paulista (UNESP). Contato: ricardo.santana@unesp.br

²⁹ Universidade Federal do Pará (UFPA). Contato:fernando@rodrigues.pro.br

relação direta ao uso de dados em ambientes informacionais digitais. Os conceitos foram selecionados a partir do critério de maior presença em publicações, em especial livros, e, no final, apresentamos um exemplo de uso de dados para ilustrar, exatamente, esta questão da presença do termo ‘DATA’ em livros. Cabe lembrar que são definições simplificadas de cada um dos conceitos, e que tiveram como objetivo uma visão geral de cada termo, mas que permita ao leitor se posicionar quando encontrar os tão presentes ‘DATA-alguma-coisa’.

Conceitos

As atividades da fase inicial de planejamento e execução da obtenção dos dados e que representam a fase de coleta do ciclo de vida dos dados é denominada de **Data Collection** e tem se configurado como a fase de maior desenvolvimento com a evolução das Tecnologias de Comunicação e Informação (TIC). Este desenvolvimento é um dos responsáveis, junto com a fase de armazenamento, pelo aumento exponencial de volume de dados disponíveis, pela facilidade de compartilhar dados em um clique e que tem proporcionado todo um novo cenário na forma como interagimos com a informação. Coletamos dados em todos os lugares, nos mais diversos formatos, dos mais diversos tipos, e a todo momento, uma verdadeira presença ubíqua de processos de coleta: sistemas administrativos, aplicativos móveis, serviços de redes sociais on-line, câmeras, sensores por todos os lados - onde todos somos geradores cientes e inscientes de dados (CRISTESCU et al., 2006; GOG et al., 2007; RODRIGUES; SANT’ANA, 2016, 2018; WICKRAMASURIYA et al., 2004).

Assim, este desenvolvimento das tecnologias digitais propiciou também um aumento nas capacidades de coletar, armazenar e de disponibilizar dados, inclusive superando nossas capacidades para tratar e interpretá-los. Este contexto é caracterizado por um aumento no potencial de acesso a grandes Volumes de dados, criados e acessados a grandes Velocidades, e disponíveis em grande Variedade de tipos e formatos, de onde identificamos os 3 Vs (Volume, Velocidade e Variedade). Este fenômeno vem sendo denominado de **Big Data** e retrata novas perspectivas de uso da tecnologia para acesso e interpretação desta

abundância de dados com novos e promissores resultados esperados para as mais variadas aplicações. Tanto o setor empresarial quanto a academia têm usado este termo para destacar a utilização de grandes quantidades de dados e como forma de valorizar seus resultados (COX; ELLSWORTH, 1997; DOBBS et al., 2011; DUMBILL, 2012; HOWE et al., 2008; KOLB; KOLB, 2013; LLAVE, 2018; LOH, 2014; LYCETT, 2013; O'REILLY, 2012).

Esta abundância de dados armazenados e disponíveis para recuperação também apresenta desafios para o processo de individualização de dados. O **Small Data** está relacionado com a capacidade de identificar tendências e padrões a partir de pequenos conjuntos de dados como, por exemplo, a personalização de propagandas a partir de históricos de pesquisa, tendências de doenças a partir de dados de sensores corporais, entre outros. As análises voltadas ao Small Data permitem a segmentação de parte dos dados armazenados em bancos de dados para personalizar as experiências de um grupo ou apresentar novas percepções sobre um determinado comportamento de um indivíduo (LINDSTROM; HEATH, 2016).

Atrrelada a este aumento do potencial de acesso a dados, a evolução da capacidade e eficiência dos recursos de comunicação proporcionou um repensar sobre onde poderíamos alocar nossos recursos físicos para armazenar nossos dados e até mesmo nossas aplicações. Os investimentos antes feitos para manter estruturas locais, principalmente para armazenamento de dados, a partir do desenvolvimento da conectividade, poderiam ser poupados por meio do acesso remoto a eles. Assim, o investimento em infraestrutura pode ser centralizado, inclusive com a possibilidade de terceirização, permitindo que todo o acesso ocorra por meio de comunicação. Esta infraestrutura centralizada (hardware e software), passa a ser denominada **Data Center** e promete redução de custos de aquisição, de implementação e até de manutenção. Como ganho adicional, pode-se pensar ainda nas reduções de custo dos dispositivos utilizados pelos usuários e a facilitação para mobilidade de acesso (CISCO INC., 2020; GREENBERG et al., 2008).

Já o **Data Warehouse** é um armazém de dados, formado por um banco de dados integrado e orientado a assuntos, armazenados em ambiente e formato diferentes daqueles utilizados nos sistemas de suporte às atividades corriqueiras da

empresa. Tem como principal função atender consultas que requerem como resposta cálculos sobre grandes quantidades de dados sistematizados. Seu formato diferenciado é justamente para que se tenha um bom tempo de resposta a estas consultas. Portanto, seu funcionamento é todo pensado para oferecer respostas aos usuários, geralmente gestores ou responsáveis por decisões. Assim, os dados das operações do dia a dia da empresa (transacionais) são sistematizados e copiados para estes grandes armazéns de dados, daí a origem do termo (BARBIERI, 2011; INMON, 2005; KIMBALL; ROSS, 2011; MARIBEL; RAMOS, 2009; RODRIGUES, 2017; RODRIGUES; SANT'ANA, 2018).

