

# Interloquções bibliográficas e epistemológicas entre a ciência de dados e a ciência da informação

**Jorge Henrique Cabral Fernandes**

Doutor, Universidade Federal de Pernambuco (UFPE), Recife, PE, Brasil

Professor Adjunto IV, Universidade de Brasília (UnB), Brasília, DF, Brasil

<http://lattes.cnpq.br/7151669913805328>

E-mail: [jhcf@unb.br](mailto:jhcf@unb.br)

Submetido em: 30/04/2020. Aprovado em: dd/mm/yyyy. Publicado em: dd/mm/yyyy .

## RESUMO

Diante da grande evidência alcançada recentemente por métodos, ferramentas e práticas da ciência de dados, este artigo verifica a possibilidade de que existam novos fundamentos em uma suposta ciência chamada de Ciência de Dados e de que forma ela impactaria ou seria impactada pela Ciência da Informação. A metodologia é exploratória e usa pesquisa bibliográfica baseada em análise de redes de co-ocorrências e de acoplamento bibliográfico envolvendo os termos “data science”, “information science” e “library and information science”. As interloquções e contribuições dos dois campos de práticas, evidenciadas pelas análises produzidas, bem como de algumas implicações epistemológicas, apontam não apenas para uma oportunidade de intercâmbios, mas para uma necessidade de que eles ocorram com a maior brevidade.

**Palavras-chave:** Big data. Biblioteconomia. Análise de co-ocorrência. Acoplamento bibliográfico. Desenvolvimento profissional.

## *Bibliographic and epistemological exchanges between data science and information science*

### ABSTRACT

*In view of the great recent evidence achieved by methods, tools and practices of data science, the article verifies the possibility that there are new foundations in a supposed science called Data Science, and how it would impact or be impacted by Information Science. The methodology is exploratory, and uses bibliographic research based on analysis of networks of co-occurrences and bibliographic coupling involving the terms “data science”, “information science” and “library and information science”. The exchanges and contributions of the two fields of practices, evidenced by the analyzes produced, as well as some epistemological implications, point not only to an opportunity for exchanges, but to a need for them to occur as soon as possible.*

**Keywords:** *Big data. Libray and information science. Analysis of co-occurrences. Bibliographic coupling. Professional development.*

## **Interlocuciones bibliográficas y epistemológicas entre ciencia de datos y ciencia de la información**

### **RESUMEN**

*En vista de la gran evidencia alcanzada recientemente por los métodos, herramientas y prácticas de la ciencia de datos, el artículo verifica la posibilidad de que haya nuevas bases en una supuesta ciencia llamada Ciencia de Datos, y de qué manera impactaría o sería impactada por la Ciencia de Información. La metodología es exploratoria y utiliza investigación bibliográfica basada en análisis de redes de coincidencias y acoplamiento bibliográfico que involucra los términos “ciencia de datos”, “ciencia de la información” y “biblioteca y ciencia de la información”. Las interlocuciones y contribuciones de los dos campos de prácticas, evidenciadas por los análisis producidos, así como algunas implicaciones epistemológicas, apuntan no solo a una oportunidad para intercambios, sino a la necesidad de que ocurran lo antes posible.*

**Palabras clave:** Big data. Biblioteconomía. Análisis de concurrencia. Acoplamiento bibliográfico. Desarrollo profesional.

### **INTRODUÇÃO**

É inegável a intensa valorização recente da área de atividade profissional associada ao termo “Ciência de Dados” (DAVENPORT; PATIL, 2012). Seria a ciência de dados uma área científica em si ou trata-se apenas de uma atividade passageira que, devido à pouca consistência, coesão ou abrangência de propósito, estaria fadada a desaparecer, sendo absorvida por áreas já estabelecidas, como computação, estatística, ciência da informação ou administração? Traria, a área de ciência de dados, novos fenômenos dignos de investigação nos campos filosóficos, epistemológicos e científicos da atividade humana, bem como numa prática profissional distintiva?

Embora exista uma razoável confusão sobre o que de fato faz um cientista de dados e sobre o que seja a ciência de dados, pode-se recorrer à pesquisa em bases de dados bibliográficas para que se tenha uma compreensão da atividade. Também é admissível o aprofundamento filosófico-epistemológico da ciência de dados, a fim de identificar se, de fato, um conjunto distintivo de fenômenos de relevância científica pode ser agregado a essa suposta nova ciência. Diante disso, este artigo busca verificar se existem novos fundamentos a explorar na presumida ciência de dados e, se for esse o caso, quais seriam as interlocuções possíveis dessa nova ciência com a já estabelecida ciência da informação.

