Garantindo imparcialidade, precisão, confidencialidade e transparência aos dados na perspectiva da Ciência dos Dados Responsáveis

Morgana Andrade[1], Paula Regina Gonçalez[2], Decio Wey Berti Junior[4], Ana Alice Baptista[5], Caio Saraiva Coneglian[3]

[1] https://orcid.org/0000-0002-4055-5882
Universidade Federal do Espírito Santo (UFES), Vitória, Espírito Santo
morganaandrade@hotmail.com

[2] https://orcid.org/0000-0002-5480-4106
Universidade Federal do Espírito Santo (UFES), Vitória, Espírito Santo.
paulaventuramorim@gmail.com

[3] http://orcid.org/0000-0003-4313-2727
Universidade Estadual de Londrina (UEL), Londrina, Paraná, Brasil.
deciowbj@gmail.com

[4] http://orcid.org/0000-0003-3525-0619
Centro Algoritmi – Universidade do Minho, Guimarães, Portugal
analice@dsi.uminho.pt

[5] https://orcid.org/0000-0002-6126-9113
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Marília, Brasil
caio.coneglian@gmail.com

Resumo

A demanda por acesso e análise de dados, sejam privados, sejam públicos, que promove a tomada de decisão comercial e a ciência, tem impulsionado a Economia e os avanços científicos, de forma a atrair a atenção de vários segmentos da sociedade. No contexto Big Data, surge, como necessidade urgente, a aplicação de direitos individuais e empresariais e de normas regulatórias que resguardem a privacidade, a imparcialidade, a precisão e a transparência. Este artigo aborda possíveis alternativas que podem assegurar a aplicação da ética, bem como a regulamentação para acesso e reúso de dados, a Responsible Data Science (Ciência de Dados Responsáveis). Nesse cenário, a Responsible Data Science desponta como uma iniciativa que tem como base as diretrizes FACT, que correspondem à adoção de quatro princípios: imparcialidade, precisão, confidencialidade e transparência. Para a implementação dessas diretrizes, deve-se considerar o uso de técnicas e abordagens que estão sendo desenvolvidas pela Green Data Science. Para tanto, foi desenvolvida investigação exploratória e descritiva com abordagem qualitativa. Quanto à literatura sobre os temas, foram realizadas pesquisas nas bases de dados bibliográficas Web of Science, Scopus e pelo motor de busca Scholar Google. Foram adotados os termos "Responsible Data Science", "Fairness, Accuracy, Confidentiality, Transparency + Data Science", FACT e FAT relacionados com Data Science. Após a análise dos documentos selecionados, concluiu-se que a Green Data Science e as diretrizes FACT contribuem significativamente para a salvaguarda dos direitos individuais, não sendo necessário recorrer a medidas que impeçam o

acesso e a reutilização de dados. A iniciativa Responsible Data Science é vista pelos autores como boa prática no contexto do Big Data, principalmente pela possibilidade de reúso dos dados. Os desafios para implementar as diretrizes FACT requerem estudos, condições *sine qua non* para que as ferramentas para análise e disseminação dos dados sejam desenvolvidas ainda na fase de design, ou seja, na concepção de metodologias que garantam a imparcialidade, a precisão, a transparência e a confidencialidade. Caso contrário, a sua não efetivação poderá resultar em consequências indesejáveis para a sociedade.

Palavras-chave: Ciência de dados. Ciência Responsável de Dados. Big Data. Ética. FACT.

Abstract

The demand for access and analysis of data be it private or public, and that promotes business decision-making and science, has driven the economy and scientific advances, attracting the attention of various segments of society. In the Big Data context, there is an urgent need to apply individual and corporate rights and regulatory standards that safeguard privacy, impartiality, precision and transparency. This article discusses possible alternatives and applications of ethics, as well as the regulation for access and reuse of data (Responsible Data Science). In this scenario, Responsible Data Science emerges as an initiative based on the FACT guidelines, which correspond to the adoption of four principles: impartiality, precision, confidentiality and transparency. For the implementation of these guidelines, the use of techniques and approaches being developed by Green Data Science should be considered. With this in mind, an exploratory and descriptive research with a qualitative approach was carried out. As for the literature on this topic, we searched the bibliographic databases Web of Science, Scopus and the Scholar Google search engine. The terms "Responsible Data Science", "Fairness, accuracy, confidentiality, transparency + Data Science", FACT and FAT related to Data Science were used. After reviewing the selected documents, we concluded that Green Data Science and the FACT guidelines contribute significantly to safeguarding individual rights and there is no need to resort to measures that prevent access to and reuse of data. The Responsible Data Science initiative is seen by the authors as a good practice in the context of Big Data, mainly because of the possibility of data reuse. The challenges to implementing the FACT guidelines require studies and *sine qua non* conditions for data analysis and dissemination tools to be developed at the design stage, i.e. designing methodologies that ensure impartiality, accuracy, transparency and confidentiality. Not doing so may result in undesirable consequences for society.

Keywords: Data Science. Responsible Data Science. Big Data. Ethic. FACT

1 Introdução

O aumento do volume de dados tem modificado a maneira como pesquisas, governança, socialização e negócios estão sendo realizados (Hilbert & López, 2011, Kemper & Kolkman, 2018). Esses dados são gerados por diferentes segmentos da sociedade, governo, indústria, institutos de pesquisa e universidades. A forma como são coletados, armazenados e

divulgados também é realizada de diferentes maneiras e por diferentes organismos (European Data Science Academy, 2019, Stoyanovich & Howe, 2018, Taylor, 2017).

Exemplos recentes de coletas e utilização de dados têm sido amplamente noticiados: o uso de dados referentes ao registro digital e bibliométrico pela Índia e pela China com o intuito de rastrear e monitorar os cidadãos (Taylor, 2017); a coleta de dados pessoais do Facebook, utilizados para fins políticos, sem consentimento dos usuários, pela Cambridge Analytica (Facebook, 2019); a utilização dos dados dos usuários do Google Maps e Youtube pelo Google para enviar anúncios personalizados (Satariano, 2019); e o compartilhamento de dados por meio de Application Programming Interface (API) (Piersma, 2018). Ações como essas estão relacionadas com a preocupação quanto ao uso e à publicação de dados no que concerne ao acesso e à divulgação de informações privadas ou públicas e à privacidade ou conclusões injustas e/ou tendenciosas por parte de diferentes atores (AIMS, 2017, Ohm, 2014, Taylor, 2017). São fatos associados às novas "Webs", à "Internet das Coisas (IoT)", à "Web de Dados", ao Big Data.

A demanda por acesso e análise de dados privados e públicos tem aumentado de forma a chamar a atenção de vários segmentos da sociedade. O panorama atual mostra que empresas obtêm lucro, governos influenciam a tomada de decisão em relação aos seus cidadãos e pessoas são induzidas a conceitos cujo agravante é que nem sempre os dados que subsidiam essas ações são imbuídos dos princípios da ética e da justiça (Kemper & Kolkman, 2018, Stoyanovich & Howe, 2018, Taylor, 2017)

Em meio a essa problemática, surge, como forte demanda, a aplicação de direitos individuais e empresariais e de normas regulatórias que resguardem a privacidade do indivíduo e a transparência das informações (Moerel & Prins, 2016). Teve início, nos anos de 2016-2017, um movimento em busca da imparcialidade, responsabilidade e transparência (FAT) na tomada de decisão algorítmica e na ciência dos dados de forma mais ampla (Stoyanovich, Howe, & Jagadish, 2018). Nesse sentido, surge o projeto Responsible Data Science (RDS), Ciência Responsável dos Dados, apresentado como possibilidade de equacionar questões em relação à imparcialidade, precisão, confiabilidade e transparência, no que se refere ao Big Data e à Ciência dos Dados. Esses quatro princípios são representados pelo acrônimo FACT (Fairness, Accuracy, Confidentiality, Transparency) (van der Aalst, 2016).

