

# Modelos de ensino e aprendizagem de ciência de dados: em foco a área da Ciência da Informação

Data science teaching and learning models: focus on the Information Science area

**Dalton Lopes Martins**

Universidade de Brasília, Brasil

**E-mail:** daltonmartins@unb.br

**ORCID:** <https://orcid.org/0000-0002-6244-6791>

## RESUMO

O tema Ciência de Dados tem se tornado um foco de interesse para o ensino e a pesquisa na área da Ciência da Informação na última década. No entanto, pela especificidade da área ainda existem grandes desafios quando se trata de construir currículos para disciplinas de graduação e pós-graduação. O que se observa é que os estudantes da área apresentam dificuldades de aprendizagem quando ensinados pelas abordagens tradicionais oriundas das ciências exatas. Uma das mais importantes revistas científicas da área da Ciência da Informação internacional, a JASIST, abriu uma chamada de artigos para um número especial da revista com foco em discutir as especificidades da Ciência de Dados em suas relações com a Ciência da Informação. A chamada disponibilizou uma extensa bibliografia especializada sobre tema, de onde se ressaltam na presente pesquisa nove trabalhos diretamente indicados ou referenciados a partir dos trabalhos indicados que são focados na questão do ensino de Ciência de Dados para a área. A análise dos trabalhos aponta para uma iniciativa específica de formação de currículo de Ciência de Dados para a área da Ciência da Informação, a iniciativa *iSchool Data Science Curriculum Committee* (iDSCC), e mais 8 modelos de currículos que poderiam ser adaptados.

**Palavras-chave:** ciência de dados; ensino e aprendizagem; modelos de currículo.

---

**Como citar:** Martins, D. L. (2022). Modelos de ensino e aprendizagem de ciência de dados: em foco a área da Ciência da Informação. Em T. M. R. Dias (Ed.), *Informação, Dados e Tecnologia. Advanced Notes in Information Science, volume 2* (pp. 140-148). Tallinn, Estonia: ColNes Publishing. DOI: 10.47909/anis.978-9916-9760-3-6.100.

**Copyright:** © 2022, The author(s). This is an open access work distributed under the terms of the CC BY-NC 4.0 license which permits copying and redistributing the material in any medium or format, adapting, transforming and building upon the material as long as the license terms are followed.

## ABSTRACT

*The Data Science field has become a focus of interest for teaching and research in the Information Science area in the last decade. However, due to the specificity of the area, there are still significant challenges when it comes to building curricula for undergraduate and graduate disciplines. What is observed is that students in the area have learning difficulties when taught by traditional approaches from the hard sciences. One of the most important international scientific journals in Information Science, the Journal of the American Society for Information Science and Technology (JASIST), opened a call for articles for a special issue of the journal with a focus on discussing the specificities of Data Science in its relations with Information Science. The call made available an extensive specialized bibliography on the subject. Nine works stand out in this research, directly indicative or referenced from the works indicated that are focused on the issue of teaching Data Science for the area. The analysis of the results points to a specific initiative to form a Data Science curriculum for the Information Science area, the School Data Science Curriculum Committee (iDSCC) initiative, and seven other curriculum models that could be adapted.*

**Keywords:** data science; teaching and learning; curricula models

## 1. INTRODUÇÃO

A CIÊNCIA de Dados (CD) tem se tornado um tema de grande relevância para a área da Ciência da Informação (CI) em termos mundiais. Seja por suas perspectivas técnicas e operacionais, seja por suas perspectivas metodológicas, a CD parece apoiar no desenvolvimento de novas abordagens de pesquisas para problemas característicos da área da CI. Wang (2018) define a CD como sendo um campo interdisciplinar que consiste na identificação e extração de valor dos dados, convertendo dados em informação e conhecimento a partir de sua análise e mineração. Vale aqui ressaltar a importância da definição ao destacar o processo de conversão de dados em informação, tema altamente central e relevante para a área da CI. Desse modo, compreender de forma mais sistemática os fatores que envolvem esse processo de conversão de dados em informação, as contribuições que a CD pode aportar para a CI e de que maneira isso pode ser incorporado nos processos de formação e pesquisa da área tornam-se elementos de interesse para os pesquisadores e ainda pouco compreendidos em suas especificidades.