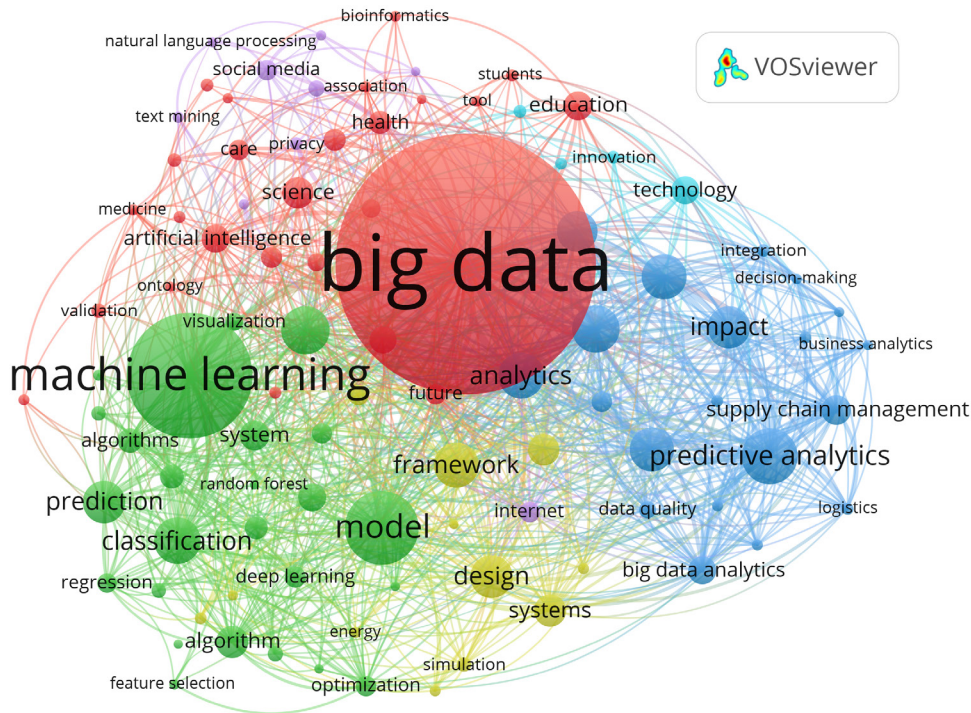
### **METODOLOGIA**

Este trabalho adota, como metodologia, a análise de registros bibliográficos obtidos na *Core Collection* da base *Web of Science* (WoS), ofertada por meio do portal de periódicos da CAPES e associada às publicações de artigos científicos até abril de 2020. Também emprega a ferramenta VOSViewer (VAN ECK; WALTMAN, 2010) para geração de redes e análises de co-ocorrência de palavras-chave e acoplamento bibliográfico.

Nas redes de co-ocorrência de palavras-chave, duas palavras-chave são vinculadas uma à outra, caso apareçam simultaneamente em um documento. Quanto maior a quantidade de documentos em que esse par ocorre, maior é a associação entre essas duas palavras-chave. A suposição é de que quanto maior for a co-ocorrência dessas palavras, mais afinidades elas possuem entre si.

Nas redes de acoplamento bibliográfico, considera-se que duas unidades de análise (documentos, fontes de publicação, autores, organizações ou países) estão vinculadas a partir do momento em que citam a(s) mesma(s) referência(s) bibliográfica(s). Nessa perspectiva, quanto maior a quantidade de referências comuns citadas, maior é a suposição de que as unidades de análise tratam do mesmo assunto.

Figura 1 – Rede de co-ocorrência de palavras-chave em publicações em ciência de dados



Fonte: Elaborada pelo autor.

## ANÁLISE E DISCUSSÃO DOS RESULTADOS

### A CONDIÇÃO BIBLIOGRÁFICA ATUAL DA DATA SCIENCE (CIÊNCIA DE DADOS)

A pesquisa na base WoS, conforme as abordagens apresentadas na metodologia, obteve, entre 2011 e 2020, 1.880 registros de artigos científicos associados aos termos “data science” ou “data-science”. A rede de co-ocorrência das 96 palavras-chave mais co-citadas nesses 1.880 registros, gerada pelo software VOSViewer, é apresentada na figura 1.

Para a apresentação de uma rede com, no máximo, 100 elementos, aumentou-se gradualmente o limite mínimo de co-citações para cada palavra-chave, alcançando-se a quantidade de 98 palavras-chave com um mínimo de 16 co-ocorrências. Foram excluídos os termos “data science” e “data-science” e tamanho das palavras-chave na figura 1 é proporcional à quantidade de vezes em que aparecem. Por fim, a rede da figura 1 é composta de seis *clusters* de palavras, delimitando os seguintes temas característicos da ciência de dados, no geral:

**Big data:** termo mais proeminentemente associado à ciência de dados e ao maior *cluster* da rede, composto de 26 termos. O *big data* representa a oportunidade que é perdida ou aproveitada pelas organizações humanas, na medida em que geram ou têm acesso a dados em grande volume, grande variedade, produzidos em alta velocidade, com ampla variação em veracidade e elevado potencial de produção de valor (WANG, 2018). O *cluster* dominado pelo *big data* está fortemente relacionado a todos os outros. Nesse *cluster*, ainda aparecem termos como “education”, “artificial intelligence”, “statistics”, “health” e “data analysis”, sugerindo a vinculação do *big data* à aplicação da inteligência artificial e de técnicas estatísticas para análises e sínteses de modelos nos campos de aplicação da educação e saúde;

