

## Ciência de Dados na Graduação em Arquivologia

**Henrique Monteiro** Universidade Federal do Espírito Santo, Espírito Santo, Brasil

<https://orcid.org/0000-0003-2011-7022>

**Cristóvão** henrique.cristovao@ufes.br

**Luciana Itida Ferrari** Universidade Federal do Espírito Santo, Espírito Santo, Brasil

<https://orcid.org/0000-0002-8223-039X>

luciana.ferrari@ufes.br

**Margarete Farias de Moraes** Universidade Federal do Espírito Santo, Espírito Santo, Brasil

<https://orcid.org/0000-0002-4110-4610>

margarete.moraes@ufes.br

**Resumo** Esta pesquisa apresenta um relato de experiência sobre ofertas da disciplina Ciência de Dados na graduação em Arquivologia. Tem como objetivos discutir a importância da disciplina Ciência de Dados para a formação do arquivista, e apresentar uma proposta de disciplina relatando procedimentos pedagógicos adotados em experiências de ofertas de turmas. Possui abordagem qualitativa, natureza aplicada, objetivo exploratório-descritivo, com relato de experiência realizado por observação participante e pesquisa documental. Como resultado, apresenta uma proposta de ementa da disciplina Ciência de Dados, bem como os objetivos, conteúdo, metodologia e bibliografia. Apresenta também a experiência de aplicação da metodologia de ensino-aprendizagem PBL, caracterizada pelo uso de problemas da vida real para estimular o desenvolvimento do pensamento crítico e das habilidades de solução de problemas. Conclui-se que a disciplina Ciência de Dados, modelada para a formação do arquivista, tem grande potencial de contribuição para a sua formação, possibilitando a compreensão de aplicações em outras áreas do conhecimento por meio da integração com alunos de outros cursos, além de fomentar motivação e engajamento do aluno no curso de graduação em Arquivologia.

**Palavras-chave** Ciência de Dados. Arquivologia. Arquivista. Arquivista de dados. Aprendizagem baseada em problemas.

## Data Science in Archival Science Undergraduate Course

**Abstract** This research presents an experience report on offers of the Data Science discipline in the Archival Science undergraduate course. It aims to discuss the importance of the Data Science discipline for the training of archivists, and to present a subject proposal reporting pedagogical procedures adopted in experiences of offering classes. It has a qualitative approach, applied nature, exploratory-descriptive objective, with experience report carried out by participant observation and documentary research. As a result, it presents a proposal for a syllabus of the Data Science discipline, as well as the objectives, content, methodology and bibliography. It also presents the experience of applying the PBL teaching-learning methodology, characterized by the use of real-life problems to stimulate the development of critical thinking and problem-solving skills. It is concluded that the Data Science discipline, modeled for the training of archivists, has great potential to contribute to their training, enabling the understanding of applications in other areas of knowledge through integration with students from other courses, in addition to promoting student motivation and engagement in the undergraduate course in Archival Science.

**Keywords** *Data Science. Archival science. Archivist. Data Archivist. Problem-based learning.*



Licença de Atribuição BY do Creative Commons  
<https://creativecommons.org/licenses/by/4.0/>

Submetido em 18/07/2022

Aprovado em 26/09/2022

Publicado em 01/11/2022

## 1 INTRODUÇÃO

Ciência de Dados (CD) é um campo interdisciplinar que visa transformar dados em valor real (AALST, 2016) e possui diversos fenômenos, como *data driven*, *datafication* e *dataism*, que já fazem parte da atual sociedade da informação. Espera-se que o arquivista dos tempos atuais tenha compreensão desses fenômenos e saiba aplicá-los conforme as competências inerentes à sua área de estudo, na transformação de dados para agregação de valor. Espera-se também que ele atue na concepção de informação estratégica demandada por administradores de organizações para auxiliar em decisões (BELLOTTO, 2004), desempenhando as funções necessárias para uma gestão bem-sucedida em coleções de dados digitais de longa duração (NATIONAL SCIENCE FOUNDATION, 2005).

Atuando em equipes interdisciplinares, o arquivista pode gerir dados, classificando-os em categorias, utilizando-se de metadados adequados para garantir o seu registro preciso, a proveniência e a organicidade com foco na melhoria de sua recuperação e na manutenção de sua autenticidade. Nesse contexto, Madeiro e Dias (2020) sinalizam a demanda pelo profissional denominado de arquivista de dados, que lida com o gerenciamento, arquivamento, preservação e reuso de dados e metadados de pesquisa, sendo necessárias habilidades e competências analíticas, interpessoais e tecnológicas em um mundo orientado a dados.