O RDS busca desenvolver métodos e técnicas para apoiar a publicação e o uso de dados de acordo com o FACT, com o objetivo de propiciar tecnologia para garantir esses princípios ainda na etapa de desenvolvimento e criar plataformas multidisciplinares. A adoção do FACT, ainda na fase de concepção das tecnologias e métodos aplicados ao Big Data e à Ciência dos Dados, apresenta-se como alternativa para garantir a qualidade e segurança na utilização de dados (Stoyanovich & Howe, 2018, van der Aalst, 2016).

Diante dos desafios de incorporar valores éticos de forma justa e sustentável aos direitos individuais conforme as inovações tecnológicas que envolvem a Ciência dos Dados, o presente estudo busca identificar orientações e iniciativas que viabilizem a implementação das diretrizes FACT por aqueles que trabalham nesse contexto.

Este artigo está estruturado em quatro seções: Introdução; Seção 1, onde é apresentado o tema; Seção 2, que contém a abordagem teórica em torno da Responsible Data Science e das diretrizes FACT. Os procedimentos metodológicos são descritos na Seção 3; os resultados obtidos são elencados na Seção 4; e, ao final, são apresentadas as Considerações Finais, Referências e Glossário.

2 Referencial teórico

Para Mayo (1996) e Chalmers (2013), citados por van der Aalst (2016), no contexto da Web há uma mudança na forma como o conhecimento é gerado. Quando o conhecimento é baseado em dados, "[...] segue a lógica do novo experimentalismo, em que o conhecimento é derivado de observações experimentais, não da teoria". Nesse novo paradigma, surgem preocupações em relação aos dados, principalmente como eles podem ser utilizados de forma irresponsável.

Um grupo de pesquisadores de diferentes instituições na Holanda deu início a um novo preceito: RDS que, segundo van der Aalst (2016), visa a produzir soluções positivas ao invés de evitar o uso dos dados. O consórcio RDS foi lançado em 2016 com o objetivo de "[...] combater desafios éticos e legais, e desenvolver técnicas de ciência de dados, infraestruturas e abordagens responsáveis por processos justos, precisos e dados confidenciais, transparentes por concepção" (AIMS, 2017, Moerel & Prins, 2016).

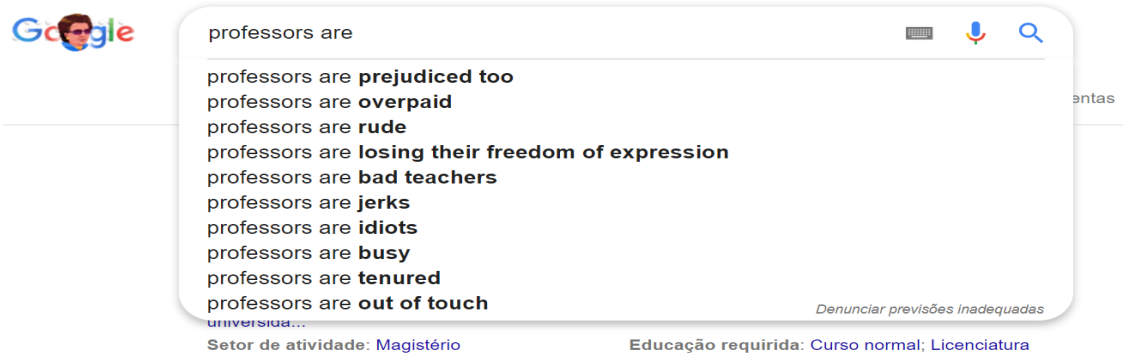
Essa iniciativa envolve disciplinas, como processo de mineração de dados, humanidades digitais, ética, recuperação da informação, representação do conhecimento, direito, *machine learning*, processamento de linguagem natural, segurança, estatística e visualização (van der Aalst, 2016, van der Aalst, Bichler, & Heinzl, 2017) e pode ser aplicada nos domínios de mídia, segurança e saúde.

Para a Academia Europeia de Ciência dos Dados (http://edsa-project.eu/overview/aboutedsa/), "A 'Era dos Dados' produz uma diversidade de dados de forma exponencial cujos desafios estão relacionados à recolha, armazenamento e análise desses dados". Nesse ambiente, têm sido desenvolvidas técnicas poderosas, como mineração de dados, aprendizado de máquina, banco de dados e visualização, que objetivam melhorar a vida de pessoas com a oferta de novos serviços e produtos com maior qualidade e eficiência (GE 301 Group 7, 2017, Stoyanovich et al., 2017, Stoyanovich, Howe, & Jagadish, 2018). Na visão de Piersman (2018), a RDS "[...] é um dos fundamentos da transformação digital".

No entanto, decisões automáticas de dados podem ser injustas e não transparentes; dados confidenciais podem ser compartilhados involuntariamente ou utilizados de forma indevida por terceiro; imprecisões podem ser geradas a partir de incorporação de vieses por parte de algoritmos, entre outros problemas (Ohm, 2014, Stoyanovich & Howe, 2018). Esses fatos provocam grande preocupação em relação à produção e uso dos dados, nomeadamente no que concerne a informações falsas, conclusões preconceituosas, divulgação de informações privadas e não transparentes, o que pode prejudicar a aplicação de dados científicos (AIMS, 2017), por exemplo, ou ainda comprometer a segurança das pessoas e a vida pessoal e profissional do indivíduo (GE 301 Group 7, 2017, Ohm, 2014, Taylor, 2017).

Exemplos sobre algumas dessas ocorrências são identificados no nosso dia a dia da forma mais simples a mais complexa e podem provocar sérias consequências tanto para os prestadores de serviços como para os proprietários dos dados (Facebook, 2019, Satariano, 2019). Um exemplo de como facilmente se pode chegar a uma conclusão preconceituosa pode ser observado a partir de uma simples busca realizada no Google para definição do termo professores (Figura 1).

Figura 1- Pesquisa no Google



Fonte: Google

Richard & King (2014), referenciados por Lodder (2016), professor da Vrije Universiteit, Amsterdam, considera que a construção dessa nova sociedade digital, a qual ele denomina de Sociedade Big Data, e o equilíbrio de valores humanos, como privacidade, confidencialidade, transparência, identidade, livre arbítrio, associados às novas estruturas digitais, serão impactantes na construção dos indivíduos. Nesse sentido, alerta sobre a relevância de se preservar esses valores em benefício da inovação e da conveniência.