**Machine learning:** segundo termo mais evidente em associação com a ciência de dados, domina o segundo maior *cluster* da rede, com 25 termos. A aprendizagem de máquina está associada aos termos “model”, “prediction”, “classification”, “algorithms” e compreende basicamente o uso de dados do passado para treinar sistemas algorítmicos e computacionais para identificação de padrões em dados, tendo por fim criar modelos de predição de eventos futuros ou de classificação de situações presentes, baseados em dados recém-disponíveis. Sheble (2016) ilustra situações em que a aprendizagem de máquina pode facilitar a seleção de artigos mais relevantes em pesquisas bibliográficas. Kazakci (2015) ilustra o modo como a aprendizagem de máquina pode auxiliar na identificação de sinais relevantes em um estudo para detecção de partículas de alta energia em física;

**Predictive analytics:** termo mais evidente do terceiro mais proeminente *cluster* do mapa de co-ocorrência de palavras-chave associadas às publicações em ciência de dados. A “analytics” preditiva consiste na aplicação da ciência de dados pela gestão (“management”), inclusive na cadeia de suprimento de empresas (“supply chain management”), visando ao melhor desempenho (“performance”) e gerando grande impacto (“impact”) nos negócios por meio da informação (“information”) para desenho de estratégia e da obtenção de vantagens competitivas em sistemas logísticos. Ortiz-Repiso, Greenberg e Calzada-Prado (2018) vinculam o uso de *analytics* em empresas à imensa valorização da atividade profissional dos chamados cientistas de dados;

**Framework:** termo mais evidente associado ao quarto *cluster* de palavras-chave em ciência de dados. Sua vinculação aos termos “systems”, “design”, “data analytics”, “simulation”, “dynamics” e “uncertainty”, evidencia a necessidade de desenvolvimento de arcabouços computacionais (frameworks) de sistemas (de informação) baseados em técnicas de simulações de sistemas dinâmicos, com vistas a reduzir a incerteza na tomada de decisão (“decision making”).

Kazakci (2015) apresenta de que forma o *design* coletivo de uma aplicação de aprendizagem de máquina é capaz de produzir inovações de forma muito ágil;

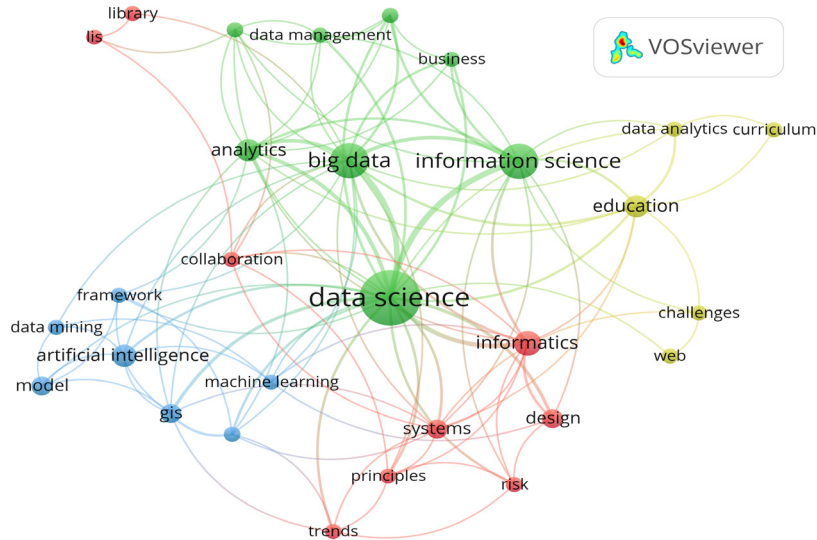
**Social media:** termo mais evidente do segundo menor *cluster*, com nove itens, e que destaca o intenso uso de abordagens em ciência de dados para analisar dados gerados pelas mídias sociais, a exemplo do *Twitter*, por meio de processamento de linguagem natural e mineração de textos. Essa utilização apresenta implicações éticas, ligadas a ameaças à privacidade na sociedade. A reduzida dimensão desse *cluster* evidencia a necessidade de avanços na compreensão dos problemas relativos à exploração de dados da sociedade sem o devido cuidado ético; e

**Tecnologia:** termo que representa o menor dos *clusters* de palavras-chave e está associada à produção de inovações. A ciência de dados, representada pela aprendizagem de máquina e pela inteligência artificial, é uma grande indutora de inovações.