Segundo Martins (2022), a CD é muito nova no contexto da Ciência da Informação (CI) e a "[...] compreensão de suas especificidades ainda é um problema de pesquisa em aberto suscitando discussão, experiências e análise de propostas em andamento". O autor complementa que, o ensino da CD no Brasil e no contexto da CI, ainda parece estar em sua fase inicial de apropriação pelos pesquisadores e professores da área, apesar da importância desse tipo de ensino para melhoria da formação dos alunos. Especificamente sobre a formação de arquivistas no Brasil, tanto Oliveira (2014) quanto Chagas, Negreiros e Silva (2021) sinalizam que se deve atender às demandas do mercado de trabalho adequando-se às constantes variações percebidas na sociedade. Além disso, o Comitê Gestor da Internet no Brasil fez uma pesquisa em 2018 sobre o uso das tecnologias de informação e comunicação em espaços informacionais brasileiros e constatou que 77% dos arquivos digitalizam parte do acervo e apenas 65% disponibiliza consulta para o público (CETIC.BR, 2019).

Com foco no atendimento às demandas, recomendações de pesquisas e percepções sobre possibilidades de melhorias, foi feita uma reformulação na matriz curricular do curso de graduação em Arquivologia da Universidade Federal do Espírito Santo (UFES) e implantada no contexto do

novo Projeto Pedagógico do Curso (UFES, 2017) a partir do segundo semestre letivo de 2017. Dentre as mudanças, foi incluída a disciplina optativa CD. Observada posteriormente, a nova matriz curricular "[...] atendeu muitas demandas do mercado de trabalho, dos egressos, dos discentes e dos docentes da instituição, mas de forma especial, trouxe direcionamentos para estudar o documento arquivístico na era digital" (GAVA; FERRARI; MORAES, 2019).

Os objetivos do presente relato de experiência são (i) discutir a importância da disciplina CD para a formação do arquivista, (ii) apresentar uma proposta de programa de disciplina para o curso de graduação em Arquivologia e (iii) relatar alguns procedimentos pedagógicos adotados em experiências de ofertas de turmas da disciplina no curso de graduação em Arquivologia da UFES até o semestre letivo de 2022/1.

A pesquisa tem abordagem qualitativa, natureza aplicada e objetivo exploratório-descritivo. Houve um levantamento bibliográfico atualizado para a temática 'arquivista de dados' e possíveis intersecções da formação em Arquivologia e da prática do arquivista com a CD nas principais bases de publicações científicas. O relato sobre a disciplina foi feito por observação participante e pesquisa documental.

O artigo apresenta duas seções para embasamento da área CD destacando possíveis relações com a atuação do arquivista. As justificativas para os objetivos descritos são apresentadas e exploradas na terceira seção do artigo, onde é tratado o papel do arquivista na CD e, conseqüentemente, a importância do ensino de elementos dessa área na formação do arquivista. A quarta seção apresenta e discute a oferta da disciplina CD para o curso de graduação em Arquivologia detalhando ementa, objetivos, conteúdo e metodologia. Por fim, apresentam-se as considerações finais.

## 2 CIÊNCIA DE DADOS

Grande quantidade de dados são gerados a cada minuto em cliques de anúncios, curtidas em mídias sociais, compartilhamentos, navegações, transações, conteúdo de *streaming* etc., como mostra o site DATA NEVER SLEEPS<sup>1</sup>. As organizações estão cada vez mais direcionando esforços para aproveitar, em suas atividades, as potencialidades advindas do uso de dados coletados. Esse fenômeno é denominado de *Data Driven* (orientado a dados). Contudo, há necessidade que eles estejam disponíveis e acessíveis, conforme orientam os princípios FAIR (WILKINSON et al, 2016). Segundo Mason e Patil (2015), a democratização dos dados é uma das ideias mais poderosas da

---

<sup>1</sup> Disponível em: <https://www.domo.com/learn/infographic/data-never-sleeps-8/>.

CD e, segundo o autor, os colaboradores de uma organização deveriam ter acesso ao máximo de dados desde que, é claro, legalmente possível.

Segundo Mayer-Schonberger e Cukier (2013), *datafy* (datafocar), é uma ação aplicada a um fenômeno para colocá-lo em um formato quantificado com o objetivo de ser tabulado e analisado. *Datafication* (dataficação) é uma tendência atual e tecnológica capaz de transformar aspectos da vida cotidiana em dados que podem ser transformados em valores e insights. Apesar do termo *datafication* ser antigo, Mayer-Schonberger e Cukier (2013) o contextualizaram à era *Big Data* explorando as possibilidades computacionais com velocidade de processamento e ferramentas para análise preditiva. *Dataism* (dataísmo), foi primeiramente descrito por David Brooks em 2013 como uma vertente filosófica para denotar que "[...] tudo que pode ser medido deve ser medido, que os dados são uma lente transparente e confiável que nos permite filtrar o emocionalismo e a ideologia, e que esses dados nos ajudarão a fazer coisas notáveis, como prever o futuro" (BROOKS, 2013, tradução nossa). O autor ainda destaca que os dados podem iluminar padrões de comportamento que ainda não se consegue observar. Diferentemente dos computadores, nossos cérebros geralmente não percebem padrões verbais sutis e a revolução dos dados está nos dando maneiras maravilhosas de entender o presente e o passado (BROOKS, 2013). Harari (2016) complementa afirmando que, para o *dataism*, todas as estruturas políticas, sociais e econômicas podem ser interpretadas como sistemas de processamento de dados e, portanto, podem ser mensuradas e registradas como dados.