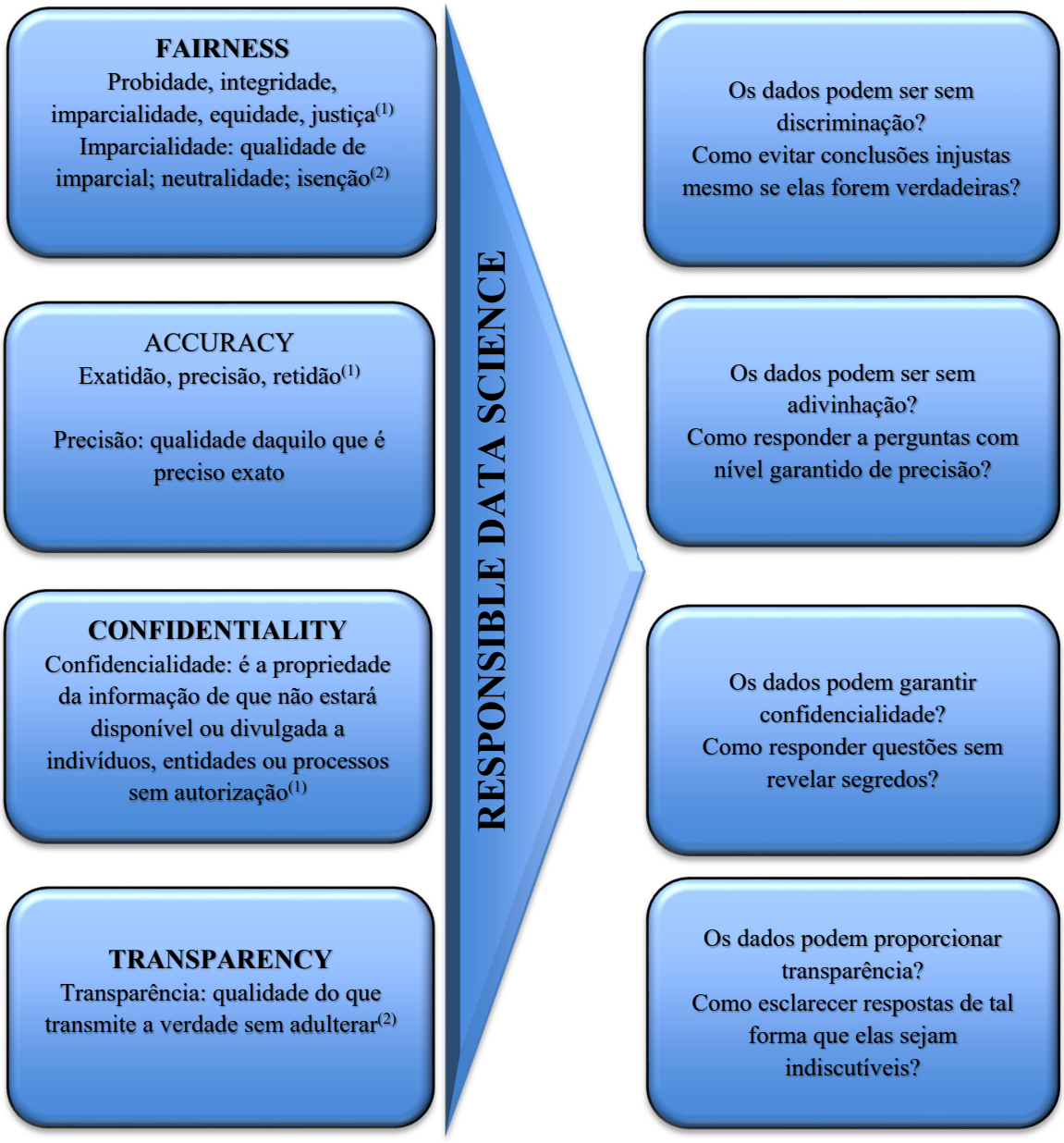
Esses valores são reiterados pela a iniciativa RDS que, como esboçado, tem por objetivo o desenvolvimento de um modelo sustentável para a ciência de dados com base nos valores éticos e aplicação de técnicas em conformidade com os requisitos de imparcialidade, responsabilidade, transparência e confidencialidade por concepção (Data Science Center Eindhoven, 2018).

Para o Data Science Center Eindhoven Research Program,

> [...] os valores da sociedade são uma parte intrínseca do valor do Big Data. Construir valores na ciência de dados 'por design' é um campo de pesquisa interdisciplinar desafiador e fascinante, com muitas aplicações práticas. Criar valor social através da ciência de dados requer uma compreensão do contexto e uma governança eficaz dos dados! (Data Science Center Eindhoven, 2018).

Assim, a Research Data Science emerge baseada em quatro pilares - imparcialidade, precisão, confidencialidade e transparência -, com o objetivo de atender às questões que afetam a sociedade em diferentes aspectos, éticos, econômicos e/ou sociais (Figura 2).

Figura 2 - Research Data Science



Nota: (1) Michaelis.uol.com.br; (2) Wikipédia; (3) Infopedia.pt
Fonte: Adaptado de van der Aalst (2016)

Esses princípios podem ser assim definidos:

IMPARCIALIDADE: "[...] vocábulo formado pelo prefixo 'in', privativo, + 'parcial'. É a atitude de quem considera, objetivamente, sem paixões ou preconceitos, um determinado fato na totalidade de seus aspectos. Quem é imparcial não sacrifica a verdade e a justiça à própria

conveniência ou à conveniência de outros, para tirar proveito pessoal, nem faz pesar no julgamento fatos anteriores ou informações que possam prejudicá-lo ou favorecê-lo individualmente" (Ávila, 1967).

A título de imparcialidade, justiça dos dados, Taylor (2017, p. 1) afirma que a disponibilidade de dados digitais, principalmente a partir do uso de dispositivos e serviços tecnológicos pelas pessoas, "[...] tem implicações políticas e práticas na maneira como as pessoas são tratadas pela comunidade, Estado e pelo setor privado". No entanto, o nível de conscientização e a implementação de mecanismos que combatam possíveis discriminações é menor, se comparados com a rapidez com que são desenvolvidas as tecnologias de processamento de dados.

PRECISÃO: a precisão das informações ou medições é a sua qualidade de ser verdadeira ou correta, mesmo em pequenos detalhes (Collins Dictionary, 2019).

A produção de resultados imprecisos independe da quantidade de dados analisados. Variáveis são usadas para prever um resultado. Para isso deve haver correlação positiva ou negativa entre variáveis e resultado com o objetivo de obter melhor precisão (GE 301 Group 7, 2017).

CONFIDENCIALIDADE: a confidencialidade é considerada como o dever de resguardar todas as informações que dizem respeito a uma pessoa, isto é, a sua privacidade. A confidencialidade inclui a preservação das informações privadas e íntimas (Goldim, 1997).

TRANSPARÊNCIA: a transparência pode ser entendida como a "'[...] compreensibilidade de um modelo específico'" (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018) e é vista como um requisito para a responsabilização algorítmica. Ainda pode significar "[…] que todos os códigos e todos os dados devem estar públicos" (Stoyanovich et al., 2018, p. 2165).

Annany e Crawford (2016)

> [...] sugerem que a transparência não pode ser uma característica de um modelo algorítmico. Em vez disso, a opacidade dos algoritmos deve ser considerada com sensibilidade para os contextos de seu uso; a transparência é realizada por conjuntos sociotécnicos de algoritmos e pessoas.

No entanto, a transparência nem sempre é total, seja em razão de garantias associadas a direitos comerciais, seja em razão da privacidade dos cidadãos que, ao não serem respeitadas, violam as leis (Stoyanovich et al., 2018).

Para Kemper e Kolkman (2018),

> A ciência de dados só pode ser eficaz se as pessoas confiarem nos resultados e puderem inferir e interpretar corretamente os resultados. A ciência de dados não deve, portanto, ser vista como uma caixa preta que transforma magicamente dados em valor. Muitas escolhas de projeto precisam ser feitas em um típico 'data science pipeline' […].

Os princípios aqui citados são defendidos pela RDS e estão associados à legalidade, ao normativo e ao ético que, ao juntar-se aos princípios FAIR (Findability, Accessibility, Interoperability and Reusability), surgem como os principais temas de pesquisa em Big Data.

Nota-se que as diretrizes FACT vêm complementar o FAIR, principalmente por focar aspectos essenciais para a reutilização dos dados.