Assim como no caso do Data Warehouse o **Data Mart** também tem como função atender a demandas de consultas. O que os diferencia é que um Data Mart atende a um foco mais específico, como, por exemplo, valores relacionados a vendas, produção, inadimplência, compras ou qualquer outro processo relacionado às operações da entidade. Portanto o Data Mart é como que um recorte do Data Warehouse, sendo que alguns autores chegam a ponderar que um Data Warehouse poderia ser considerado um conjunto de Data Marts. Portanto os dados que compõem o Data Mart têm origem em banco de dados transacionais. A sua composição é alicerçada em bases de dados que organizam seus conteúdos, em função de estruturas bem específicas, com um conjunto de dados denominado fato, que identifica seu foco (vendas por exemplo) e outros conjuntos de dados que são as dimensões (tempo, região, produto, representante seriam exemplos para o fato vendas). Estes conjuntos denominados como fato se configuram como ponto de partida, um centro, na construção de modelos multidimensionais, em que são armazenados somente conteúdos quantificáveis e conteúdos que permitam vinculação (chaves estrangeiras) com os conjuntos de dados denominados como dimensões (INMON, 2005; KIMBALL; ROSS, 2011; RODRIGUES, 2017; RODRIGUES; SANT'ANA, 2013, 2018; SANTOS; SANT'ANA, 2015).

As tecnologias digitais ampliaram o volume de dados que podemos ter acesso e estes volumes crescentes permitem que um volume também crescente de interpretações, possa ser estabelecido a partir deles. Esta busca por interpretações, realizadas com o auxílio da própria tecnologia, pode resultar em respostas a questões

postuladas previamente e pode, ainda, resultar em descobertas valiosas e inesperadas. Tal como em um processo de mineração em que se têm estimativas e indícios de quais resultados podemos esperar de uma exploração, na mineração de dados podemos ser surpreendidos por informações inesperadas que podem vir na identificação de padrões não previstos, ou indícios de fatos desconhecidos ou qualquer outra contribuição valiosa para o conhecimento sobre uma determinada situação ou contexto. Este potencial de exploração nos dados, uma forma de mineração de dados é denominada **Data Mining**, processo de aplicação de algoritmos com métodos estatísticos e probabilísticos em conjuntos de dados. Estas descobertas como, por exemplo, a identificação de uma sazonalidade ou de uma correlação entre duas variáveis ganham relevância no entendimento de fenômenos, situações ou determinados contextos, agregando valor ao processo de acesso a dados, em especial para seu uso estratégico nas organizações (HAN; KAMBER, 2012; HILL; LEWICKI, 2006; PROVOST; FAWCETT, 2013a).

Com o aumento da capacidade de armazenamento de dados, surgem repositórios desenvolvidos para grandes quantidades dados estruturados (como os formatos gerados por editores de planilhas e bancos de dados) ou não estruturados (como arquivos gerados por processadores de texto), em formatos variados. Esse tipo de recurso, denominado de **Data Lake**, o que em uma tradução literal seria um lago de dados, oferece tanto o local de armazenamento dos dados como a capacidade de processá-los para a realização de análises (ANNE LAURENT, 2020; GORELIK, 2019; LLAVE, 2018).

E o volume de dados armazenados cresce tanto, que parte deles acabam por passar despercebidos não só pelos usuários mas pelos próprios profissionais envolvidos em sua gestão. Tais conjuntos de dados pode ser denominados como **Dark Data** e, apesar de disponíveis para acesso, acabam não sendo utilizados. Representam parcela dos dados coletados que estão além da nossa capacidade de utilizá-los, seja por limitações tecnológicas, de formato, de suporte, de custo, de divergências conceituais ou mesmo de aderência às nossas necessidades informacionais. Descartar tais dados não é uma decisão fácil já que podem ser ricos recursos de informação confiável e vir a ganhar relevância

em momentos futuros. Na ciência, o conceito de Dark Data está relacionado à quantidade de dados que é gerada nas pesquisas científicas que, apesar de estar em evidência o apoio à disponibilização destes conjuntos de dados à comunidade e a importância das informações que estes contém, ainda é pouco aproveitada (HEIDORN, 2008; PLATTS, 2018; SACOLICK, 2013; SCHEMBERA; DURÁN, 2020).

Todo este desenvolvimento tecnológico com o acesso a dados abrindo tantas possibilidades configurou, também, um cenário complexo o suficiente para que estivesse ao alcance de uma só área, o que levou à composição, de certa forma espontânea, de uma articulação interdisciplinar envolvendo competências como as tratadas pela Ciência da Computação, Matemática, Administração, Ciência da Informação e outras, compondo novos campos de atuação. Passa-se, então a denominar **Data Science** a este novo amálgama de perfis e de técnicas necessários para se alcançar resultados para o acesso e uso de tamanha abundância de dados. Com a presença crescente do uso de dados nas mais distintas esferas do cotidiano, a Ciência dos Dados passa a ser cada vez mais usada para se referir a estes esforços interdisciplinares (vide Quadro 1) e por sua recente existência ainda não alcançou consenso sobre seu campo conceitual ou sobre a parcela de participação de cada área na sua composição, o que não impede o surgimento de pesquisas, de cursos e de profissões relacionadas a ela (GRUS, 2015; PROVOST; FAWCETT, 2013a; SCHUTT; O'NEIL, 2013).

Todo este esforço para aplicação da Ciência dos Dados no acesso e interpretação dos dados, contando com a participação de competências advindas de distintas áreas como a Computação, a Matemática, Administração e outras, permitem que estes dados sejam analisados e disponibilizados no atendimento de necessidades informacionais dos usuários, e é denominado **Data Analytics**, compreendendo, portanto, técnicas e conhecimentos necessários para este processo de uso dos dados (DAMIEN, 2019; MAHESHWARI, 2014; PROVOST; FAWCETT, 2013a; SHARDA; DELEN; TURBAN, 2018). O processo de análise de dados propriamente dito também pode ser denominado de análise de dados ou **Data Analysis** e também tem forte interesse

(vide Quadro 1) (GRUS, 2015; HILL; LEWICKI, 2006; PROVOST; FAWCETT, 2013a).