Nesse contexto, a CD é uma área interdisciplinar ampla e com um campo de atuação também amplo que pode responder a perguntas que ainda não foram formuladas, diferentemente de outras áreas conhecidas como BI (*business intelligence*) que tem foco em responder perguntas previamente estabelecidas. Ela também pode trabalhar com bases de dados criadas e alimentadas para atender a propósitos já pré-estabelecidos, como feito em análises estatísticas, ou em bases que foram criadas para outro propósito, como em processos de mineração de dados.

A CD visa transformar dados (estruturados ou não estruturados, grandes ou pequenos, estáticos ou *streaming*) em valor real (na forma de previsões, decisões automatizadas, modelos baseados em dados ou visualizações de dados que ofereçam informações úteis) (AALST, 2016). Segundo Aalst (2016), a CD lida com: (i) extração de dados, (ii) preparação de dados, (iii) exploração de dados, (iv) transformação, armazenamento e recuperação de dados, (v) infraestruturas de computação, (vi) vários tipos de mineração e aprendizado de máquina, (vii) apresentação de explicações e previsões e (viii) exploração de resultados. Aalst ainda chama

atenção de que é necessário sempre levar em consideração questões éticas, sociais, legais e os aspectos inerentes ao negócio.

Existe grande quantidade de metodologias e técnicas usadas na CD. A descoberta de conhecimento, conhecida na literatura mundial por KDD (*Knowledge Discovery in Databases*) é um campo de pesquisa da inteligência artificial e que envolve a *data mining* (mineração de dados) (HAND et al., 2001). Fayyad (1996) propôs fases para o processo de descoberta de conhecimento (Figura 2 na seção 4.4): (i) **Selection** - seleção da base de dados e identificação de possíveis problemas e/ou objetivo da análise; (ii) **Preprocessing** - pré-processamento dos dados, preparação limpeza e ajustes diversos; (iii) **Transformation** - transformação dos dados para uma base que seja aceita pelos algoritmos de data mining; (iv) **Data Mining** - mineração de dados pela aplicação dos algoritmos escolhidos conforme o contexto, problema e dados para extrair padrões e relacionamentos; (v) **Interpretation/Evaluation** - interpretação e avaliação dos resultados obtidos pela mineração de dados. Existe uma fronteira tênue nos limites dessas fases como, por exemplo, Hand et al. (2001) sinalizam que a fase da transformação de dados (*transformation*) é uma parte intrínseca da fase da mineração de dados (*data mining*).

### 3 O ARQUIVISTA NA CIÊNCIA DE DADOS

Entre as atribuições da profissão de arquivista, regulamentada pela Lei nº 6.546 (BRASIL, 1978), no art. 2º, destaca-se a "orientação do planejamento da automação aplicada aos arquivos". Essa automação aplicada aos arquivos, já citada na década de 70 pela Lei, toma agora uma dimensão muito ampla contemplando fenômenos contemporâneos de transformação digital.

Nas novas formulações para Arquivologia, segundo Cook (2012), entre as "[...] três partes que compõem qualquer documento arquivístico – a estrutura, o conteúdo e o contexto [...]" podem agora ser divididas em conjuntos de dados separados onde um documento deixa de ser um "[...] objeto físico para virar um 'objeto' conceitual de informação, controlado por metadados, que virtualmente combina conteúdo, contexto e estrutura para fornecer evidências de atividade ou função de algum criador".

Tratar documentos<sup>2</sup> como dados abre o caminho para o emprego de ferramentas analíticas poderosas para permitir novos modos de investigação e pesquisa, mas, por outro lado, tratar os

---

<sup>2</sup> Nessa pesquisa, o termo '*records*', encontrado nas referências de língua inglesa, foi traduzido como 'documentos'. Apesar da recomendação do glossário do CONARQ (2020) de traduzir '*records*' como 'documento arquivístico', existem referências que não o tratam dessa forma. Por exemplo, os autores Yeo e Lowry (2020) alertam que há diferenças terminológicas nos padrões europeu e americano e que '*records*' não é o mesmo que '*archival document*' (documento arquivístico). Luciana Duranti, em 1989, já alertava para o fato de que essa equivalência ainda era uma questão em aberto e era necessário ser cauteloso (DURANTI, 1989).

dados como documentos remete a ideia de que "[...] os dados são moldados por seus contextos culturais e que o uso e a compreensão efetivos deles só serão possíveis se o conhecimento de seus contextos for salvaguardado" (YEO; LOWRY, 2020, tradução nossa). De qualquer forma, é importante ressaltar que os dados só podem ser gerenciados se forem registrados de alguma forma, ou seja, constar em documentos (HJØRLAND, 2018). Embora Furner (2016) destaque que documentos não são compostos de dados e um documento não é uma espécie de conjunto de dados, mas o contrário: "[...] um conjunto de dados é feito de documentos; e o conjunto de dados é uma espécie de documento" (tradução nossa). Nessa mesma linha, Oliveira (2015) afirma que o conjunto de tabelas de um banco de dados, que registra e apoia uma atividade-meio ou fim desenvolvida por um órgão, é um documento arquivístico. Yeo e Lowry (2020) chamam atenção para o fato de que quando dados persistem de forma estável além de seu momento de criação, eles passam a ter características de documentos.