A discussão sobre FAIR tem envolvido profissionais e instituições de diferentes domínios e encontra-se em um estágio mais avançado, portanto com maior número de publicações. No que tange às diretrizes FACT, por terem sido apresentadas recentemente ou por ainda os atores que atuam na área não compreenderem a dimensão desses elementos, não atingiram o mesmo nível de discussão, embora suas implicações envolvam a sociedade de forma mais presente e impactante. Os defensores das diretrizes FACT enfatizam que elas devem ser aplicadas não apenas no uso de modelos algorítmicos, mas também nas fases de projeto e desenvolvimento.

Esses princípios vêm norteando ações, principalmente, dentro da Comunidade Europeia; a exemplo da recente penalidade infligida à empresa Google por "[...] falta de transparência, informação inadequada e falta de consentimento válido em relação à personalização de anúncios" (Euronews, 2019).

A seguir, serão apresentados os procedimentos metodológicos aplicados a este estudo, cujo objetivo é identificar contribuições e iniciativas baseadas nas diretrizes FACT.

3 Procedimentos metodológicos

Este trabalho refere-se a uma investigação descritiva em que a pesquisa bibliográfica foi realizada a partir das bases de dados Scopus e Web of Science e com o motor de busca Google Scholar. A pesquisa documental foi desenvolvida em sites de organizações que abordam a presente temática: Responsible Data Science, Data Science Center Eindhoven e GE 301-Group 7. As buscas foram realizadas no período de 15 de dezembro de 2018 a 15 de janeiro de 2019, utilizando os termos “Responsible Data Science”, “Fairness, accuracy, confidentiality, transparency + Data science”, FACT e FAT relacionados com Data Science.

4 Resultados

A pesquisa bibliográfica proporcionou a identificação de 54 documentos que abordam o termo Responsible Data Science. Ao usar os termos FACT, FAT e “*Fairness*, *accuracy, confidentiality, transparency*”, associados à “Data Science”/”Big Data”, foram identificados 20 documentos. Consulta a sites de organizações/instituições proporcionaram a recuperação de mais 5 documentos. A partir da leitura dos documentos selecionados, encontrou-se o termo “Green Data Science”, que passou a ser incluído na pesquisa bibliográfica, resultando na identificação de mais 16, totalizando 95 documentos. Após a eliminação dos itens duplicados, análise do título e resumo para identificação da pertinência com o objetivo do estudo, foram selecionados 24 documentos para leitura (Tabela 1).

Tabela 1 - Resultado da pesquisa bibliográfica

Termos	Scopus	Web of Science	Google scholar	Medline
Responsible Data Science	7	1	40	6
FACT + Data Science	0	0	10	0
FAT + Data Science	0	0	0	0
Fairness, accuracy, confidentiality, transparency + Data Science	5	2	2	1

Green data Science	1	0	15	0

Fonte: Autores

Esclarecemos que a busca do termo "Responsible Data Science" no Google resultou na recuperação de 14.600 itens, que incluem *blogs*, *lectures*, apresentações em *slide*, entrevistas, *posts* no Twitter, artigos (recuperados via Google Scholar). Em vista da grande quantidade, com pouca precisão, e pela diversidade de tipo de publicações, não foram incluídos os resultados obtidos com a busca no Google, apenas os identificados no Google Scholar.

A partir da análise dos 24 documentos, foram extraídas informações quanto ao tipo de publicação, de pesquisa, conceitualização dos termos "Responsible Data Science", "Green Data Science", "Fairness", "Accuracy", "Confidentiality", "Transparency" e identificação de iniciativas/projetos alinhados à RDS.

Observamos que, embora os termos que representam o acrônimo FACT possam trazer conceitos ambíguos, a depender do contexto e do domínio, eles foram pouco discutidos em relação à semântica pelos autores dos artigos. Nesse sentido, ao final deste estudo, é apresentado um glossário com esses termos e termos transversais.

A iniciativa Responsible Data Science representa um grupo que procura alternativas para a implantação do FACT e as ações em busca de resultados são contempladas pela Green Data Science, termo cunhado por um dos fundadores do grupo RDS.

Dos 24 documentos analisados, foram identificados 9 artigos científicos e 7 *papers*/apresentações (conferências), 2 *slide*, 4 *blogs* e 2 capítulos de livro (Quadro 1). Na categoria "Tipo de estudo": 10 são artigos de pesquisa, 7 são artigos de revisão, 1 estudo de caso, 2 livros e 4 apresentam opiniões de *experts* postadas em *blogs* ou sites de instituições.

Quadro 1 - Research Data Science: documentos analisados

Id	**Autoria**	**Título**
1	AIMS (2017)	Responsible Data Science. Ensuring fairness, accuracy, confidentially, transparency
2	de Jong et al. (2018)	CLARIN: towards FAIR and Responsible Data Science using language resources
3	de Smedt et al. (2018)	Towards an open science infrastructure for the digital humanities
4	Fišer, Lenardic, & Erjavec (2018)	CLARIN's Key resource families
5	GE 301 Group 7. (2017)	Responsible Data Science
6	Kemper & Kolkman (2018)	Transparent to whom? No algorithmic accountability without a critical audience
7	Moerel (2016)	GDPR conundrums - the data protection officer requirement
8	Ohm (2014)	Changing the rules: general principles for data use and analysis
9	Piersma (s.d.)	Data in urban environments
10	Srivastava, Scannapieco, & Redman (2019)	Ensuring high-quality private data for Responsible Data Science
11	Stoyanovich et al. (2017)	Fides: towards a platform for Responsible Data Science
12	Stoyanovich et al. (2018)	Panel: a debate on data and algorithmic ethics
13	Stoyanovich, & Howe (2018)	Follow the Data!
14	Stoyanovich, Howe, & Jagadish (2018)	Special session: a technical research agenda in data ethics and responsible data management
15	Stoyanovich, Yang, & Jagadish (2018)	On line set selection with fairness and diversity

Id	Autoria	Título
16	Taylor (2018)	Data, visibility and justice
17	Taylor, & Broeders (2015)	In the name of development
18	van Berchum, & Trippel (2015)	CLARIN data management Activities
19	van der Aalst (2016)	Green Data Science
20	van der Aalst et al. (2018)	Views on the past, present, and future of business information systems engineering
21	van der Aalst et al.	Responsible data science: using event data in a "people friendly manner"
22	van der Hoven	SoBigData ethics unpacking privacy
23	Veuger	Attention to disruption and blockchain creates
24	Veuger	Trust in a viable real estate economy with disruption and blockchain

Fonte: Autores

Apesar de os relatos de experiências ainda serem poucos, orientações e ações que podem atender às diretrizes FACT são apresentadas pela maior parte das publicações consultadas. A seguir, expomos direcionamentos e iniciativas que podem contribuir para a aplicação do FACT (Quadros 2 e 3).