Os profissionais preparados para participar dos processos de análises de dados são os Analistas de Dados, Data Analysts, e precisam ser capazes de aplicar as competências necessárias e ainda ter disponibilidade para trabalhar colaborativamente em equipes multidisciplinares ajudando pessoas de toda a empresa a entender consultas específicas com relatórios e gráficos ad-hoc (BERMAN, 2013; MAHESHWARI, 2014; RASMUSSEN et al., 2019).

Com esta diversidade de competências envolvidas nos processos de análises de dados, alguns perfis profissionais acabam por se adaptar em determinadas dimensões das áreas envolvidas. Assim, mesmo com uma eminente característica interdisciplinar, as especificidades de cada indivíduo envolvido permitem que suas contribuições tenham um viés mais para a dimensão matemática, outros para a dimensão de processos, enquanto outros para a dimensão tecnológica. Estes últimos, que tendem a se aproximar dos requisitos computacionais, se aproximam da definição do Data Engineer, por consequência, do que se denomina **Data Engineering**. Esses profissionais colaboram em atividades como o desenvolvimento de arquiteturas de sistemas de informação distribuídas para aumentar a capacidade de processamento e diminuir o tempo de resposta, a combinação de conjuntos de dados de origens diferentes, bem como apresentação de novas soluções a partir dos conjuntos de dados disponíveis. Com o crescimento das demandas por este tipo de especialidade, surgiram cursos livres e de especialidades para formar profissionais com as competências necessárias ao desempenho destas atividades (CHAN; TALBURT; TALLEY, 2010; SHIVE, 2013; WOLKENHAUER, 2001).

Todo este cenário de acesso e uso de dados também impacta nas áreas envolvidas. Na Ciência da Computação, o termo **Data-driven** é utilizado para identificar processos que tenham como diretriz o uso de dados, seja para tomada de decisões, para controle e acompanhamento ou mesmo para planejamento. Com a escalada do uso de dados em atividades que até então eram predominantemente realizadas com interferência humana, tais como nos processos de decisão, esta perspectiva de uso dos dados vem sendo implementada

nos principais sistemas das organizações, tais como os Sistemas Integrados de Gestão (Enterprise Resource Planning - ERP), Gestão de Relacionamento com Clientes (Customer Relationship Management - CRM) e Gestão de Cadeia de Suprimentos (Supply Chain Management - SCM), definindo um conjunto de requisitos que estes aplicativos devem atender ao utilizar dados para sustentar a automatização de ações que exigem tomadas de decisão, muitas vezes em tempo de execução (BRYNJOLFSSON; HITT; KIM, 2011; PROVOST; FAWCETT, 2013a, 2013b; TRELEAVEN; BROWNBRIDGE; HOPKINS, 1982).

A participação dos dados em tantos processos também interfere na nossa relação com a tecnologia. O **Data-ism**, por exemplo, está relacionado à existência de uma tendência natural a aceitar um resultado obtido pelo processamento de uma máquina como sendo mais confiável do que aquele feito pela ação humana. Se você vê uma conta realizada por uma calculadora, dificilmente vai questionar se o resultado é válido. Por ter sido projetado para tal finalidade e por já ter sido testado inúmeras vezes em processos similares, o senso comum assume o pressuposto de que a confiabilidade de um mecanismo tecnológico é inquestionável. Essa crescente aceitação nos resultados obtidos por processos eletrônicos pode gerar dependência e levar a situações em que o impacto de resultados gerados por processamento automatizado pode interferir em decisões sobre aspectos das mais variadas esferas tais como saúde, educação, jurídica, financeira, política e cultural. No entanto, é preciso considerar que estes resultados dependem em sua essência de processos anteriores relacionados à coleta, tratamento e visualização destes dados. Ou seja, os resultados obtidos dependem diretamente da seleção dos dados de entrada (denominados como *inputs*), bem como na forma que foi feito o tratamento, a análise e a disponibilização dos dados. Em um contexto em que há uma quantidade abundante de dados, a confiança sobre um fato pode estar baseada mais em dados e análises, e menos na intuição e na experiência, o que pode transformar também a natureza da liderança e da gestão, o comportamento do consumidor, o diagnóstico médico, entre outros (LOHR, 2015, 2016).

O fenômeno Big Data, associado à disponibilidade de conexão com a Internet e a adoção de TIC nas atividades de entretenimento e profissionais, influenciaram a forma que nos informamos. Para áreas da Comunicação, a disponibilidade de conjuntos de dados traz uma nova fonte informacional, o que amplia a sua área de atuação e, ao mesmo tempo, exige novas competências para manipular os dados e as TIC. O jornalismo de dados, ou **Data Journalism**, é um termo que define esta nova abordagem na forma de produção, onde também está relacionado com a forma dos jornalistas de coletar, elaborar relatórios e publicar dados e interpretações sobre estes, a partir do uso de ferramentas de TIC (GRAY et al., 2012; LIMA JUNIOR, 2012; ZION; CRAIG, 2015).