A partir da análise dos *clusters* pode-se sintetizar a atividade de ciência de dados como sendo, atualmente, aquela que busca explorar as oportunidades do *big data* por meio da criação de sistemas computacionais de análise, inclusive de natureza preditiva, visando ao desenvolvimento de inovações tecnológicas para empresas em um ambiente competitivo, especialmente de cadeia de suprimentos, ou para usos em áreas como saúde e educação, mediante aplicação de técnicas como aprendizado de máquina e inteligência artificial. A atividade dos cientistas de dados, quando aplicada às mídias sociais, pode gerar implicações éticas de ameaça à privacidade.

## INTERLOCUÇÃO BIBLIOGRÁFICA DA CIÊNCIA DA INFORMAÇÃO COM A CIÊNCIA DE DADOS

Figura 2 – Mapa de co-ocorrência de palavras-chave que combinam ciência de dados e ciência da informação ou “library science”



Fonte: Elaborada pelo autor.

Uma segunda pesquisa na base de dados bibliográfica foi efetuada pelo autor, dessa vez usando a *string* de busca “((‘data science’ or data-science) and (‘information science’ or ‘library science’ or lis))”, aplicável todos os anos. Foi retornado um pequeno conjunto de documentos em todos os tipos disponíveis na base, 34 ao todo, sendo 25 artigos de *journal*, seis *proceeding papers* e três revisões.

### ANÁLISE DE CO-OCORRÊNCIAS

A figura 2 apresenta a rede de co-ocorrência de palavras-chave usadas nos 34 documentos que associam ciência de dados e ciência da informação, e que possuem pelo menos duas associações. A rede possui quatro *clusters*, sendo que os termos “data science”, “information science” e “big data” continuam a dominar o *cluster* principal, que possui oito elementos. Além disso, esse *cluster* apresenta ainda a ocorrência de termos ligados à gestão de dados e a negócios.

O *cluster* mais numeroso da rede, com nove elementos, apresentado em vermelho, possui o termo “informatics” como mais dominante, apesar de próximo dos termos “design” e “systems”.

Desse mesmo *cluster*, fazem parte os termos ligados às “library sciences” e “lis”, que, embora agrupados com “informatics”, se encontram posicionados próximos ao “big data”, “data management”. Nesse *cluster*, a ocorrência de vários termos sugere preocupações com fenômenos em contextos sociais, tais como riscos (“risks”), colaborações (“collaboration”), princípios (“principles”) e tendências (“trends”).

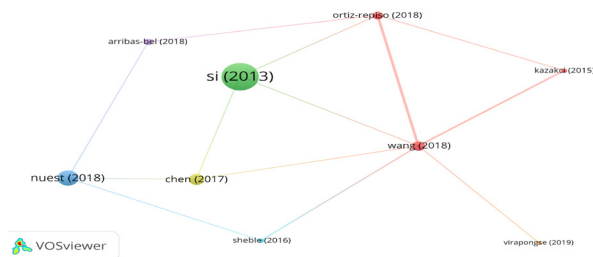
O terceiro mais importante *cluster* da rede apresenta predominância de termos de origem computacional e estatística, como inteligência artificial, aprendizagem de máquina, mineração de dados e sistemas de informação georreferenciados. Termos ligados a fatores humanos ou sociais estão ausentes nesse *cluster*.

O último e menor dos *clusters* de palavras-chave que aparecem na relação entre ciência de dados e ciência da informação enfatiza “education” e seus desafios (“challenges”), e o “curriculum” de “data analytics” apresenta-se como foco. Os termos sugerem a existência de desafios à incorporação de temas como *data science*, *big data* e *analytics* à ciência da informação ou, de outro modo, a possibilidade de também existirem desafios educacionais relacionados à incorporação de temas da ciência da informação à ciência de dados.

## ACOPLAMENTO BIBLIOGRÁFICO

A fim de melhor detalhar as principais discussões dos pontos de interseção entre a ciência de dados e a ciência da informação, pode-se analisar a rede de acoplamento bibliográfico dos 34 documentos anteriormente recuperados, buscando-se observar de que forma se associam os itens mais relevantes na interlocução entre a ciência de dados e a ciência da informação. Dentre esses 34 documentos, 17 são citados pelos demais e apenas nove formam um componente de rede, isto é, fazem citação de bibliografia comum. A rede é apresentada na figura 3.

Figura 3 – Mapa de acoplamento bibliográfico entre itens que combinam ciência de dados e ciência da informação ou “library science”



Fonte: Elaborada pelo autor.

A leitura e análise dos nove documentos acoplados pela bibliografia é apresentada a seguir, em busca de aprofundar e sugerir os futuros caminhos de desenvolvimento de uma interlocução entre a ciência de dados e a ciência da informação. Os artigos serão apresentados na ordem de maior centralidade na rede apresentada na figura 3.

**Twinning data science with information science in schools of library and information science**, de Wang (2018), apresenta a ciência de dados como uma nova corrente vital na educação em “library and information science”, onde a ciência de dados seria uma “irmã gêmea” da ciência da informação, com um grande número questões em comum. Wang descreve de que forma os conceitos da ciência da informação contribuiriam para aprimorar a ciência de dados.