Quadro 2 – Direcionamentos relativos às diretrizes FACT

Princípios **Abordagens**	**Imparcialidade**	**Precisão**	**Confidencialidade**	**Transparência**
DIRECIONAMENTO	Três pilares são apontados como base para a imparcialidade internacional de dados: (in) visibilidade, (des) engajamento com tecnologia e antidiscriminação Esses pilares integram direitos positivos e negativos e liberdades (16)	Os líderes devem defender as abordagens a priori de gestão de dados (10)	O controle de acesso à identidade e/ou papel do indivíduo que busca a identificação e o acesso aos dados precisa se efetivar por meio de uma política e de tecnologia (10)	As agências devem disponibilizar ao público informações sobre a coleta de dados e a metodologia de pré-processamento, em termos de suposições, critérios de inclusão, fontes de viés conhecidas e qualidade dos dados (13)
	A economia política de dados prevê o envolvimento metodológico em que se inclui a identificação sobre o que é, quem é importante e como eles se relacionam com os resultados desejados (16)	Os algoritmos para várias configurações de problemas com dados de *streaming* são desenvolvidos e uma decisão on-line deve ser feita em cada item à medida que forem apresentados (15)	As diretrizes para a privacidade: • consentimento informado • direito de ser esquecido • gerenciamento de identidade • privacidade recíproca • granulação grosseira, anonimização • vigilância, contravigilância • detecção de violação / intrusão • aplicativos de Big Data para detectar violações de privacidade em Big Data (21)	As agências governamentais devem disponibilizar publicamente resumos com as propriedades estatísticas relevantes do conjunto de dados que possam auxiliar na tomada de decisão mediante a preservação de dados e a privacidade dos indivíduos
	No domínio das	Os cientistas de dados	Os conjuntos de	Os técnicos devem

Princípios Abordagens	Imparcialidade	Precisão	Confidencialidade	Transparência
DIRECIONAMENTO	humanidades digitais, a utilização de dados deverá ser realizada com a preocupação de evitar dados tendenciosos que inflinjam a privacidade, a confidencialidade e a transparência de modo a comprometer os resultados por ausência de confiabilidade (3)	devem enfatizar a abordagem qualitativa adotando métodos e técnicas a priori para avaliar a qualidade de dados (10)	dados sintéticos que preservam a privacidade, quando apropriados, podem ser liberados em vez de conjuntos de dados reais para expor certas características dos dados (ex.: conjuntos de dados reais sensíveis) (13)	construir a priori métodos de qualidade de dados na Internet das Coisas; ferramentas para limpeza, transporte e integração de dados (10)
	Se os dados de treinamento são tendenciosos ou têm erros, é lógico que o resultado algorítmico seja injusto ou errado (14)	O uso de Big Data envolve a análise remota para tornar populações visíveis. Tal visibilidade cria uma força sobre os assuntos dos dados por meio de volume de dados ao invés de precisão (accuracy) e detalhe (17)	Os empresários devem focar no potencial dos mercados de dados que equilibram a variação dos preços com os *trade-offs* de qualidade de dados privados (12)	Os cientistas sociais devem contribuir para o entendimento das relações de confiança e qualidade de dados e para a pertinência da Ciência de Dados Responsáveis (21)
	As técnicas de ciência de dados precisam garantir justiça. As decisões e percepções automatizadas não devem ser usadas para discriminar de maneira que são inaceitáveis do ponto de vista legal ou ético (19)	A incerteza dos dados em Big Data está relacionada com o volume, velocidade, variedade e veracidade. A veracidade está associada à confiabilidade dos dados de entrada. Para melhor interpretação dos resultados das análises, podem ser adotados a imprecisão explícita ou os diagnósticos de confiança mais explícitos (19)	Quando apropriados, conjuntos de dados sintéticos que preservam a privacidade, podem ser liberados em vez de conjuntos de dados reais para expor certas características dos dados, se os conjuntos de dados reais forem sensíveis e não puderem ser liberados para o público (13)	Os dados coletados devem ser mantidos e disponibilizados para tornar o processo de tomada de decisão transparente (13)
	A prevenção da discriminação com a criação de algoritmos de decisão automatizados que não discriminam, usando variáveis sensíveis (sexo, idade, nacionalidade etc.), utilizando um método predefinido; pré-processamento ou pós-processamento (19)		A confidencialidade dos dados é um dos aspectos mais difíceis de se atingir. Os princípios de proteção de dados não devem ser aplicados a informações anônimas ou que não estejam relacionadas com uma pessoa física identificada ou identificável ou com dados tornados anônimos de forma que não seja mais possível identificar a pessoa (19)	A interpretabilidade deve apresentar as propriedades estatísticas de um conjunto de dados, a metodologia usada para produzi-los. Em última análise, fundamentar sua "adequação ao uso" no contexto de um sistema ou tarefa de decisão automatizada específica é fundamental para a transparência dos dados (13)
			A confidencialidade pode estar ameaçada	O cientista de dados deve trabalhar de forma

Princípios Abordagens	Imparcialidade	Precisão	Confidencialidade	Transparência
DIRECIONAMENTO			ao longo do processo de análise dos resultados. É sugerido o uso de métodos de desidentificação como: remoção de variáveis, randomização, *hashing*, *shuffling*, subamostragem, agregação, truncamento, generalização, adição de ruído, entre outros. Não obstante, essas ações podem impactar a qualidade dos resultados (19)	responsável nas fases de recebimento, manipulação, modelagem e implantação de dados e deve persistir quando os resultados do modelo algorítmico forem interpretados e o modelo algorítmico mantido (11)
			O uso de dados tendenciosos viola a privacidade ou confiabilidade, e a ausência de transparência, pode distorcer conclusões e afetar as relações de confiança (2)	As agências devem: a) disponibilizar ao público informações sobre a coleta de dados e a metodologia de pré-processamento, em termos de suposições, critérios de inclusão, fontes de viés conhecidas e qualidade dos dados; b) anotar adequadamente os conjuntos de dados, quando eles forem compartilhados, e manter informações sobre como os conjuntos de dados são adquiridos e manipulados; c) explicar as propriedades estatísticas dos conjuntos de dados; d) descobrir fontes de preconceitos e fazer declarações sobre qualidade de dados e adequação para uso (13)
			As licenças podem ser estabelecidas entre provedor de dados, repositórios e usuários finais em que se incluem restrições e responsabilidades e assegurem o compromisso da utilização de dados bem descritos, como forma de resguardar a privacidade (2)	As abordagens qualitativas podem ajudar a elaborar relatos detalhados e ricos do uso de modelos algorítmico, possibilitando a análise do contexto em que eles se inserem (10)
			A criação de leis de privacidade deveria abranger o contexto	As etapas de conversão de dados devem ser explicitadas para o

Princípios Abordagens	Imparcialidade	Precisão	Confidencialidade	Transparência
DIRECIONAMENTO			de determinado conteúdo e escopo adequado (8)	resultado final na forma de pipeline (5)
				As decisões (baseadas em model os de processo ou padrões frequentes e regras de decisão automatizadas) e os resultados da análise devem ser transparentes para garantir a aceitação e o uso adequado das técnicas de ciências dos dados (19)
				A transparência deve envolver a tomada de decisão automatizada, a explicação de decisões individuais, a inteligibilidade, clareza e a compreensão dos resultados da análise. O acesso aos dados e a técnica de análise utilizada devem ser explicitados por meio de link (19)
				A existência de um vínculo claro entre os dados e os resultados (histórias da análise) favorece a transparência. É necessário o detalhamento e a inspeção dos dados na perspectiva do modelo, de modo que seja possível reproduzir os resultados da análise dos dados originais (19)
				As diretrizes FACT devem ser aplicadas desde a fase que os cientistas de dados recebem os dados, passando pela fase de disputa de dados, modelagem, implantação até a interpretação dos resultados do modelo algorítmico (11)
				Uma infraestrutura de pesquisa pode ser considerada "responsável" apenas se acomoda usuários de dados, controladores e processadores como responsáveis pela produção de resultados epistêmicos (6)