Ao pensar nesta grande abundância de dados e nos resultados que podemos obter a partir deles, temos que considerar também a questão da definição de ‘quem’ pode acessá-los. Tornar os dados acessíveis ao maior número possível de pessoas e usos, não só agrega valor aos dados com também se configura como importante contribuição para a sociedade. Dados preparados de tal forma que possam ser acessados e utilizados por qualquer indivíduo ou instituição são denominados como **Open Data**, ou Dados Abertos e podem beneficiar estados, instituições e indivíduos ao reduzir as assimetrias no acesso, propiciando assim uma maior participação e acompanhamento em questões sociais, econômicas e ambientais. É preciso salientar que os dados não podem ser disponibilizados em sua totalidade, devendo haver o cuidado com questões relacionadas com a privacidade, tais como a impessoalidade dos dados, ou seja, garantindo que não contenham informações suficientes para a identificação de indivíduos ou entidades quando se tratar de dados sensíveis (CHARALABIDIS et al., 2018; EUROPEAN COMMISSION, 2020a; MORIN, 2013; OPEN KNOWLEDGE FOUNDATION, 2015, 2020).

O mesmo conceito de dados abertos, quando aplicado a dados que estejam sob responsabilidade de agentes públicos é denominado como Dados Abertos Governamentais ou **Open Government Data**. Este conceito ganha relevância a partir do princípio de que grande parte dos dados gerados pela gestão da coisa pública deveria estar ao alcance de todos os cidadãos e, portanto, estruturados como dados abertos. Para que sejam considerados como dados abertos, é preciso levar em conta

critérios que garantam que estes dados possam ser acessados de tal forma que sejam processados automaticamente, sem a exigência de identificação do usuário que os procura, sem a necessidade de uso de aplicativos que também não sejam abertos, entre outros critérios, garantindo assim que estes dados realmente tenham seu acesso facilitado e otimizado (OPEN GOVERNMENT WORKING GROUP, 2008; RODRIGUES; SANT'ANA, 2017a, 2017b; THE WORLD BANK GROUP, 2014).

Quanto a facilitar a interpretação dos dados disponibilizados, principalmente na internet, é preciso lembrar que os dados em sua essência não contém carga semântica suficiente para que sejam interpretados autonomamente. Muitas são as propostas de estruturar estes dados de tal forma que possam receber tais conteúdos semânticos e uma delas diz respeito a formas de se estabelecer regras de relacionamento entre os conteúdos e seus significados, tornando estes dados em dados conectados ou **Linked Data**. Para representar estes dados, foram adotados padrões como a estrutura para descrição de recursos (Resource Description Framework - RDF) e de linguagens de consulta e de recuperação dos conjuntos de dados (BERNERS-LEE, 2009; BIZER; HEATH; BERNERS-LEE, 2009; ISOTANI; BITTENCOURT, 2015).

Considerando mais uma vez que estes dados devem ter como uma de suas premissas a necessidade de estarem ao alcance do maior número de pessoas e necessidades possíveis, foram estabelecidos critérios para que estes dados ligados estivessem sob o formato aberto, podendo, assim, ser denominados como **Linked Open Data**, ou seja, conjuntos de dados em formato Linked Data que são publicados sobre algum tipo de licença de uso aberto, tais como o Creative Commons (na modalidade CC-BY) e a Open Data Commons Open Database License (ODbL) (BERNERS-LEE, 2009; BIZER, 2013; BIZER; HEATH; BERNERS-LEE, 2009, 2009; ISOTANI; BITTENCOURT, 2015; RODRIGUES; SANT'ANA, 2017a).

Nem todo acesso aos dados é desejado. Existem as muitas situações em que os dados precisam ser protegidos e o comprometimento da segurança de dados, configurado como uma brecha nesta segurança, é denominado como **Data Breach**. Estas brechas podem resultar em efeitos como a destruição acidental ou ilegal dos dados; a perda; a alteração; a exposição não

autorizada ou o acesso a algum dado privado; a interceptação da transmissão ou processamento de dados privados. Como efeito de uma brecha, pode se esperar o comprometimento da confidencialidade, da integridade ou da disponibilidade dos dados, geralmente envolvendo uma situação de quebra de privacidade, com o acesso a dados que identificam um fato, uma pessoa ou uma instituição, incluindo a possibilidade de utilização de tais informações para outros tipos de atividades ilegais. Incidentes relacionados com o conceito de Data Breach podem ser intencionais (por exemplo, quando há uma ação intencional para acessar os dados) ou acidentais (por exemplo, quando ocorre uma divulgação não intencional de dados confidenciais por funcionários ou colaboradores) (CHENG; LIU; YAO, 2017; INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2015; ROMANOSKY; HOFFMAN; ACQUISTI, 2014; SEN; BORLE, 2015; SOLOVE, 2009; SOLOVE; CITRON, 2017; STEVENS, 2012).

Quando se considera os dados armazenados e a interpretação de seus conteúdos, é preciso ter em mente que serão compostos por valores, descritivos ou quantitativos, e, ainda, conteúdos identificadores que permitem a contextualização destes valores. Dentre estes identificadores, podemos ter aqueles que possibilitam vincular os valores a instâncias internas ou externas aos dados, ou seja, vincular estes valores a outros conjuntos de dados ou, mesmo que de forma indireta, a indivíduos referenciados por estes dados, o que torna estes conjuntos de dados como dados pessoais ou **Personal Data**. Estes valores agora referenciados a um indivíduo podem ser formados por informações como suas características naturais (como as informações sobre o seu corpo) e artificiais (como os números de seus documentos), pessoais e profissionais. Para proteger estes dados, são aplicadas técnicas como as de anonimização (processo de remoção ou modificação de atributos que permitam a sua identificação) e de encriptação dos dados (BRASIL, 2018; EUROPEAN COMMISSION, 2020b; FORS et al., 2019; MADSEN, 1992; MAI, 2016; PECK, 2020).