Sant'Ana (2016) apresenta, como principais fases do ciclo de vida dos dados em uma perspectiva da CI, a coleta, o armazenamento, a recuperação e o descarte. Para cada uma dessas fases o autor ainda destaca a presença dos fatores: privacidade, integração, qualidade, direitos autorais, disseminação e preservação. Segundo Sant'Ana, na fase da recuperação de dados tem-se esforços voltados para verificar como os dados podem ser encontrados, acessados e interpretados. Dessa forma, o arquivista colabora diretamente na organização dos documentos considerando necessidades de recuperação, tendo como base os seus dados e metadados constituintes. Além disso, a organização adequada dos documentos e seus respectivos metadados desde a coleta impacta diretamente no fator de qualidade, trazendo confiança no uso dos dados ali contidos.

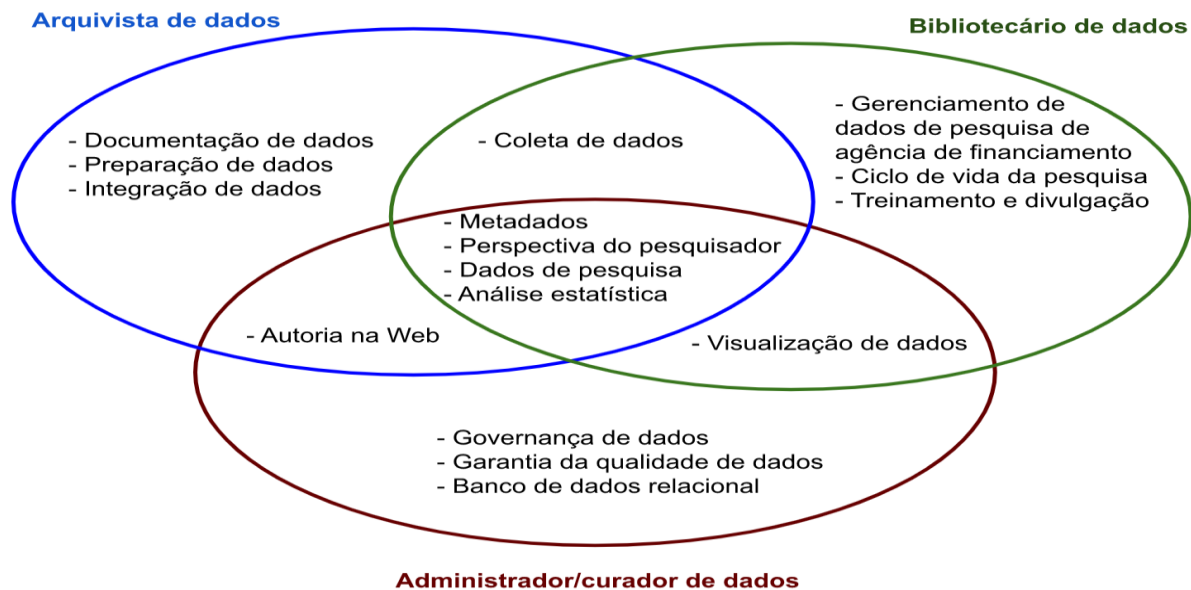
Quanto aos fatores privacidade, direito autoral, disseminação e preservação, citados por Sant'Ana (2016), pode-se inferir sobre possíveis intersecções com a prática do arquivista. O fator privacidade, necessário desde a fase da coleta, traz preocupações sobre a identificação de dados que poderiam levar à quebra da privacidade de pessoas ou instituições e, nesse sentido, o arquivista poderia fazer um tratamento dos dados quanto a anonimização ou pseudonimização conforme estabelece a LGPD (BRASIL, 2018) utilizando-se de ferramentas de preparação de dados da CD e, assim, evitar a solução simplória, feita por muitos pesquisadores, de eliminar todos os dados com intuito de proteger os participantes, conforme alerta Humphrey et al. (2000).

O fator direito autoral requer a identificação da autoria dos dados por meio de metadados adequados a fim de respeitar os elementos jurídicos vigentes, principalmente quanto à

disseminação pública. O fator disseminação requer do arquivista uma formatação de modelagem dos dados e metadados dos documentos para uso de ferramentas específicas de software para as áreas de *information and knowledge visualization*. O fator preservação pode exigir do arquivista uma descrição dos dados por meio de metadados adequados para uma organização e posterior recuperação com melhores níveis de revocação e precisão.