Nota: *() Identificação da fonte relacionada no Quadro 1.
Fonte: Autores

Quadro 3 – Iniciativas relativas às diretrizes FACT

Princípios **Abordagens**	**Imparcialidade**	**Precisão**	**Confidencialidade**	**Transparência**
INICIATIVAS	O repositório Clarin inclui ações como contrato de licença de usuário final, termos de serviços dos dados, adição de restrições e responsabilidades para o usuário final, principalmente no que se refere à privacidade (3)	O compartilhamento de dados com preservação de privacidade utilizou Secure Multiparty Computation (SMC) (10)	O anonimato de dados, a partir do uso de modelo de privacidade diferencial local (LDP), combina a garantia estatística de privacidade diferencial e a segurança de que a informação nunca será visível em sua forma bruta (10)	O registro de dados de proveniência nos metadados CLARIN torna os dados rastreáveis e o uso de PIDs torna os dados passíveis de serem citados e seu uso replicável (3)
	O repositório Clarin requer a proveniência dos dados a partir dos metadados, que possibilitam a rastreabilidade dos dados e a inclusão de PIDs que viabilizam a citação a e replicabilidade dos dados. Está em desenvolvimento a inclusão de amostras melhoradas, enriquecimento dos dados com o uso de metadados (extralinguísticos), ligação com fontes externas (ex. Gazetters) e anotações em nível conceitual (3)		O artigo versa sobre o General Data Protection Regulation (GDPR) e os requisitos para a contratação de um Data Protection Officer (DPO) propõem uma coleção de técnicas e abordagens que buscam facilitar o acesso aos benefícios da Ciência de Dados e Big Data enquanto garante imparcialidade, confidencialidade, precisão (correção) e transparência (21)	A aprovação de transparência algorítmica é apontada como uma forma de proporcionar a transparência (13)
			A análise e visualização de dados são realizadas pela abordagem de banco de dados gráficos (Graph Vision) e implementação de *machine learning* como tecnologias de análise de dados (5)	As agências devem disponibilizar resumos publicamente disponíveis de propriedades estatísticas relevantes dos conjuntos de dados que possam auxiliar na interpretação das decisões tomadas com os dados, aplicando métodos de última geração para preservar os dados e a privacidade dos indivíduos, além de liberar conjuntos de dados de treinamento e validação sempre que possível (13)
			O sistema de avaliação de Impacto Privado (PIA) ajuda a entender melhor como as informações pessoais podem ser usadas, armazenadas, compartilhadas e também	A geração de linguagem natural é utilizada para transformar os resultados da análise selecionada em relatórios individualizados, concisos e fáceis de ler (19)

Princípios Abordagens	Imparcialidade	Precisão	Confidencialidade	Transparência
INICIATIVAS			diminui os riscos de privacidade (5)	
				A transparência do algoritmo não é suficiente para garantir a precisão dos resultados. É preciso conhecer os dados utilizados (14)
				A transparência dos dados, por si só, não garante o sucesso das tomadas de decisões. São necessárias a contextualização e a literacia de dados, por parte dos gestores, que impactem negativamente nas decisões de gestores e legisladores (14)

Nota: *() Identificação da fonte relacionada no Quadro 1.
Fonte: Autores

Ainda que de importância ímpar para que sejam atendidas questões referentes à ética, à economia e aos aspectos sociais, é tímida a sinalização das diretrizes do FACT na literatura. Nos artigos consultados, a transparência é o princípio mais comentado pela maior parte dos artigos consultados, o que denota a preocupação dos autores com a demanda da sociedade sobre o tema, intimamente relacionado com a disponibilização pública de dados. O princípio da confidencialidade foi o segundo mais abordado, seguido por imparcialidade e precisão. Ao analisar esses resultados, observamos que existem dificuldades em adotar os princípios de forma igualitária e eficaz. Como exemplo, têm-se as ações que, ao preservarem a confidencialidade e a imparcialidade, incidem nos níveis de transparência.

O posicionamento de Kemper & Kolkman (2018) é de que deve haver mais estudos empíricos sobre modelos algorítmicos para testar a responsabilidade algorítmica, pois só assim será possível desenvolver diretrizes sólidas adequadas às práticas existentes. Esse deve ser o grande desafio, já que a existência de legislação não assegura a efetividade do uso desses princípios (Moerel, 2016, Ohm, 2014).

De acordo com Ohm (2014), a maioria das leis de privacidade se concentra na coleta e na divulgação, e não no uso. Por outro lado, ainda existem diferenças nas leis de um país para outro. A lei de privacidade dos Estados Unidos é diferente da Diretiva de Proteção de Dados da União Europeia, mesmo se tratando de aspectos cujo danos possam ser globais. A regra de que "os dados devem ser processados para um fim específico e subsequetemente, usados ou comunicados apenas na medida que não são incompatíveis com a utilidade da transferência" também parece intangível. Trabalhar com grandes dados é trabalhar com o imprevisível. Sendo assim, como assegurar o resultado final da análise dos dados? Questionamento que motiva discussão e investigação no universo dos dados.

Por outro lado, agentes, como os analistas, programadores, cientistas de dados e o governo, têm papel imprescindível para que esses princípios sejam resguardados de forma a promover uma sociedade igualitária. No Quadro 3 são apresentadas algumas iniciativas relativas às diretrizes FACT. Como exposto neste estudo, as ações para garantir os quatro princípios são

concebidas mais lentamente do que as tecnologias desenvolvidas para a análise e disseminação dos dados.

Pode-se observar que, em relação aos direcionamentos e iniciativas identificados na pesquisa, eles ainda demandam um grande esforço conjunto e interdisciplinar das áreas sociais e tecnológicas para que possam acontecer de fato. O lado social das nossas vidas é referido por Bauman (2001) como modernidade líquida, metáfora que ele adota para afirmar que tudo é fluído. Ou seja, Bauman considera que a modernidade que vivemos se caracteriza pela incapacidade de manter a forma.

Observando o período histórico que antecede essa fase, denominado por Bauman (2001) de "modernidade sólida", o autor apresenta ideias que apontam uma constância e segurança, pois a transformação de valores ocorria de forma lenta e perceptível, dando a sensação de certeza e controle sobre o que acontecia no mundo.

Assim, quando consideramos, por meio dessa abordagem da modernidade líquida, os direcionamentos que denotam mais o lado social, constatamos que eles podem ser mais complexos, especialmente em questões de discriminação e sensibilidade dos dados. De acordo com as iniciativas que mostram como vem sendo trabalhada a tecnologia em relação ao RDS, infere-se que os sistemas de tratamento dos dados demandam maior flexibilidade no sentido de adaptar-se às vertiginosas mudanças no direcionamento, provocadas pelo fluxo resultante da pressão exercida pelas condições comuns à "vida líquida" (Bauman, 2001).

5 Conclusão

O conceito de RDS apresenta um significativo avanço no contexto do Big Data e da Data Science, trazendo importantes elementos que favorecem uma reflexão por parte dos cientistas de dados. Em especial, o FACT identifica claramente os aspectos e facetas vinculados a esse conceito, bem como demonstra o que os cientistas de dados devem considerar durante os processos que compõem as fases do Ciclo de Vida dos Dados.

Observamos que as ações promovidas pelo grupo RDS, em relação à Green Data Science ou às diretrizes FACT, podem contribuir para salvaguardar os direitos individuais, sem que seja necessário adotar medidas que impeçam sumariamente o acesso e a reutilização de dados.

Dessa forma, destacamos que as técnicas de análise de dados, em conjunto com as tecnologias de Internet das Coisas, vêm sendo amplamente utilizadas sob uma perspectiva tecnicista, visando a gerar valor para as organizações. Nesse sentido, o estudo e a aproximação das ciências sociais a essas temáticas são relevantes para que questões sociais sejam consideradas durante os processos tecnológicos. Embora haja movimentos relevantes para aprovação de leis que garantam direitos individuais no contexto do Big Data, o uso da tecnologia é imprescindível para que esses direitos sejam assegurados.

Em vista do exposto, há necessidade de estudos empíricos sobre modelos de algoritmos, em que as medidas de transparência sejam analisadas na prática, de forma que os envolvidos na geração, armazenamento e tratamento dos dados adotem mecanismos que lhes assegurem confiança quanto aos serviços prestados, assim como aos dados acessados.