Também é preciso considerar as preocupações jurídicas acerca do local de armazenamento dos conjuntos de dados pessoais, institucionais e governamentais. Com o uso intensivo de inúmeras plataformas e serviços de terceiros - tais como os

serviços de e-mail, de redes sociais online, de armazenamento de arquivos, de comércio eletrônico, entre outros - surge o problema sobre o local em que os conjuntos de dados estão armazenados, pois a legislação que sustenta as condições de coleta, de acesso, de processamento, de uso e de compartilhamento podem variar de país para país e muitos destes serviços são oferecidos por empresas que operam globalmente. O **Data Localization** trata sobre este tema, definindo regras para prevenir a transferência de dados entre diferentes localidades. As regras podem ter objetivos diferentes, dependendo do tipo de dado, variando desde ações de proibição de transferência de dados para outro país, até regras que exigem o consentimento prévio do titular dos dados antes da transferência ou que exigem que cópias dos dados sejam armazenadas no país em que foram geradas (BLOKDYK, 2020; CHANDER; LE, 2014; HON, 2017).

Outro conceito relacionado com a localização dos dados e que se complementa com o fato de que a terceirização da infraestrutura de armazenamento de dados tem se tornado um procedimento cada vez mais comum é a questão da Soberania dos Dados ou **Data Sovereignty**, que trata sobre a necessidade de inclusão de novas informações nos conjuntos de dados disponíveis para garantir que existam informações sobre a integridade, a autenticidade e a localização dos dados na nuvem. Nesse sentido, é importante o estabelecimento de elementos informacionais detalhados que permitam identificar a localização dos dados, garantindo ao proprietário dos dados que os provedores estão cumprindo as regras (FANG, 2018, 2018; PETERSON; GONDREE; BEVERLY, 2011).

Identificando o interesse nas publicações pelos principais termos

Já que estamos falando de dados, podemos utilizar um recurso bastante interessante para perceber o grau de interesse dos principais conceitos vistos neste texto, tomando como base sua utilização em livros. Este levantamento foi feito pelo recurso Books Ngram Viewer do Google, que pode ser utilizado de forma gratuita e apresenta gráficos com indicações de citações do termo informado em livros.

Quadro 1 - Informações do Books Ngram Viewer sobre os principais conceitos neste texto, ordenados pelo pico de interesse, de forma decrescente

Conceito	Ponto de maior interesse		Situação atual	
	Ano	Índice ¹	Índice em 2019 ¹	Tendência
Data Collection	1993	0,8540749	0,6695922	Queda
Big Data	2019	0,7374097	0,7374097	Alta
Data Analysis	2019	0,5167581	0,5167581	Estável
Data Mining	2008	0,3259558	0,2628871	Estável
Data Center	1994	0,2821684	0,0984718	Queda
Personal Data	2019	0,2577607	0,2577607	Alta
Data Analytics	2019	0,1605424	0,1605424	Alta
Data Warehouse	2002	0,1345805	0,0364595	Queda
Data-driven	2019	0,1311215	0,1311215	Alta
Data Science	2019	0,1290053	0,1290053	Alta
Open Data	2019	0,0974848	0,0974848	Alta
Data Breach	2019	0,0344231	0,0344231	Alta
Linked Data	2015	0,0343318	0,0226086	Queda
Data Engineering	2003	0,0291547	0,0148295	Queda
Small Data	2019	0,0160955	0,0160955	Alta
Data Mart	2002	0,0152036	0,0033919	Queda
Data Lake	2019	0,0129638	0,0129638	Alta
Open Government Data	2018	0,0100055	0,0100055	Alta
Data Journalism	2019	0,0074127	0,0074127	Alta
Linked Open Data	2019	0,0069044	0,0069044	Alta
Data Sovereignty	2019	0,0028027	0,0028027	Alta
Data Localization	2019	0,0023196	0,0023196	Alta
Dark Data	2019	0,0015533	0,0015533	Alta
Data-ism	2015	0,0001607	0,0001270	Estável

¹ Os índices foram multiplicados por 1000 para facilitar a visualização.

Fonte: autores a partir de Google Books Ngram Viewer (2020).

Por meio dados obtidos e apresentados no Quadro 1, podemos perceber os conceitos com maior interesse no universo de livros publicados, quando foi seu momento de maior interesse e se a tendência é de estabilidade, de queda ou de alta, permitindo estabelecer em que ponto de maturidade encontra-se o seu uso.

Quadro 2 - Informações do *Books Ngram Viewer* sobre conceitos tecnológicos do século XX

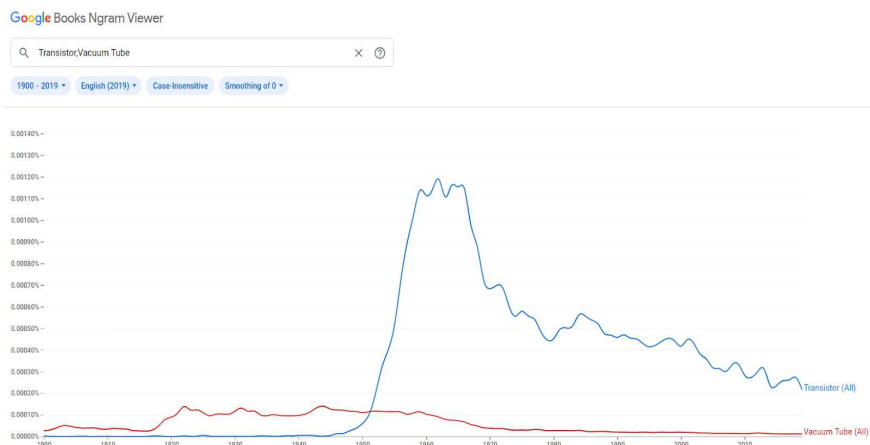
Conceito	Ponto de maior interesse		Situação atual	
	Ano	Índice ¹	Índice em 2019 ¹	Tendência
Vacuum Tube	1922	0,1540554	0,2160811	Queda
Transistor	1962	1,2372511	0,0111190	Queda

¹ Os índices foram multiplicados por 1000 para facilitar a visualização.