No Brasil, a CD no contexto da CI ainda parece estar em sua fase inicial de apropriação pelos pesquisadores da área. Em busca recente na base de dados BRAPCI<sup>1</sup> com a expressão “ciência de dados”, pode se

<sup>1</sup> [https://brapci.inf.br/?q=%22Ci%C3%Aancia+de+Dados%22&type=1&year\\_s=1972&year\\_e=2021&order=0](https://brapci.inf.br/?q=%22Ci%C3%Aancia+de+Dados%22&type=1&year_s=1972&year_e=2021&order=0). Acesso em 29/08/2021.

encontrar apenas 20 trabalhos indexados. Em torno de 95% dos trabalhos foram publicados do ano de 2017 para cá, mostrando que não apenas a produção ainda é tímida como bastante recente. Sabe-se que outros termos correlatos podem ser usados para se compreender a CD no contexto da CI, tais como “curadoria digital”, “bibliotecário de dados”, entre outros. No entanto, o presente trabalho se mantém com foco específico na CD em sua concepção original com interesse de compreender suas relações com a CI.

Preocupados em entender como o tema da CD tem sido compreendido pela CI a partir de sua produção científica, Virkus e Garoufallu (2019) realizaram um estudo bibliométrico na base de dados *Web of Science* com a expressão “data science” no intervalo de tempo de 1980 a 2019. Foram identificadas 2350 publicações, sendo que apenas 3,4% (80) delas são relativas à área da CI. No estudo, os pesquisadores apontaram as 3 principais revistas científicas onde a maior dos trabalhos foi publicada pelos pesquisadores da CI. São elas: *Journal of the American Medical Informatics Association*, *International Journal of Information Management* e *Journal of the Association for Information Science and Technology (JASIST)*. Dando continuidade à pesquisa, agora em busca de avaliar tematicamente a produção científica da área, Virkus e Garoufallu (2020) classificaram tematicamente os 80 trabalhos da área da CI. Os trabalhos foram divididos em 6 categorias principais: (1) educação e treinamento em CD, (2) conhecimentos e habilidades do profissional de dados, (3) o papel das bibliotecas e dos bibliotecários no movimento da CD, (4) ferramentas, técnicas e aplicações de CD, (5) CD na perspectiva da gestão do conhecimento e (6) CD na perspectiva das ciências da saúde.

Destacam-se, para o foco da presente pesquisa, dos temas apresentados acima o (1), (2) e (3). Os 3 temas são voltados para definir aspectos que buscam contextualizar a CD em relação a CI, seja em busca de propor e analisar o impacto de currículos para educação e treinamento de profissionais e pesquisadores, seja no sentido de definir quem é esse profissional e que conhecimentos e habilidades ele deveria ter para lidar com o novo cenário que se apresenta ou, por fim, no sentido de ressignificar explicitamente o papel das bibliotecas e dos bibliotecários frente ao movimento da CD. Os temas revelam que a CD é ainda bastante nova no contexto da CI e que a compreensão de suas especificidades ainda é um problema de pesquisa em aberto suscitando discussão, experiências e análises de propostas em andamento. O desafio de discutir isso no contexto brasileiro carece de discussões que levem em consideração a realidade da área da CI no Brasil, o perfil de seus estudantes, dos seus docentes e as possibilidades de produção de currículos e experiências educacionais que dialoguem com essa realidade.

## 2. OBJETIVOS

A presente pesquisa consiste nos resultados iniciais de uma iniciativa do Laboratório de Inteligência de Redes da Faculdade de Ciência da Informação da Universidade de Brasília. O laboratório tem realizado nos últimos 4 anos diversas pesquisas em níveis de iniciação científica, trabalhos de conclusão de curso, mestrado e doutorado que envolvem temáticas diretamente relacionadas ao campo da CD no contexto da CI. Como maneira de formar estudantes, sobretudo dos cursos de graduação da Biblioteconomia e Museologia, para atuarem em pesquisas em andamento, foi desenvolvida uma disciplina optativa denominada “Introdução a Ciência de Dados para Biblioteconomia”.

A disciplina se encontra em sua quinta edição no segundo semestre de 2021. Para a preparação da disciplina, pesquisadores do laboratório têm estudado modelos de currículo voltados para a formação de CD no contexto de estudantes da CI. Desse modo, o objetivo da presente pesquisa é apresentar os modelos de currículo que estão em análise pelos pesquisadores do laboratório. Espera-se, desse modo, estimular o debate em nível nacional sobre o tema ampliando a interlocução entre os demais pesquisadores da CI que estão produzindo experiências similares e ainda em estágios iniciais de elaboração.