O desenvolvimento de tal tipo de tecnologia de análise de dados, de acordo com o modelo FACT de abordagem de ciência de dados responsável, é de importância crucial, porque a falta

de implementação dessa abordagem em tecnologias de dados pode resultar em consequências indesejáveis para a sociedade, causando impacto em três elementos: privacidade, segurança e responsabilidade. Nota-se que dificilmente serão contemplados os quatro princípios FACT simultaneamente, entretanto deve-se buscar um equilíbrio em que sejam consideradas as compensações – priorizar um em detrimento de outro(s). No caso de garantir a confidencialidade, provavelmente a precisão e a transparência serão comprometidas. Logo, a busca por um equilíbrio entre esses quatro princípios deve ser almejada tanto pelos desenvolvedores como pelos gestores de dados.

Referências

AIMS. (2017). *Responsible data science*. Ensuring, fairness, accuracy, confidentially, transparency. Recuperado em 10 janeiro 2019 de http://aims.fao.org/activity/blog/responsible-data-science-ensuring-fairness-accuracy-confidentially-transparency-fact.

Annany, M., & Crawford, K. (2016). Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Soc.*, *20*(3): 973–989.

Ávila, F. B. S.J. (1967). *Pequena enciclopédia de moral e civismo*. Rio de Janeiro: MEC.

Baurman, Z. (2001). *Modernidade líquida*. Rio de Janeiro: Zahar.

Collins dictionary. (2019). Recuperado em 2 março 2019 de https://www.collinsdictionary.com/pt/dictionary/english/accuracy.

Data Science Center Eindhoven. (2019). *Responsible Data Science*. Ensuring fairness, accuracy, confidentiality & transparency by design. Recuperado em 10 janeiro 2019 de https://www.tue.nl/en/research/research-areas/data-science/responsible-data-science/

de Jong, F. M. G., Maegaard, B., De Smedt, K., Fišer, D., & Van Uytvanck, D. (2018). CLARIN: towards FAIR and responsible data science using language resources. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. (p. 3259–3264). Recuperado em 10 janeiro 2019 de https://dspace.library.uu.nl/handle/1874/364776.

de Smedt, K., de Jong, F., Maegaard, B., Fišer, D., & Van Uytvanck, D. (2018). Towards an open science infrastructure for the digital humanities: the case of CLARIN. *CEUR-WS.org*, *2084*: 1–12. Recuperado em 10 janeiro 2019 de http://ceur-ws.org/Vol-2084/paper11.pdf.

Euronews (2019). France fines Google €50 million using EU's transparency and consent law. Recuperado em 10 janeiro 2019 de https://www.euronews.com/2019/01/21/france-fines-google-50-million-using-eu-s-transparency-and-consent-law.

European Data Science Academy. (2019) *About EDSA*. Recuperado em 28janeiro 2019 de http://edsa-project.eu/overview/about-edsa/.

Facebook: Cambridge analytica data scandal (2019). *Wikipedia*: The Free Encyclopedia. Apr. 26, 2019. Recuperado em 26 de abril de 2019 de https://en.wikipedia.org/wiki/Facebook% E2%80%93Cambridge_Analytica_ data_scandal

Fišer, D., Lenardič, J., & Erjavec, T. (2018). CLARIN's key resource families. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. Recuperado em 10 janeiro 2019 de http://www.lrec-conf.org/proceedings/lrec2018/pdf/829.pdf.

GE 301 Group 7. (2017). *Responsible Data Science*. Recuperado em 11 janeiro 2019. Disponível em: http://ge301.bilkent.edu.tr/fall2017group7/.

Goldim, J. R. *Confidencialidade* (1997-2003). Recuperado em 15 março 2019 de https://www.ufrgs.br/bioetica/confiden.htm.

Haggerty, K.D., Ericson, R.V., 2000. The surveillant assemblage. *Br. J. Sociol*. 51 (4), 605–622

Hilbert, M., López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Sciencexpress*. 1-7. Disponível em:Recuperado em 18 março 2019 de http://www.ris.org/uploadi/editor/1304938275129769729 4Science-2011-Hilbert-science.1200970.pdf

Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Inf. Commun. Soc.* 1–16. Recuperado em 11 janeiro 2019 de https://doi.org/10.1080/1369118X.2018.1477967

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes the premise, the proposed solutions, and the open challenges. *Philosophy and Technology*. 31(4):611-627. doi: 10.1007/s13347-017-0279-x.

Lodder, G. M. A., Scholter, R. H. J., Goossens, L., Engels, R. C. M. E., Verhagen, M. (2016). Loneliness and the social monitoring system: emotion recognition and eye gaze in a real-life conversation. *Br. J. Psychol*. 107(1):135-153. Recuperado em 18 março 2019 de https://doi.org/10.1111/bjop.12131.

Moerel, L. (2016, julho 19). GDPR conundrums: the data protection officer requirement. Recuperado em 10 janeiro 2019 de https://research.tilburguniversity.edu/en/ publications/gdpr-conundrums-the-data-protection-officer-requirement.

Moerel, L., Prins, C. (2016). Privacy for the homo digitalis: proposal for a new regulatory framework for data protection in the light of Big Data and the Internet of Things. May 25, 2016. Recuperado em 10 janeiro 2019 de http:??dx.doi.org/10.2139/ssrn.2784123.

Ohm, P. (2014). Changing the rules: general principles for data use and analysis. In: Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Ed.). *Privacy, big data, and the public good*: frameworks for engagement. Cambridge: Cambridge University Press. v.1, p. 96-111.

Pennock, M. (2007): Digital curation: a life-cycle approach to managing and preserving usable digital information. *Library & Archives Journal*, (1). Recuperado em 10 janeiro 2019 de http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf.

Piersma, N. (2018). Data in urban environments. In: Piersma, N. (Ed.). *Through the clouds: urban analytics for smart cities*. Amsterdam: Hogeschool van Amsterdam. p. 11-21.

Satariano, A. (2019 Jan 21). Google is fined $57 million under Europe's Data Privacy Law. *New York Times*. Recuperado em 21 março 2019 de https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html.

Srivastava, D., Scannapieco, M., & Redman, T. C. (2019). Ensuring high-quality private data for responsible data science: vision and challenges. *ACM Journal of Data and Information Quality (JDIQ)*. *11*(1), 1.

Stoyanovich, J., & Howe, B. (2018 Nov 27). *Follow the data*! Algorithmic transparency starts with data transparency. Recuperado em 21 março 2019 https://ai.shorensteincenter.org/ideas/2018/11/26/follow-the-data-algorithmic-transparency-starts-with-data-transparency.

Stoyanovich, J., Howe, B., & Jagadish, H. V. (2018). Special Session: A technical research agenda in data ethics and responsible data management. In *Proceedings of the 2018 International Conference on Management of Data* (p. 1635–1636). pp. 1635–1636.

Stoyanovich, J., Howe, B., Abiteboul, S., Miklau, G., Sahuguet, A., & Weikum, G. (2017). Fides: Towards a platform for responsible data science. In SSDBM'17. *29th International Conference on Scientific and Statistical Database Management*, Jun 2017, Chicago, United States. 10.1145/3085504.3085530. hal-01522418. Recuperado em 10 janeiro 2019 de https://hal.inria.fr/hal-01522418/document

Stoyanovich, J., Howe, B., Jagadish, H. V., & Miklau, G. (2018). Panel: a debate on data and algorithmic ethics. *Proceedings of the VLDB Endowment*, *11*(12): 2165–7.