Fonte: Autores a partir de Google Books Ngram Viewer (2020).

Para efeito de comparação, podemos utilizar o mesmo processo para perceber a relação de interesse por tecnologias que até continuam a ser utilizadas mas que já não despertam o interesse ou não merecem o mesmo destaque que em momentos de maior impacto (Quadro 2). Assim, por meio de dados, podemos facilmente acompanhar os ciclos de maturidade relacionados a determinados conceitos como, por exemplo, Vacuum Tube e Transistor - em língua portuguesa, Tubo de Vácuo e Transistor, respectivamente. Ou seja, a possibilidade de uso destes dados nos dá a compreensão de quando determinado conceito está ou não sendo foco de interesse dos autores.

Figura 1 - Comparativo entre os conceitos ‘Transistor’ e ‘Vacuum Tube’



Fonte: Google Books Ngram Viewer (2020).

Inclusive, este uso dos dados sobre os conceitos nos livros publicados nos permite visualizar a evolução dos conceitos ao longo do tempo na forma gráfica, como o apresentado na Figura 1, em que podemos analisar o comportamento dos conceitos e o interesse destes ao longo do tempo. Quanto mais alta a linha maior o número de ocorrências do conceito nos livros. Já o eixo horizontal representa a linha do tempo (no exemplo entre os anos de 1900 e 2019).

Considerações finais

Após este breve passeio sobre os conceitos ‘DATA-alguma-coisa’, como citado no início deste capítulo, espera-se que esta jornada ajude na redução de dúvidas na interpretação de tais conceitos, empoderando o leitor quando este se deparar com o emprego de tais conceitos em textos que vão desde a divulgação de produtos de TIC, passando por textos acadêmicos e até mesmos em discursos políticos, quando buscam se aproximar deste novo cenário tecnológico.

Recomendamos aos leitores, que se interessarem pelos temas, que busquem leituras para se aprofundar nas referidas definições sobre o universo de dados. As que apresentamos aqui representam apenas uma pequena parte das opções que podem ser encontradas em livros e até mesmo na internet.

Referências

ANNE LAURENT. **Data Lakes**. 1. ed. Hoboken, EUA: Wiley Publishing, Inc., 2020.

BARBIERI, C. **BI2: Business Intelligence**. 1. ed. Rio de Janeiro: Elsevier, 2011.

BERMAN, J. J. **Principles of big data: preparing, sharing, and analyzing complex information**. 1. ed. Waltham, EUA: Elsevier, Morgan Kaufmann, 2013.

BERNERS-LEE, T. **Linked Data**. Disponível em: <http://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 1 jun. 2015.

BIZER, C. Interlinking Scientific Data on a Global Scale. **Data Science Journal**, v. 12, p. GRDI6–GRDI12, 2013.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far: **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1–22, mar. 2009.

BLOKDYK, G. **Data Localization A Complete Guide - 2020 Edition**. [s.l.] 5STARCook, 2020.

BRASIL. 13709. **Lei número 13709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). 2018.

BRYNJOLFSSON, E.; HITT, L. M.; KIM, H. H. **Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?** Rochester, Nova Iorque, EUA: Social Science Research Network, 22 abr. 2011.

CHAN, Y.; TALBURT, J.; TALLEY, T. M. (EDS.). **Data engineering: mining, information and intelligence**. Nova Iorque, EUA: Springer, 2010.

CHANDER, A.; LE, U. P. **Breaking the Web: Data Localization vs. the Global Internet**. Rochester, Nova Iorque, EUA: Social Science Research Network, 1 abr. 2014.

CHARALABIDIS, Y. et al. **The world of open data: concepts, methods, tools and experiences**. 1. ed. Nova Iorque, EUA: Springer Science+Business Media, 2018.

CHENG, L.; LIU, F.; YAO, D. (DAPHNE). Enterprise data breach: causes, challenges, prevention, and future directions. **WIREs Data Mining and Knowledge Discovery**, v. 7, n. 5, p. 1–14, 2017.

CISCO INC. **What Is a Data Center?** 2020. Disponível em: <https://www.cisco.com/c/en/us/solutions/data-center-virtualization/what-is-a-data-center.html>. Acesso em: 11 ago. 2020.

COX, M.; ELLSWORTH, D. **Application-controlled demand paging for out-of-core visualization**. Proceedings of the 8th conference on Visualization '97. **Anais...**: VIS '97. Washington, DC, USA: IEEE Computer Society Press, 1 out. 1997.

CRISTESCU, R. et al. Network Correlated Data Gathering with Explicit Communication: NP-completeness and Algorithms. **IEEE/ACM Trans. Netw.**, v. 14, n. 1, p. 41–54, fev. 2006.

DAMIEN, L. **DATA ANALYTICS: A Comprehensive Beginner's Guide to Learn the Realms of Data Analytics**. [s.l.] Publicação independente, 2019.

DOBBS, R. et al. **Big Data**. McKinsey Global Institute, 2011.

DUMBILL, E. **Planning for Big Data**. 1. ed. Sebastopol, EUA: O'Reilly Media, Inc., 2012.

EUROPEAN COMMISSION. **What is open data?** 2020a. Disponível em: <https://www.europeandataportal.eu/elearning/en/module1/#/id/co-01>. Acesso em: 11 ago. 2020a.