## 3. PROCEDIMENTOS METODOLÓGICOS

Para a realização da presente pesquisa, optou-se por um procedimento bibliométrico realizado a partir das referências disponibilizadas no texto de uma chamada<sup>2</sup> da revista JASIST para um número especial com foco a discutir a CD no contexto da CI. A revista é considerada uma das 3 mais importantes (VIRKUS & GARAFLOU, 2019) pelos pesquisadores da CI em âmbito internacional para publicação de suas pesquisas relacionadas a CD. O número especial é denominado “*Data Science in the iField*” e apresenta um conjunto de referências como forma de estimular a pesquisa e dar um contexto mais preciso do que a revista espera de contribuições para o número. Foram indicadas 26 referências no corpo da chamada e uma lista de tópicos de interesse.

Um dos tópicos de interesse refere-se explicitamente a educação e modelos de currículo. Nesse contexto, são indicadas 11 referências (42%) que tratam especificamente de trabalhos que discutem experiências de ensino, modelos de currículo e perfis dos profissionais da CI no contexto das inovações e possibilidades ofertadas pela área da CD. Considerando

<sup>2</sup> <https://www.asist.org/2021/02/16/cfp-2-21/>. Acesso em 29/08/2021.

a importância da revista para área no Brasil (a revista é Qualis A1 na área da CI na classificação atualmente vigente) bem como sua importância internacional (a revista é a 15 mais citada de um total de 235 revistas indexadas da área da CI na base Scopus),<sup>3</sup> inferiu-se para a presente pesquisa que a curadoria de referências apresentada na chamada do número especial poderia funcionar como um ponto de partida altamente especializado para se obter um panorama do estado da arte de referências a respeito das iniciativas de educação e modelos de currículo de CD no contexto da CI.

A presente pesquisa adotou o seguinte procedimento metodológico. Inicialmente, foram analisadas de forma exploratória as 11 referências apresentadas consideradas específicas para o tema da pesquisa. Delas, foram identificadas novas referências de potencial interesse para a pesquisa, obtendo-se mais 3 trabalhos analisados. De todos os trabalhos, foram identificadas iniciativas que mencionavam explicitamente modelos de currículo para a formação de CD no contexto da CI ou consideradas adaptáveis para o contexto da CI pelos pesquisadores. Apenas foram citadas na presente pesquisa as referências consideradas relevantes no apontamento de modelos de currículos ou questões específicas relativas a CI.

As iniciativas são apresentadas a seguir com comentários específicos sobre cada uma delas.

#### 4. RESULTADOS

Sha *et al.* (2021) discutem a especificidade da CD no contexto da CI, ressaltando os aspectos humanos que devem ser levados em consideração com destaque para os aspectos éticos e sociais dos profissionais da informação. Os pesquisadores discutem a problemática dos vieses que os processos de CD podem gerar quando operados a partir de bases de dados de má qualidade ou em projetos pouco preocupados com os aspectos curatoriais e inclusivos dos dados. Defendem que esses aspectos humanos devem fazer parte de currículos que tratem da CD no âmbito da CI.

Oh *et al.* (2019) apresentam uma iniciativa denominada *iSchool Data Science Curriculum Committee* (iDSCC). A iniciativa foi criada em março de 2019 pelos diretores das *iSchools* no contexto dos EUA com 3 objetivos principais: coletar e sistematizar informação sobre os cursos de CD oferecidos pelas *iSchools*, analisar a informação recebida com o objetivo de propor um currículo comum que possa servir de referência para as escolas e, por fim, propor aspectos chave da formação que possam ser usados para processos de certificação. Vale ressaltar que até o momento

<sup>3</sup> <https://www.scopus.com/sourceid/21100307484>. Acesso em 28/08/2021.

de realização desta pesquisa não foi publicado pela iniciativa um modelo de currículo para as *iSchools*. No painel reportado no texto, foram apresentados alguns modelos de currículos que estão sendo estudados no âmbito da iniciativa. São eles: *ACM<sup>4</sup> Computing Competencies for Undergraduate Data Science Curricula*, *Park City Models<sup>5</sup> Curriculum Guidelines for Undergraduate Programs in Data Science*, *EDISON Data Science Framework<sup>6</sup>*, *National Academy's DS for UG programs<sup>7</sup>*, *BHEA* e *INFORMS<sup>8</sup>*.