Stoyanovich, J., Yang, K., Jagadish, H. V. (2018). Online set selection with fairness and diversity constraints. In *Proceedings of the 21st International Conference on Extending Database Technology (EDBT*), March 26-29, 2018. Recuperado em 10 janeiro 2019 de https://openproceedings.org/2018/conf/edbt/paper-98.pdf.

Taylor, L. (2017 July/Dec). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society.* pp. 1-14. Recuperado em 10 janeiro 2019 de https://doi.org/10.1177/2053951717736335.

Taylor, L. (2017). *Data, visibility and justice*. Recuperado em 10 janeiro 2019 de https://redasci.org/wp-content/uploads/2016/10/Linnet-Taylor-RDS-16.3.17.pdf.

Taylor, L., & Broeders, D. (2015). In the name of development: power, profit and the datafication of the global south. *Geoforum. 64*. pp. 229–37.

van Be.rchum, M., & Trippel, T. (2018). CLARIN data management activities in the PARTHENOS context. In *CLARIN Annual Conference 2018*. pp. 95-99. Recuperado em 10 janeiro 2019 de https://ris.utwente.nl/ws/portalfiles/portal/63914609/CE_2018_1292_CLARIN2018_ConferenceProceedings.pdf#page=102.

van der Aalst, W. M. (2016). Green data science: using big data in an" environmentally friendly" manner. In *18th International Conference on Enterprise Information Systems (ICEIS 2016).* Apr. 25-28, 2016. Rome: SciTePress. pp. 9-21. Recuperado em 10 janeiro 2019 de https://pdfs.semanticscholar.org/5889/68dd392ae93b1524aa7a491917d839bca050.pdf.

van der Aalst, W. M., Becker, J., Bichler, M., Buhl, H. U., Dibbern, J., Frank, U., … Hui, K.-L. et al. (2018). Views on the past, present, and future of business and information systems engineering. *Bus. Inf. Syst. Eng*. *60*(6): 448–450. Recuperado em 10 janeiro 2019 de https://doi.org/10.1007/s12599-018-0561-1

van der Aalst, W. M., Bichler, M., & Heinzl, A. (2017). Responsible data science. *Bus. Inf. Syst. Eng. 59*(5): 311–313. Recuperado em 10 janeiro 2019 de https://link.springer.com/article/10.1007%2Fs12599-017-0487-z.

van der Hoven, J. (2018). *SoBigData ethics unpacking privacy designing for responsibility*. Recuperado em 10 janeiro 2019 de https://slideplayer.com/slide/13113648/.

Veuger, J. (2017). Attention to disruption and blockchain creates a viable real estate economy. *J. US-China Public Adm*. *14*(5): 263–85. Recuperado em 10 janeiro 2019 de https://davidpublisher.org/Public/uploads/Contribute/ 5a3c644925d78.pdf.

Veuger, J. (2018). Trust in a viable real estate economy with disruption and blockchain. *Facilities*. 36(1/2):103–20. Recuperado em 10 janeiro 2019 de https://doi.org/10.1108/F-11-2017-0106.

Wilkinson, M. D., Dumontier, M., Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data,* 3(160018).

Agradecimentos

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UID/CEC/00319/2019.

À Universidade Federal do Espírito Santo pela liberação para participar do Evento.

Glossário

Termo	Sigla	Tradução	Definição
Green Data Science	GDS	Ciência de Dados Verde	Coleção de técnicas e abordagens que busca facilitar o acesso aos benefícios da ciência de dados e big data enquanto garante imparcialidade, confidencialidade, exatidão (correção) e transparência (van der Aalst, 2016).
Fairness		Imparcialidade, tratamento justo, igualdade	É a atitude de quem considera, objetivamente, sem paixões ou preconceitos, um determinado fato na totalidade de seus aspectos (Ávila, 1967).
Confidentiality		Confidencial, privado, manter em segredo informação privada	Dever de resguardar todas as informações que dizem respeito a uma pessoa, isto é, à sua privacidade (Goldim, 1997).
Accuracy		Exato, correto	Precisão das informações ou medições é a sua qualidade de ser verdadeira ou correta, mesmo em pequenos detalhes (Collins Dictionary, 2019)
Transparency		Transparência	“'compreensibilidade de um modelo específico'” (Stoyanovich et al., 2018)
Business Process Model and Notation	BPMN		É um padrão para modelagem de processo de negócio que fornece uma notação gráfica para especificar processos de negócios em um Business Process Diagrama (BPD) baseado em uma técnica de *flowcharting* muito parecida com diagramas de atividade da Unified Modeling Language (UML)
Datafication		Datificação	Refere-se a um conjunto de ferramentas, tecnologias e processos utilizados para transformar uma organização em uma empresa orientada por dados. Também se utiliza DATAFY. Uma empresa que implementa datafication é chamada DATAFIED.
Data Doubles		Dubles Dados	Representação abstrata de pessoas por meio de seus dados, com o auxílio de monitoramento e focando (alvo) pessoas para intervenção (Haggerty & Ericson, 2000).
Data Management		Gestão de Dados	É a atividade de criar, prover, manter e arquivar dados de pesquisa em todos os estágios do ciclo de vida dos dados de pesquisa (Pennock, 2007)
Dados Findable, Acessible, Interoperable and Re-Usable	FAIR	Dados Recuperáveis, Acessíveis, Interoperáveis, Re-utilizáveis (Findable, Acessible, Interoperable and Re-usable)	São dados que atendem aos padrões de localização, acessibilidade, interoperabilidade e reusabilidade (Wilkinson, Dumontier, & Mons, 2019)
Modelol algoritmico			Um modelo algorítmico é uma representação formal de um objeto que um observador pode usar para responder a perguntas sobre esse objeto. Dentro dessa definição, 'observador' refere-se a um humano ou decisor de máquina (Frigg & Hartmann, 2012, Gross & Strand,

Termo	Sigla	Tradução	Definição
			2000, Haag & Kaupenjohann, 2001, Minsky, 1965, apud Kemper & Kolkman, 2017).
Data Science			A ciência de dados é um campo interdisciplinar com o objetivo de transformar os dados em valor real. Os dados podem ser estruturados ou não estruturados, grandes ou pequenos, estáticos ou em fluxo contínuo. A ciência de dados inclui extração de dados, preparação de dados, exploração de dados, transformação de dados, armazenamento e recuperação, infraestruturas computacionais, vários tipos de mineração e aprendizado, apresentação de explicações e predições, e também exploração de resultados levando em consideração aspectos éticos e sociais, aspectos legais e comerciais (Aalst, 2016)
Big Data			A quantidade de dados (brutos) que são produzidos e o crescimento exponencial desses dados é denominado de Big Data (Piersman, 2018). Big Data também se refere a tecnologias que trabalham com grandes quantidades de dados (GE 301 Group 7, 2017).
Digital transformation		Transformação digital	A transformação digital é definida pelo uso de tecnologia digital para transformação de negócios e atividades organizacionais, processos, competências e modelos (Indústria 4.0) e seu impacto acelerado na sociedade humana (Sociedade 5.0, às vezes denominada de “A sociedade sem papel” ou “A sociedade do conhecimento”) (Piersman, 2018).
Discriminação			Tratamento prejudicial a um indivíduo baseado em sua participação em um grupo ou categoria específica (raça, gênero, nacionalidade, deficiência, estado civil ou idade)” (van der Aalst, 2016).
Desindentificação de dados			Refere-se ao processo de remoção ou obscurecimento de variáveis com o objetivo de minimizar as divulgações não intencionais. Em muitos casos, a reidentificação é possível através da ligação de dados de diferentes fontes (van der Aalst, 2016).