Lyon et al. (2015) fizeram um estudo sobre os requisitos e funções solicitadas para a contratação de *data archivist* (arquivista de dados), *data librarian* (bibliotecário de dados) e *data steward/curator* (administrador/curador de dados) nos Estados Unidos, cuja síntese é representada na Figura 1. Para o cargo de Arquivista de Dados houve ênfase em documentação de dados, preparação de dados e integração de dados, "[...] que agregam peso à afirmação de que os princípios arquivísticos estabelecidos estão refletidos na linguagem que é aplicada a novos objetos digitais de registro, ou seja, conjuntos de dados de pesquisa" (LYON et al., 2015, tradução nossa). A coleta de dados é uma função que também aparece para o bibliotecário de dados, enquanto a autoria na Web, que lida com as questões de direitos autorais, é compartilhada com o administrador/curador de dados. O cargo de administrador/curador de dados, muitas vezes associado ao arquivista, lida com a própria gestão de dados garantindo a sua qualidade e são requisitados conhecimentos em banco de dados relacional. Além disso, ele compartilha com o bibliotecário de dados a função de organizar a visualização de dados. Finalmente, aparecem para os três cargos: conhecimentos sobre padrões de metadados, experiência ou conhecimento para a compreensão na perspectiva do pesquisador, conhecimento de dados de pesquisa e familiaridade com pacotes de software estatísticos de análise de dados.

Figura 1 - Requisitos e funções solicitadas para profissionais da informação.



Fonte: Adaptado de Lyon et al. (2015)

Madeiro e Dias (2020) também fizeram pesquisa em sites de bancos internacionais de trabalho e verificaram aspectos requeridos para o perfil profissional do arquivista de dados, destacando a formação profissional orientada a dados com conhecimento de gestão documental, gerenciamento e curadoria de dados, descrição de metadados e ainda o domínio de pacotes de software estatísticos e linguagens computacionais. Enquanto a curadoria de dados de pesquisa, reconhecida pelas práticas de gestão de dados de pesquisa (GRANT, 2017), exige uma equipe com muitas expertises onde dados "[...] precisam ser preservados e arquivados de modo que propriedades arquivísticas, como proveniência, confiabilidade e autenticidade sejam mantidas" podendo ser melhores desenvolvidas por profissionais denominados de arquivistas de dados (SAYÃO; SALES, 2016). Lyon et al. (2015) vão além quando formulam que, considerando que dados são uma documentação significativa para a pesquisa, então todos os arquivistas são arquivistas de dados.

Dados de pesquisa são associados a pesquisas científicas, contudo eles podem incluir outros assuntos como, por exemplo, dados de censos demográficos, histórias orais, pesquisas longitudinais em Ciências Sociais e resultados de pesquisas em artes e humanidades (GRANT, 2017). A inclusão de arquivistas como membros em equipes multidisciplinares de projetos de diversas áreas traz vantagens na conscientização dos membros quanto à importância de se arquivar dados de pesquisa sendo seu papel o de garantir que os dados sejam preparados visando



a preservação a longo prazo (HUMPHREY et al., 2000). Além disso, a gestão e publicação de dados de pesquisa primária para a pesquisa acadêmica, considerando-se a usabilidade e documentação dos dados, é também papel do arquivista de dados (HUVILA, 2016).

Segundo estimativas, 80% do trabalho de um cientista de dados é na preparação dos dados, e os usuários sem experiência em análise de dados não estarão preparados para fazer a limpeza sozinhos (MASON; PATIL, 2015). Segundo Porto e Ziviani (2014), a gestão de dados é um dos grandes desafios da CD e inclui a representação e particionamento de dados e o tratamento dos graus de confiança dos dados quanto às fontes de informação. Nesses aspectos, o arquivista pode contribuir fortemente com as equipes interdisciplinares de projetos em CD com sua competência de gerir dados cuidando, entre outras coisas, da proveniência e organicidade. Enquanto curador de dados de pesquisa, o arquivista pode também aplicar métodos de descoberta de conhecimento por meio de mineração de dados para enriquecer a base de dados e contribuir com o trabalho de pesquisadores que buscam facetas diferenciadas do conjunto de dados para alimentar novas pesquisas científicas. Apesar da demanda por arquivistas com competências em atividades que envolvem dados estar aumentando, Lyon et al. (2015) alertam que, no âmbito dos EUA, devido ao desconhecimento formal do termo arquivista de dados, as funções associadas a esse cargo estão sendo renomeadas e reclassificadas para outros cargos.

#### **4 OFERTA DA DISCIPLINA CIÊNCIA DE DADOS PARA O CURSO DE GRADUAÇÃO EM ARQUIVOLOGIA**

Considerando que o currículo de uma disciplina "[...] não compreende apenas conteúdos, mas um curso a ser seguido e uma construção social - em nível da prescrição, do processo e da prática - que tem como objetivo promover a formação do aluno" (CHAGAS; NEGREIROS; SILVA, 2021), foi proposta a disciplina CD, no âmbito do curso de graduação em Arquivologia da UFES, e com foco na formação de um arquivista sintonizado às reflexões feitas no presente artigo. A disciplina é optativa e faz parte do grupo de disciplinas denominado de 'Tecnologia e Comunicação', com carga horária de 60h e duas disciplinas como pré-requisito (UFES, 2017): (i) Raciocínio Lógico, obrigatória de 60h e ofertada no primeiro período do curso, que contribui diretamente para a formação da competência de criação e análise de expressões lógicas necessárias em algumas construções na CD; e (ii) Sistemas Gerenciadores de Banco de Dados, optativa de 60h, que contribui na compreensão e organização de banco de dados relacional e,

principalmente, quanto a escrita de consultas na linguagem SQL<sup>3</sup> que fornece base para análise comparativa com outros métodos de descoberta de conhecimento usados na CD.