EUROPEAN COMMISSION. **What is personal data?** 2020b. Disponível em: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en. Acesso em: 11 ago. 2020.

FANG, B. **Cyberspace sovereignty**. 1. ed. Nova Iorque, EUA: Springer Berlin Heidelberg, 2018.

FORS, V. et al. **Imagining Personal Data: Experiences of Self-Tracking**. 1 ed. [s.l.] Bloomsbury Academic, 2019.

GOG, T. V. et al. Data Collection and Analysis. In: SPECTOR, J. M. et al. (Eds.). **Data Collection and Analysis**. 3. ed. Nova Iorque: Lawrence Erlbaum Associates, 2007. p. 763–806.

GOOGLE BOOKS NGRAM VIEWER. **Google Books Ngram Viewer**. Alphabet, Inc., 2020. Disponível em: <https://books.google.com/ngrams>. Acesso em: 1 ago. 2020

GORELIK, A. **The Enterprise Big Data Lake**. Sebastopol, EUA: O'Reilly Media, Inc., 2019.

GRAY, J. et al. (EDS.). **The data journalism handbook**. 1 ed. Sebastopol, EUA: O'Reilly Media, 2012.

GREENBERG, A. et al. **Towards a next generation data center architecture: scalability and commoditization**. Proceedings of the ACM workshop on Programmable routers for extensible services of tomorrow - PRESTO '08. **Anais...** In: THE ACM WORKSHOP. Seattle, WA, USA: ACM Press, 2008.

GRUS, J. **Data science from scratch: first principles with Python**. 1 ed. Sebastopol, EUA: O'Reilly, 2015.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 3 ed. Burlington, MA, EUA: Elsevier, 2012.

HEIDORN, P. B. Shedding Light on the Dark Data in the Long Tail of Science. **Library Trends**, v. 57, n. 2, p. 280–299, 2008.

HILL, T.; LEWICKI, P. **Statistics: methods and applications: a comprehensive reference for science, industry, and data mining**. Tulsa, EUA: StatSoft, 2006.

HON, W. K. **Data localization laws and policy: the EU data protection international transfers restriction through a cloud computing lens**. Cheltenham, Reino Unido: Edward Elgar Publishing, 2017.

HOWE, D. et al. Big data: The future of biocuration. **Nature**, v. 455, n. 7209, p. 47–50, 2008.

INMON, W. H. **Building the data warehouse**. 4. ed. Indianapolis, EUA: Wiley, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION.
ISO/IEC 27040:2015.2015.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. 1. ed. São Paulo: Novatec, 2015.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit The Complete Guide to Dimensional Modeling**. Nova Iorque, Estados Unidos da América: John Wiley & Sons, 2011.

KOLB, J.; KOLB, J. **The big data revolution: the world is changing. Are you ready?** Charlseton, EUA: CreateSpace, 2013.

LIMA JUNIOR, W. T. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. **Estudos em Comunicação**, v. 12, p. 207–222, dez. 2012.

LINDSTROM, M.; HEATH, C. **Small Data: The Tiny Clues That Uncover Huge Trends**. [s.l.] St. Martin's Press, 2016.

LLAVE, M. R. Data lakes in business intelligence: reporting from the trenches. **Procedia Computer Science**, CENTERIS 2018 - International Conference on ENTERprise Information Systems / ProjMAN 2018 - International Conference on Project MANagement / HCist 2018 - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018. v. 138, p. 516–524, 1 jan. 2018.

LOH, S. **BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões**. 1. ed. Porto Alegre, Brasil: Stanley Loh, 2014.

LOHR, S. **Data-ism: the revolution transforming decision making, consumer behavior, and almost everything else.** First edition ed. Nova Iorque, EUA: HarperCollins Publishers, 2015.

LOHR, S. **Data-ism: inside the big data revolution.** 2. ed. Londres, Inglaterra: Oneworld, 2016.

LYCETT, M. 'Datafication': making sense of (big) data in a complex world. **European Journal of Information Systems**, v. 22, n. 4, p. 381–386, 1 jul. 2013.

MADSEN, W. **Handbook of personal data protection.** Nova Iorque, EUA: Macmillan Publishers Ltd., 1992.

MAHESHWARI, A. **Data Analytics Made Accessible: 2020 edition.** [s.l.] 1, 2014.

MAI, J.-E. Big data privacy: The datafication of personal information. **The Information Society**, v. 32, n. 3, p. 192–199, 26 maio 2016.

MARIBEL, Y. S.; RAMOS, I. **Business Intelligence: Tecnologias da Informação na Gestão de Conhecimento.** 2. ed. [s.l.] FCA, 2009.

MORIN, P. **Open Data Structures: An Introduction.** 31st ed. edition ed. [s.l.] Athabasca University Press, 2013.

OPEN GOVERNMENT WORKING GROUP. **Principles of open Government data.** Disponível em: https://public.resource.org/8_principles.html. Acesso em: 10 ago. 2020.

OPEN KNOWLEDGE FOUNDATION. **Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge.** Disponível em: <https://opendefinition.org/od/2.1/en/>. Acesso em: 11 ago. 2020.

OPEN KNOWLEDGE FOUNDATION. **What is Open Data?** Disponível em: <https://opendatahandbook.org/guide/en/what-is-open-data/>. Acesso em: 11 ago. 2020.

O'REILLY. **Big Data Now.** 2. ed. EUA: O'Reilly Media, Inc., 2012.

PECK, P., Patricia. **Proteção de Dados Pessoais: Comentários à Lei n. 13.709/2018 -LGPD.** São Paulo, Brasil: Saraiva, 2020.