Hagen *et al.* (2019) e Hagen *et al.* (2020) apresentam duas iniciativas que ocorrem na iConference de 2019 e 2020. Em 2019, o painel teve como objetivo avançar na criação de uma comunidade de interessados na produção de currículos de CD, discutir os currículos, melhores práticas e o ensino de CD no nível de cursos de graduação e propor a criação de um currículo mínimo para as *iSchools*. Já em 2020, os objetivos eram basicamente os mesmos, mas o painel contou com a apresentação de algumas iniciativas. Bishop *et al.* (2019) apresentam a iniciativa do *Institute of Museum and Library Services* (IMLS) no âmbito dos EUA para a proposição de um modelo curricular<sup>9</sup> para a formação de profissionais da informação na área da CD.

Outras duas iniciativas de modelos de currículo foram mencionadas como referências nos trabalhos analisados. No entanto, elas não possuem relação direta com a área da CI, mas sim uma preocupação em oferecer modelos de currículo voltados para estudantes do equivalente ao ensino médio no Brasil. São elas: *International Data Science in Schools Project* (IDSSP),<sup>10</sup> desenvolvida por estatísticos australianos e apoiada pela Sociedade de Estatística da Austrália, e *Introduction to Data Science* (IDS),<sup>11</sup> desenvolvida pela Universidade da Califórnia em Los Angeles (UCLA). Esse tipo de iniciativa pode ser de especial interesse para os estudantes de graduação da área da CI considerando que atuam em aspectos básicos de matemática e estatística que geralmente não são tratados por disciplinas dos currículos específicos dos cursos.

<sup>4</sup> <https://dstf.acm.org/DSReportDraft2Full.pdf>. Acesso em 28/08/2019.

<sup>5</sup> <https://www.amstat.org/asa/files/pdfs/EDU-DataScienceGuidelines.pdf>. Acesso em 28/08/2021.

<sup>6</sup> <https://edison-project.eu/edison/edison-data-science-framework-edsf/>. Acesso em 28/08/2021.

<sup>7</sup> <https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options>. Acesso em 28/08/2021.

<sup>8</sup> <http://info.informs.org/abok>. Acesso em 28/08/2021.

<sup>9</sup> <https://www.imls.gov/sites/default/files/grants/re-70-17-0094-17/proposals/re-70-17-0094-17-full-proposal-documents.pdf>. Acesso em 28/08/2021.

<sup>10</sup> [http://www.idssp.org/files/IDSSP\\_Frameworks\\_1.0.pdf](http://www.idssp.org/files/IDSSP_Frameworks_1.0.pdf). Acesso em 28/08/2021.

<sup>11</sup> <https://www.idsucla.org/introduction-to-data-science-curriculum>. Acesso em 28/08/2021.

No total, os trabalhos revisados mencionaram 9 iniciativas de construção de currículos específicos para a formação de estudantes de graduação em CD. No entanto, apenas uma das iniciativas apresentadas foi construída especificamente levando em consideração o perfil do estudante da área da CI, a iniciativa promovida pelo IMLS. Compreende-se que outras iniciativas específicas para a área da CI ainda se encontram em andamento, caso da iDSCC, onde os primeiros movimentos em direção a tal tarefa datam de março de 2019 no âmbito do evento anual ASIS&T 2019. No contexto da presente pesquisa, não cabe uma análise detalhada dos modelos propostos, algo a ser realizado em pesquisas futuras. No entanto, apenas a título de uma visão inicial, apresenta-se na tabela 1 o ano de publicação das 9 iniciativas e o tamanho de páginas do documento de cada um dos modelos de currículos.

INICIATIVA	ANO PUBLICAÇÃO	NÚMERO DE PÁGINAS
ACM4 Computing Competencies for Undergraduate Data Science Curricula	2019	126
Park City Models <sup>5</sup> Curriculum Guidelines for Undergraduate Programs in Data Science	2017	16
EDISOM Data Science Framework	2017	74
National Academy's DS for UG programs	2018	138
BHEA	Não disponível	Não disponível
INFORMS	2018	400
IMLS – LIS Education and Data Science-for-the National Digital Platform	2018	18
Introduction do Data Science Curriculum	2014	403
Curriculum Frameworks for Introductory Data Science IDSSP	2019	96

**Tabela 1.** Ano de publicação e número de páginas das iniciativas de modelos de currículos de CD. **Fonte:** elaborada para a pesquisa.