Baseado na oferta da disciplina para o semestre letivo 2022/1, e já considerando algumas atualizações previstas, essa seção apresenta a ementa, objetivos, bibliografia, o conteúdo, as ferramentas computacionais sugeridas para os trabalhos práticos, a proposta de estudos de casos e o emprego da metodologia PBL em seu processo de ensino-aprendizagem.

#### 4.1 EMENTA, OBJETIVOS E BIBLIOGRAFIA

A ementa, Quadro 1, contém uma introdução à CD para fundamentar o aluno com conceitos básicos e aplicações da área na sociedade e nas organizações. Em seguida situa o arquivista na área enquanto um profissional com competência para lidar com dados estruturados. Depois, estuda a fase que consome maior tempo na CD, que é a preparação e transformação de dados. Finalmente, propõe o uso de alguma metodologia para mineração de dados e a sua visualização e interpretação.

Quadro 1 - Ementa para a disciplina Ciência de Dados.

Introdução à Ciência de Dados. O arquivista na Ciência de Dados. Preparação e transformação de dados. Descoberta de conhecimento em bases de dados. Mineração de dados. Visualização e interpretação de dados

Fonte: Atualização proposta pelos autores a partir do PPC publicado em UFES (2017)

Os objetivos são mostrados no Quadro 2 e tem três focos, o primeiro é o conhecimento dos elementos básicos da CD, o segundo é situar o arquivista na CD, e o terceiro é compreender e praticar, em algum contexto de dados estruturados, as típicas fases de um processo de descoberta de conhecimento em bases de dados (KDD).

Quadro 2 - Objetivos para a disciplina Ciência de Dados.

1. Conhecer elementos básicos da Ciência de Dados.
2. Estabelecer possíveis atuações do arquivista na Ciência de Dados.
3. Compreender e aplicar ações de preparação e transformação de dados, descoberta de conhecimento, mineração de dados, visualização e interpretação de dados em contextos sociais ou organizacionais representados por bases de dados.

Fonte: Atualização proposta pelos autores a partir do PPC publicado em UFES (2017)

<sup>3</sup> SQL (Standard Query Language) é uma linguagem para escrita de consultas em bancos de dados relacionais.

Quanto à bibliografia, parte dela cobre assuntos relacionados ao arquivista e sua relação com dados, outra parte sobre a CD propriamente dita e, a terceira parte, sobre a Análise de Redes Complexas enquanto método para a CD. A lista de obras completa encontra-se em uma biblioteca dinâmica e virtual de referências<sup>4</sup>, organizada e mantida na plataforma Zotero<sup>5</sup>.

#### 4.2 CONTEÚDO E FERRAMENTAS COMPUTACIONAIS

O conteúdo, mostrado no Quadro 3, é organizado em oito unidades e está descrito nas notas de aula da disciplina, disponível em Cristovão (2022a). A primeira unidade, que trata do arquivista na CD, tem o objetivo de situar e motivar o aluno a conhecer e explorar o conteúdo que ele irá estudar ao longo do semestre letivo, destacando reflexões acerca de funções do profissional denominado de arquivista de dados e da curadoria digital. Em seguida, as unidades 2, 3 e 4 preparam o estudante com conceitos fundamentais e que são também pré-requisitos para um entendimento mais significativo da CD considerando a vertente que será trabalhada. A unidade 5 aborda alguns elementos selecionados da CD. As unidades 6 e 7 apresentam um dos métodos de mineração de dados da CD, denominado de Análise de Redes Complexas. Esse método, auxiliado por ferramentas apropriadas, foi escolhido por ser possível de ser aplicado sem necessitar de conhecimentos aprofundados de computação ou estatística, tendo em vista a área de conhecimento dos alunos do curso Arquivologia. Finalmente, os estudos de caso, detalhados na subseção 4.3, são importantes para exemplificar e auxiliar a aprendizagem de partes do conteúdo teórico.

Quadro 3 - Conteúdo da disciplina Ciência de Dados.