PETERSON, Z. N. J.; GONDREE, M.; BEVERLY, R. **A position paper on data sovereignty: the importance of geolocating data in the cloud.** Proceedings of the 3rd USENIX conference on Hot topics in cloud computing. **Anais...** HotCloud'11. EUA: USENIX Association, 14 jun. 2011. Acesso em: 11 ago. 2020

PLATTS, J. **What is Dark Data?** Disponível em: <https://medium.com/@jillplatts/what-is-dark-data-1beef317bb2e>. Acesso em: 11 ago. 2020.

PROVOST, F.; FAWCETT, T. **Data science for business: what you need to know about data mining and data-analytic thinking**. 1. ed., 2. release ed. Beijing: O'Reilly, 2013a.

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, 13 fev. 2013b.

RASMUSSEN, R. et al. **Data Analyst - BCS Guides to IT Roles**. 1. ed. Swindon, Reino Unido: BCS Learning & Development Limited, 2019.

RODRIGUES, F. A. **Coleta de dados em redes sociais: privacidade de dados pessoais no acesso via Application Programming Interface**. Dissertation—Marília, Brasil: Universidade Estadual Paulista, 3 mar. 2017.

RODRIGUES, F. A.; SANT'ANA, R. C. G. Uso de Modelos de Dados Multidimensionais para a ampliação da Transparência Ativa. **Liinc em Revista**, v. 9, n. 2, p. 469–487, 29 nov. 2013.

RODRIGUES, F. A.; SANT'ANA, R. C. G. Use of Taxonomy of Privacy to Identify Activities Found in Social Network's Terms of Use. **Knowledge Organization**, v. 43, n. 4, p. 285–295, 2016.

RODRIGUES, F. A.; SANT'ANA, R. C. G. A study on actions to make government datasets available in linked open data. In: **Knowledge Organization and Cultural Diversity**. Recife: ISKO e Editora UFPE, 2017a. v. 1p. 522–532.

RODRIGUES, F. A.; SANT'ANA, R. C. G. Elaboração de estratégia para mensuração de critérios de qualidade na recuperação de datasets disponíveis em sítios governamentais. In: **Comunicação e Transformações Sociais**. Ciência da Informação, Comunicação e Educação. 1. ed. Coimbra: Associação Portuguesa de Ciências da Comunicação, 2017b. v. 1p. 155–169.

RODRIGUES, F. A.; SANT'ANA, R. C. G. Contextualização de conceitos teóricos no processo de coleta de dados de Redes Sociais Online. **Informação & Tecnologia**, v. 5, n. 1, p. 18–36, 2018.

ROMANOSKY, S.; HOFFMAN, D.; ACQUISTI, A. Empirical Analysis of Data Breach Litigation. **Journal of Empirical Legal Studies**, v. 11, n. 1, p. 74–104, 2014.

SACOLICK, I. **Dark Data - A Business Definition**, 2013. Disponível em: <<https://blogs.staricio.com/2013/04/dark-data-business-definition.html>>. Acesso em: 11 ago. 2020

SANTOS, P. L. V. A. DA C.; SANT'ANA, 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, p. 11, maio 2015.

SCHEMBERA, B.; DURÁN, J. M. Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer. **Philosophy & Technology**, v. 33, n. 1, p. 93–115, 1 mar. 2020.

SCHUTT, R.; O'NEIL, C. **Doing data science**. First edition ed. Sebastopol, EUA: O'Reilly Media, 2013.

SEN, R.; BORLE, S. Estimating the Contextual Risk of Data Breach: An Empirical Approach. **Journal of Management Information Systems**, v. 32, n. 2, p. 314–341, 3 abr. 2015.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business intelligence, analytics, and data science: a managerial perspective**. Fourth edition ed. Nova Iorque, EUA: Pearson, 2018.

SHIVE, B. **Data engineering: a novel approach to data design**. 1. ed. Basking Ridge, EUA: Technics Publications, LLC, 2013.

SOLOVE, D. J. **Understanding privacy**. First Harvard University Press paperback edition ed. Cambridge, Massachusetts Londres, Inglaterra: Harvard University Press, 2009.

SOLOVE, D. J.; CITRON, D. K. Risk and Anxiety: A Theory of Data-Breach Harms. **Texas Law Review**, v. 96, p. 737, 2018 2017.

STEVENS, G. **Data Security Breach Notification Laws**: Congressional Research Service. Washington, EUA: National Conference of State Legislatures, 10 abr. 2012.

THE WORLD BANK GROUP. **Open Government Data Toolkit**. Disponível em: <http://opendatatoolkit.worldbank.org/en/>. Acesso em: 1 ago. 2020.

TRELEAVEN, P. C.; BROWNBRIDGE, D. R.; HOPKINS, R. P. Data-Driven and Demand-Driven Computer Architecture. **ACM Computing Surveys**, v. 14, n. 1, p. 93–143, 1 mar. 1982.

WICKRAMASURIYA, J. et al. **Privacy Protecting Data Collection in Media Spaces**. Proceedings of the 12th Annual ACM International Conference on Multimedia. **Anais...: MULTIMEDIA '04**. In: 12TH ANNUAL ACM INTERNATIONAL CONFERENCE ON MULTIMEDIA. Nova Iorque, NY, USA: ACM, 2004.

WOLKENHAUER, O. **Data engineering: fuzzy mathematics in systems theory and data analysis**. Nova Iorque, EUA: Wiley, 2001.

ZION, L.; CRAIG, D. (EDS.). **Ethics for digital journalists: emerging best practices**. Nova Iorque, EUA: Routledge, Taylor & Francis Group, 2015.