Vinculado ao Departamento de Administração da Universidade Normal de Tianjin, na China, mas citando avanços da integração disciplinar nos EUA, Wang (2018) argumenta que as escolas de informação devem se tornar ambidestras na integração de ambas as “ciências”. Nesse sentido, trariam contribuições significativas à ciência de dados, os seguintes aspectos oriundos da ciência da informação (WANG, 2018):

- a) Concepção de dados nos sentidos históricos e pragmáticos, em complemento à visão positivista dominante em ciência de dados, que, do ponto de vista analítico, tende a considerar dados como sendo objetivos e neutros;
- b) Controle da qualidade de dados, visando promover técnicas para aumentar a precisão, acurácia e credibilidade de dados, especialmente quando tratando-se de *big data*;
- c) Organização bibliotecária de dados, com vistas a promover, em ciência de dados, o uso de técnicas e ferramentas voltadas a coleção, preservação, curadoria, governança, metadados, compartilhamento, visualização, suporte à análise, avaliação de qualidade, referenciação, citação e literacia de dados; e
- d) Estudo das teorias da documentação, com o objetivo de trazer para a ciência de dados a percepção de que ela trabalha com documentos e que documentos apresentam três dimensões essenciais (BUCKLAND, 2016): dimensão tecnológica de seu suporte físico; dimensão cognitiva da relação mental dos indivíduos com os documentos; e dimensão social, relacionada aos papéis econômicos e políticos dos documentos.

Enquanto contribuição inversa da ciência dos dados para a ciência da informação, o autor aponta para o *big data* como gerador de demandas por mudanças na ciência da informação, mas sem maiores detalhamentos.

**A cross-institutional analysis of data-related curricula in information science programmes: A focused look at the iSchools**, de Ortiz-Repiso, Greenberg e Calzada-Prado (2018), apresenta uma análise da ênfase em estudos sobre dados em 597 cursos que conferiram graus em 65 escolas do consórcio internacional das *i-Schools* (*information schools*), no período de fevereiro de 2016. O levantamento indicou que 14% dos programas analisados enfatizavam estudos sobre dados, sendo que 11% desses estudos voltavam-se ao *data science* em geral ou, mais especificamente, ao *big data analytics*; e 3% eram dirigidos à curadoria de dados digitais.

**The cultivation of scientific data specialists: Development of LIS education oriented to e-science service requirements**, de Si, Zhuang, Xing e Guo (2013), é o mais antigo de todos os documentos obtidos dentre os 34 identificados, e, portanto, tenderia a ser o mais citado, o que ficou constatado, com 18 citações. O artigo focaliza as qualificações dos profissionais que buscavam ser contratados em portais de emprego na China associadas aos termos “scientific data management”, “data service”, “data curation”, “e-Science”, “e-Research” e “data specialist”, e buscou apresentar seus resultados de forma útil ao desenho de novos currículos de formação de profissionais das *iSchools* aplicados ao campo de “scientific data management” e “e-science” no contexto daquele país. As habilidades, as qualificações, os deveres e as responsabilidades identificados como mais valorizados na contratação de profissionais para o “scientific data management” e “e-science” foram (SI; ZHUANG; XING; GUO, 2013): a) capacidade para trabalho em equipe; b) boa comunicação; c) relacionamento interpessoal; d) uso de ferramentas de curadoria de dados; e) padrões de metadados.

Si, Zhuang, Xing e Guo (2013) não analisaram demandas relacionadas a outros tipos de profissionais ligados à ciência de dados, como nas áreas de “big data analyst” ou mesmo “data scientist”.

**Educating Data Management Professionals: A Content Analysis of Job Descriptions**, de Chen e Zhang (2017), apresenta análise de qualificações desejáveis requeridas junto a profissionais de “library and information science” no início de 2015, nos EUA, voltados à função de “librarian data management”. Os resultados da pesquisa apontaram que “os candidatos mais bem sucedidos deveriam estar aptos a apoiar professores e alunos na coleção, gerenciamento e análise de dados de pesquisa”. Não foram analisadas por Chen e Zhang (2017), quaisquer demandas relacionadas a outros tipos de profissionais ligados à ciência de dados, como nas áreas de “big data analyst” ou mesmo “data scientist”.

**Research Synthesis Methods and Library and Information Science: Shared Problems, Limited Diffusion**, de Sheble (2016), buscou identificar, por meio de técnicas de pesquisa bibliográfica, as intersecções de temas de pesquisa que relacionam a “library and information science” com o uso dos métodos de RSM (*Research Synthesis Methods*), envolvendo a Revisão Sistemática e a Meta-análise. Os métodos de RSM são fortemente usados em pesquisa biomédica para a síntese de evidências. Os principais tópicos identificados na interseção entre “library and information science” e RSM foram: “open access”, “information retrieval”, “bias and research information ethics”, “referencing practices”, “citation patterns” e “data science”. A aplicabilidade de “data science” foi evidenciada em situações nas quais algoritmos de aprendizagem de máquina poderiam auxiliar na identificação automática das fontes de informação mais relevantes para leitura detalhada (SHEBLE, 2016), o que exemplifica o uso de *machine learning* para finalidades de classificação.