<p><b>Unidade 1 - O Arquivista na Ciência de Dados:</b> Arquivista de dados; Curadoria digital.</p> <p><b>Unidade 2 - Dados e Variáveis:</b> Dados Relacionados com Informação e Conhecimento; Categorias de dados; Classificação de Variáveis.</p> <p><b>Unidade 3 - Dados, Informações e Eventos na Internet:</b> <i>Data Driven</i>, <i>Datafication</i> e <i>Dataism</i>; Infometria; Dados Abertos; <i>Internet of Events</i> e <i>Big Data</i>; Princípios FAIR; LGPD: Anonimização e Pseudonimização de Dados Pessoais.</p> <p><b>Unidade 4 - Web de Dados:</b> Contexto; Dados Ligados; Classificação 5 estrelas de Berners-Lee.</p> <p><b>Unidade 5 - Ciência de Dados:</b> Definição; Cientista de Dados; Exemplos de aplicações de Ciência de Dados; Correlação entre variáveis; Correlação não implica em causalidade; Manipulações tendenciosas de dados; Mineração de Dados; Data Warehouse; Descoberta de Conhecimento em Bases de Dados (KDD); Softwares de Manipulação de Dados; Visualização; Metodologias e técnicas da Ciência de Dados.</p> <p><b>Unidade 6 - Ciência das Redes:</b> Redes Complexas; Tipos de Redes; Cronologia, modelos e marcos históricos.</p> <p><b>Unidade 7 - Análise de Redes Complexas:</b> Aplicações; Fundamentos; Formatação e Inspeção Visual; Conceitos Básicos; Métricas de Centralidade; Coesão, Comunidades e Contágio; Redes Bipartidas; Projeção</p>
--

<sup>4</sup> Biblioteca virtual da disciplina, contendo a bibliografia usada, disponível em:

[https://www.zotero.org/groups/4663489/cincia\\_de\\_dados/library](https://www.zotero.org/groups/4663489/cincia_de_dados/library)

<sup>5</sup> Zotero é um software gerenciador de referências bibliográficas. Disponível em: <https://www.zotero.org/>.

Bipartida; Software para Análise de Redes Complexas.

**Unidade 8 - Estudos de casos.**

Fonte: Plano de ensino publicado em <http://arquivologia.ufes.br/>.

As ferramentas computacionais usadas pelos alunos para os estudos de casos (subseção 4.3), atividades práticas e desenvolvimento do PBL (subseção 4.4) são organizadas em quatro categorias, sendo todas de uso gratuito: (1ª) Preparação de dados, com o software OpenRefine<sup>6</sup>; (2ª) Análise de redes, com os softwares VOSviewer<sup>7</sup> e Gephi<sup>8</sup>; (3ª) Visualização de dados, com os softwares Google Data Studio<sup>9</sup> ou Power BI Desktop<sup>10</sup>; (4ª) Consultas em bases de dados, com os softwares MySQL Workbench<sup>11</sup> ou SQLite Online<sup>12</sup>.

Existem diversos outros assuntos e ferramentas computacionais dentro do vasto campo da CD que poderiam ser abordados na disciplina. Contudo, escolheu-se um grupo de assuntos e de ferramentas que pudessem oferecer uma sequência coesa e lógica, sendo o seu estudo e uso factíveis para um aluno que não é da área tecnológica. Além disso, a escolha contemplou elementos destacados e discutidos na seção 3, por Lyon et al. (2015), Madeiro e Dias (2020), Sayão e Sales (2016), Grant (2017), Huvila (2016), Yeo e Lowry (2020), Humphrey et al. (2000) quanto às funções de arquivista de dados e curador de dados, tais como: preparação de dados, integração de dados, coleta de dados, análise de dados, visualização de dados, gestão de dados, qualidade de dados e manipulações com banco de dados relacional.

#### 4.3 ESTUDOS DE CASOS

Complementando os assuntos teóricos e práticas desenvolvidas no laboratório de computadores, os estudos de casos cumprem um papel importante para o reforço da aprendizagem, pois os estudantes têm oportunidade de familiarizar-se com as ferramentas computacionais em contextos reais, isto é, em bases de dados abertos coletados de sites

<sup>6</sup> OpenRefine é um software para limpeza, preparação e reconciliação de dados. Classificado como *Interactive Data Transformation* (HOOLAND; VERBORGH; WILDE, 2013), trabalha com grande massa de dados, diferentemente das populares planilhas eletrônicas. Disponível em: <https://openrefine.org/>.

<sup>7</sup> VOSViewer é um software para análise de redes bibliométricas. Disponível em: <https://www.vosviewer.com/>.

<sup>8</sup> Gephi é um software interativo para visualização, exploração e análise de vários tipos de redes e sistemas complexos, grafos dinâmicos e hierárquicos. Em CD ele oferece uma das técnicas de mineração de dados, baseada em análise de redes complexas por inspeção visual. Disponível em: <https://gephi.org/>.

<sup>9</sup> Google Data Studio é uma ferramenta de visualização de dados e geração e publicação de dashboards. Disponível em: <https://datastudio.google.com/>.

<sup>10</sup> Power BI Desktop é uma ferramenta de visualização de dados. Apesar de mais poderosa que o Google Data Studio, não oferece o serviço de publicação de dashboards de forma gratuita. Disponível em: <https://powerbi.microsoft.com/pt-br/desktop/>.

<sup>11</sup> MySQL Workbench é um Sistema Gerenciador de Banco de Dados (SGBD), baseado no MySQL, com interface visual para dar apoio às técnicas de mineração de dados em CD. Disponível em: <https://www.mysql.com/>.