Observa-se que a grande maioria das iniciativas é bastante recente, datando do ano de 2017 para cá. Apenas uma é mais antiga, tendo sua primeira versão publicada em 2014, sendo uma das iniciativas voltada para o público do ensino médio. Há uma grande variação no tamanho das propostas, oscilando de um mínimo de 16 páginas a um máximo de 403 páginas. Pode-se inferir que há uma grande variação no grau de detalhamento das propostas curriculares, sendo algumas bem detalhadas em termos dos objetivos educacionais, conteúdos propostos e atividades

sugeridas e outras mais genéricas apenas apresentando grandes blocos temáticos de conhecimento e sua relação entre si.

## 5. CONSIDERAÇÕES FINAIS

Os resultados iniciais da pesquisa permitiram identificar 9 modelos de currículos de CD. Desses, 7 iniciativas são voltadas para estudantes de graduação e 2 iniciativas para estudantes do ensino médio. Apenas uma das iniciativas tem foco específico para estudantes e profissionais da área da CI. Os resultados fornecem uma base de currículos que podem ser estudados detalhadamente em pesquisas futuras, onde se pretende detalhar as abordagens pedagógicas propostas, os conteúdos sugeridos, tipos de atividades, tecnologias e metodologias educacionais utilizadas.

Entende-se que o trabalho de desenvolvimento de currículos específicos de CD para a área da CI se encontra ainda em fase inicial. Espera-se que o conhecimento de tais iniciativas no Brasil possa auxiliar no debate e na promoção de novas iniciativas, ampliando o debate e a interlocução entre os pesquisadores, favorecendo a construção de currículos que possam melhorar a formação de estudantes da área de CI no campo da CD, eventualmente ampliando seu mercado de trabalho e áreas de atuação em pesquisa.

## CONFLITOS DE INTERESSE

O autor declara não haver conflito de interesses.

## DECLARAÇÃO DE CONSENTIMENTO DE DADOS

Nenhum dado foi gerado durante o desenvolvimento deste estudo. 

## REFERÊNCIAS

- BISHOP, B., ALLARD, S., BENEDICT, K., GREENBERG, J., HOEBELHEINRICH, N., LIN, X., & WILSON, B. (2019, September 24-26). *Curricula models and resources along the data continuum: Lessons learned in the development and delivery of research data management and data science education*. [Conference panel]. ALISE 2019, Knoxville, Tennessee, U.S.A.
- HAGEN, L., ZAMIR, M., ANDREWS, J., & HAMERLY, D. (2019, March 31 - April 3). *Undergraduate data science education in iSchools: Current practices and future directions*. [Conference panel presentation]. iConference 2019, Washington DC, United States.

- HAGEN, L., M. ABDUL-MUHAMMAD, M., ANDREWS, J., ZAMIR, H., HAMERLY, D., & CLIFFORD-BOVA, S. (2020, March 23–26). *Undergraduate data science education in iSchools: Optics and politics* [Conference panel presentation]. iConference 2020.
- OH, S., SONG, I.-Y., MOSTAFA, J., ZHANG, Y., & WU, D. (2019, October 19-23). *Data science education in the iSchool context* [Conference presentation]. 82nd Annual Meeting, ASIS&T 2019, Melbourne, Australia.
- SHAH, C. ANDERSON, T., HAGEN, L., & ZHANG, Y. (2021). An iSchool approach to data science: Human-centered, socially responsible, and context-driven – A position paper. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24444>
- VIRKUS, S. & GAROUFALLOU, E. (2019). Data science from a library and information science perspective. *Data Technologies and Applications*, 53:4, 422-441. <https://doi.org/10.1108/DTA-05-2019-0076>
- VIRKUS, S. & GAROUFALLOU, E. (2020). Data science and its relationship to library and information science: a content analysis. *Data Technologies and Applications*, 54(5), 643-663. <https://doi.org/10.1108/DTA-07-2020-0167>
- WANG, L. (2018), Twinning data science with information science in schools of library and information science, *Journal of Documentation*, 74(6), 1243-1257. <https://doi.org/10.1108/JD-02-2018-0036>