**Reproducible research and GIScience: an evaluation using AGILE conference papers**, de Nuest *et al.* (2018), desenvolve um conjunto de critérios para a produção de pesquisa reproduzível, fundamentado na plena documentação, disponibilização e permanência de acesso aos dados, métodos e critérios para produção de resultados em artigos científicos, com foco na produção científica na área de *GIS Science* (Sistemas de Informação Geográficos).

Foram então exploradas, junto aos autores de artigos de conferências na área de *GIS Science*, as razões pelas quais seus artigos apresentaram falhas no atendimento aos critérios de reprodutibilidade já reconhecidos. Dentre as cinco razões indicadas pelos autores, as mais relevantes para o cometimento das falhas foram, em ordem decrescente de relevância: (1) restrições legais; (2) falta de tempo; (3) falta de ferramentas; (4) falta de incentivos; e (5) falta de conhecimento.

**Geography and computers: Past, present, and future**, de Arribas-Bel e Reades (2018), explora de que forma a ciência de dados pode conduzir os estudos quantitativos em geografia para a criação de uma *Geographic Data Science*, que seria uma forma de ciência de dados consciente da espacialidade do dado geográfico, inclusive nas dimensões de crítica social e humanista presentes nos estudos geográficos. Segundo os autores, o desenvolvimento de uma *Geographic Data Science* combinaria a “tradição de *pensamento espacial* prevalente na geografia computacional e em sistemas de informação geográficos, com abordagens modernas para captura, transformação, processamento e análise evidentes na ciência de dados” (ARRIBAS-BEL; READES, 2018, p. 6), sem desconsiderar as questões epistemológicas, metodológicas e mesmo políticas que pautam o desenvolvimento histórico da geografia.

**Data Science as a new frontier for Design**, de Kazakci (2015), explora como as teorias do design – *design science, design research, design thinking* (BROWN, 2009; BIRKHOFER, 2011) – podem explicar as contribuições da ciência de dados para o aprimoramento do avanço científico na chamada *e-science*, ciência baseada em análise intensiva de dados. Para tal, o autor descreveu o caso chamado de *HiggsML Challenge* (Desafio de aprendizagem de máquina Higgs), no qual o desenvolvimento de soluções colaborativas de algoritmos de aprendizagem de máquina para análise dos dados gerados por um acelerador de partículas levou ao aprimoramento da compreensão do Bóson de Higgs.

O caso envolveu 1.785 times de todo o mundo em uma espécie de evento do tipo *hackaton* e foi considerado muito bem-sucedido, inclusive por envolver pessoas com muito pouco conhecimento de física na produção de resultados relevantes. A abertura de acesso aberto aos dados científicos *on-line*, o uso de uma plataforma *on-line* (Kaggle) para “gamificação” de todo o processo de formação dos times, o uso de repositórios de código aberto de programas de computador compartilhado em plataformas como o GitHub, a troca de mensagens em fóruns *on-line* e a exploração simultânea de várias estratégias distintas para análise de dados, algumas delas não convencionais, foram algumas das práticas marcantes que caracterizam os processos de trabalho dos cientistas de dados durante o *HiggsML Challenge*.

**Knowledge mobilization for community resilience: perspectives from data, informatics, and information science**, de Virapongse, Duerr e Metcalf (2019), apresenta resultado de uma discussão mediada com uma comunidade nos EUA que busca desenvolver um arcabouço (*framework*) de dados, métodos, ferramentas e processos que objetivam aprimorar a resiliência de comunidades frente à ocorrência de desastres naturais ou artificiais. A discussão visava identificar as principais recomendações para mobilização de informação no processo de aprimoramento da resiliência de comunidades. Dentre elas, foram elencadas e detalhadas:

- a) Esclarecer o conceito de resiliência de comunidades, a fim de permitir identificar as necessidades de dados (fatores de riscos de natureza diversa, especialmente ambientais);
- b) Priorizar suporte institucional para a resiliência de comunidades;
- c) Garantir a acessibilidade e usabilidade aos dados;
- d) Preencher os vazios no ciclo da informação, com vistas a promover a interação entre os provedores de dados e os praticantes de resiliência em comunidades.



## ANÁLISE E DISCUSSÃO

A breve exploração da atual condição da ciência de dados, sob a perspectiva bibliográfica, combinada com a exploração dos nove artigos que evidenciam algumas das interlocuções entre a ciência de dados e a ciência da informação mostra que a ciência de dados é uma área do conhecimento predominantemente envolvida com o desenvolvimento de inovações tecnológicas e com a produção de conhecimento científico de natureza quantitativa em um contexto de massiva produção de dados.

A ciência de dados é uma área aplicada, que se apropria de ferramental estatístico e computacional recentemente produzido e disponibilizado em repositórios de software aberto, como o GitHub, para produzir, em curto espaço de tempo, análises e sínteses, especialmente as de natureza preditiva, baseadas em dados de variam amplamente em volume, variedade e velocidade de geração.

As análises e sínteses produzidas de forma algorítmica pela ciência de dados geram elevado impacto sobre a decisão em negócios empresariais e sobre a compreensão de fenômenos científicos em todas as áreas do conhecimento humano.

A perspectiva epistemológica dominante na ciência de dados é positivista, quantitativa, social e historicamente ingênua, além de fundamentada no compartilhamento de ferramental coletivamente gerado e na participação em jogos sociais competitivos, isto é, “gamificados”.

No que se refere às implicações das decisões geradas pelas previsões e classificações de eventos feitas por modelos desenvolvidos em ciência dos dados, pairam dúvidas sobre a validade do que é produzido, sobretudo pela ausência de explicabilidade e rastreabilidade dos processos dedutivos e indutivos empregados em algumas situações.

A breve exploração da co-ocorrência de palavras-chave na interlocução entre a ciência de dados e a ciência da informação evidencia o desafio educacional que a ciência da informação enfrenta para, de um lado, incorporar à ciência de dados os valores e responsabilidades do profissional da informação e, de outro lado, gerar novo valor em ciência da informação a partir da oportunidade produzida pelos *big data*, com a incorporação de ferramentas da inteligência artificial e da aprendizagem de máquina. Nesse sentido, quais princípios, tendências, designs de sistemas e riscos surgem na transformação da ciência da informação por princípios epistemológicos, modelos, métodos, ferramentas e valores da ciência de dados?

No que se refere às mensagens latentes nos nove artigos mais acoplados bibliograficamente na relação entre a ciência da informação e a ciência dos dados, cabem algumas considerações finais:

- 1. A transformação paradigmática da ciência da informação** pela emergência dos *big data* é iminente e, para que essa transformação se processe em um sentido positivo, deve ser garantida a incorporação de valores e princípios da ciência da informação, tais como, segundo Wang (2018): a concepção de dados nos sentidos históricos e pragmáticos; o cuidado com a precisão, acurácia e credibilidade de dados; os processos e funções clássicos da biblioteconomia; além das teorias da documentação.
- 2. Existem distintas formas de neutralidade na ciência da informação e na ciência de dados, e cada uma apresenta aspectos positivos e negativos.** Por um lado, a ciência de dados tende a assumir uma condição de neutralidade perante dados, informações e conhecimentos à sua disposição, sejam ou não esses dados validados do ponto de vista científico. Nesse sentido, a ciência de dados promove ativamente uma postura inovadora, aderente ao *design thinking* (VIRAPONGSE; DUERR; METCALF, 2019), criando novos modelos de natureza estatística e algorítmica, produzidos para domínios de conhecimento específicos, visando à imediata geração de novo conhecimento acionável para a tomada de decisão, baseada nos recursos à disposição.

Por outro lado, a ciência da informação clássica se promove como mediadora das necessidades de informação de seus usuários (ZINS, 2007) frente à organização dos dados, de informações e conhecimentos científicos já produzidos, validados e registrados, com os quais essa ciência também se apresenta neutra, sem assumir caráter julgador. Entretanto, as práticas da ciência de dados não se apresentam como ativas promotoras de intervenções, em especial no processo decisório.

- 3. As escolas de ciência da informação devem se movimentar com celeridade compatível com a avalanche tecnológica dos big data** (ORTIZ-REPISO; GREENBERG; CALZADA-PRADO, 2018; VIRAPONGSE; DUERR; METCALF, 2019; WANG, 2018) na produção de novos currículos para oportunizar a sobrevivência e o reconhecimento dos profissionais da informação, ou mesmo a preservação dos valores profissionais e éticos em um contexto de geração de novas profissões.
- 4. Uma potencial dominância dos conceitos empregados na geografia durante a interlocução entre a ciência de dados e a ciência da informação merece exploração futura**, como sugerem os estudos de Nuest *et al.* (2018) e Arribas-Bel e Reades (2018). Ainda, com base em Nuest *et al.* (2018), o conjunto das razões pelas quais as pesquisas científicas em determinada área de conhecimento apresentam problemas com reprodutibilidade sugerem aspectos que podem ser considerados relevantes em contribuições da ciência da informação para o aprimoramento da ciência de dados e estariam relacionados ao desenvolvimento de epistemologias, metodologias, instrumentos legais, códigos de conduta profissional, incentivos mercadológicos e ferramentais críticos para aprimoramento da prática dos cientistas de dados, no sentido de melhor compreensão das dimensões cognitivas e sociais envolvidas com os dados, métodos e critérios que empregam na produção de suas análises.

- 5. As contribuições da ciência de dados para a evolução das demais áreas do conhecimento científico e suas correspondentes profissões são marcantes e distintas, sugerindo a possibilidade de que a ciência de dados se constitua como um novo paradigma científico.** Se os erros e acertos decorrentes da aplicação massiva dos métodos e técnicas da ciência de dados nas empresas para a criação de soluções de *data analytics* apresentam problemas éticos que devem ser endereçados com aprimoramentos legais, normativos e profissionais, a ciência de dados evidencia a oportunidade da emergência da *e-science*. Essa nova forma de fazer ciência é caracterizada no estudo de Kazakci (2015).
- 6. O problema da resiliência das comunidades vem a ser um arquétipo dos problemas enfrentados pela humanidade no atual momento em que nos encontramos, de pandemia causada pelo SARS-Cov-2.** Nesse sentido, o avanço da ciência cidadã é fundamentalmente baseado nas ferramentas da ciência de dados e depende do preenchimento dos imensos vazios no ciclo da informação, que foram levantados por Virapongse, Duerr e Metcalf (2019), no contexto da promoção da resiliência em comunidades, como:
  - Organizar oficinas de trabalho e tutoriais para auxiliar os praticantes de análise e síntese de dados em comunidades a acessarem dados e a usarem ferramentas para a sua análise;
  - Ensinar habilidades de colaboração, facilitação, liderança, gerenciamento e trabalho em equipe para produtores de dados;
  - Criar papéis ativos e distintivos para pessoas atuantes no ciclo da informação e do conhecimento, tais como intermediários no acesso aos dados, intérpretes de dados científicos, mediadores, promotores de engajamento em comunidades e contadores de histórias;
  - Ligar praticantes de análise e síntese de dados em comunidades às redes onde se encontram os produtores de dados;

- Identificar comunidades que precocemente adotam inovações, para intensificar colaborações e parcerias, promovendo exemplos de usos de dados e ferramentas junto às comunidades;
- Promover encontros para criação de projetos pilotos envolvendo produtores de dados e praticantes de análise e síntese de dados em comunidades;
- Criar fóruns para encontros regulares de distintas partes interessadas no ciclo da informação relativa à resiliência de comunidades;
- Promover organizações para mediação da comunicação entre produtores de dados e praticantes de análise e síntese de dados em comunidades.

As tradições da ciência da informação presentes na releitura dos conceitos de bibliotecas são extremamente pertinentes no preenchimento desses vazios.

## CONCLUSÕES

Este artigo buscou situar, do ponto de vista bibliográfico, o atual estado da ciência dos dados em sua relação com a ciência da informação. Foram empregadas pesquisas em uma base de dados bibliográfica, obtendo-se registros que foram analisados de forma visual, usando as técnicas de análise de co-ocorrência e acoplamento bibliográfico. As perspectivas de um breve entrelaçamento entre a ciência de dados e a ciência da informação são traçadas em ambas as direções, apontando não apenas para uma oportunidade de mútuo benefício das áreas e de seus praticantes, mas para uma necessidade de que isso ocorra em benefício de uma sociedade mais resiliente, sustentável, democrática e ética.

## REFERÊNCIAS

- ARRIBAS-BEL, D.; READES, J. Geography and computers: Past, present, and future. *Geography Compass*, [s.l.], v. 12, n. 10, p. e12403, Sep. 2018. Disponível em: <https://doi.org/10.1111/gec3.12403>. Acesso em: abr. 2021.
- BIRKHOFFER, H. (ed.). *The Future of Design Methodology*. London: Springer-Verlag, 2011.
- BROWN, T. *Change by Design: how design thinking transforms organizations and inspires innovation*. USA: Harper Collins, 2009.
- BUCKLAND, M. The physical, mental and social dimensions of documents. *Proceedings from the Document Academy*, Ohio, USA, v. 3, n. 1, p. 1-6, 2016. DOI <https://doi.org/10.35492/docam/3/1/4>. Disponível em: <https://ideaexchange.uakron.edu/docam/vol3/iss1/4>. Acesso em: maio 2021.
- CHEN, H.; ZHANG, Y. Educating Data Management Professionals: a content analysis of job descriptions. *Journal of Academic Librarianship*, [USA], v. 43, n. 1, p. 18–24, Jan. 2017. Disponível em: <https://doi.org/10.1016/j.acalib.2016.11.002>. Acesso em: maio 2021.
- DAVENPORT, T. H.; PATIL, D. J. Data Scientist: the sexiest job of the 21st Century. *Harvard Business Review*, [s.l.], v. 90, n. 10, p. 70–76, Oct. 2012. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>. Acesso em: maio 2021.
- KAZAKCI, A. O. Data science as a new frontier for design. In: INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, 15., 2015, Milan, Italy. *Proceedings [...]*. Glasgow: Design Soc, 2015. Disponível em: <https://www.designsociety.org/publication/37969/>