<sup>12</sup> SQLite Online é um SGBD leve com interface visual que funciona direto no navegador Web, sem instalação, para dar apoio às técnicas de mineração de dados quando a base de dados é pequena. Disponível em: <https://sqliteonline.com/>.

institucionais. A taxonomia proposta por Barrows (1986), sobre métodos de ensino e o quanto eles se aproximam do uso de problemas reais na aprendizagem, é composta por seis níveis, sendo o primeiro mais simples e o sexto o mais elaborado. Os estudos de casos desenvolvidos na disciplina classificam-se, nessa taxonomia, no nível dois, denominado de *case-based lectures*, onde o estudante vivencia a teoria apoiada por um estudo de caso determinado pelo professor.

Os estudos de casos atualmente disponíveis no material da disciplina fornecido aos alunos (CRISTOVÃO, 2022b) são: (i) a análise bibliométrica orientadas por redes de citação e coautoria de publicações científicas do sobre um determinado tema; (ii) manipulações de dados oriundos de um formulário online; (iii) mapeamento e análise de redes sobre documentos cadastrados e recebidos pelo órgão ANVISA; (iv) mapeamento e análise de redes, consultas via banco de dados relacionais e criação de dashboard sobre base de dados de acidentes rodoviários em rodovias federais no Brasil; (v) mapeamento e análise de redes, consultas via banco de dados relacionais e criação de dashboard sobre base de dados de emendas parlamentares do congresso brasileiro.

#### 4.4 METODOLOGIA COM FOCO EM PBL

Cerca de 40% da carga horária é usada em aulas expositivas e dialogadas sobre os temas do conteúdo. Os outros 60% são práticas no laboratório de computadores para desenvolvimento de atividades individuais e coletivas. Como apoio às aulas presenciais, existe uma sala virtual, baseada no Moodle, para concentrar os principais elementos da disciplina como: canais de comunicação e interação, plano de ensino, notas de aula, registro de aulas, atividades para fazer e os diários de bordo dos alunos.

O diário de bordo é um registro individual, cronológico e dinâmico que contém resultados e reflexões acerca dos assuntos discutidos na disciplina e atividades realizadas, dificuldades encontradas, crescimentos e amadurecimentos obtidos, interpretações e análises sobre os resultados oriundos dos experimentos realizados. É solicitado ao aluno que seja feito ao menos um registro semanal. Esse registro faz o aluno refletir sobre a sua aprendizagem, sobre o conteúdo que foi discutido e sobre o manuseio das ferramentas computacionais usadas nas aulas práticas.

O principal método empregado na disciplina é o PBL, *problem based learning* ou aprendizagem baseada em problemas. Ela é "[...] uma metodologia de ensino-aprendizagem caracterizada pelo uso de problemas da vida real para estimular o desenvolvimento do pensamento crítico e das habilidades de solução de problemas [...]" (RIBEIRO, 2008) e, conseqüentemente, para a melhoria da compreensão de conceitos fundamentais da matéria.

Segundo Ribeiro, o PBL usa “um problema para iniciar, direcionar, motivar e focar a aprendizagem, diferentemente de outras metodologias convencionais que utilizam problemas de aplicação ao final da apresentação de um conceito ou conteúdo”. Ainda segundo Ribeiro, a metodologia PBL tem cinco princípios e, na disciplina em questão, se caracterizam por: (i) Motivação Epistêmica: ocorre nas reflexões sobre a natureza do conhecimento e o quanto ele pode proporcionar segurança para fazer escolhas adequadas nos vários processos decisórios dos métodos da CD; (ii) Interação com a Vida Real: ocorre ao escolher uma base de dados abertos pertencente a alguma instituição, problematizar uma situação relativamente pequena e desenvolver a solução por meio dos elementos estudados na disciplina; (iii) Metacognição: acontece quando o aluno reflete sobre a própria forma como está aprendendo, passando a se questionar se as suas práticas são suficientes para o domínio dos métodos e ferramentas usados na resolução do problema em questão; (iv) Construção do Conhecimento: acontece por meio da vivência, ao longo do tempo, das ações executadas para resolução do problema e (v) Interação Social: é obtida nas interações dos membros do grupo de trabalho para debater entre si decisões de caminhos a serem tomados e, posteriormente com a turma nos momentos de apresentação das etapas desenvolvidas.

O problema, propriamente dito, do PBL é o desenvolvimento das fases de um processo KDD sobre uma base de dados escolhida pelo grupo de trabalho. Essas fases foram organizadas em 7 etapas, conforme mostra a Figura 2.

Etapa 1 - Investigação e escolha de uma base de dados para o PBL. São sugeridos aos estudantes portais de dados abertos para escolha de uma base que fornecerá os dados primitivos para o desenvolvimento das etapas do PBL.

Etapa 2 - Preparação e limpeza dos dados. Com apoio do *software OpenRefine* é feita uma análise na base de dados, considerando as suas peculiaridades, para corrigir problemas, padronizar nomes de variáveis e dados, excluir dados desnecessários e diminuir o escopo do problema caso necessário. No contexto das funções arquivísticas, essa etapa se enquadra na fase de produção/criação de documentos.

Etapa 3 - Transformações na base de dados. Usando o *software OpenRefine* e considerando os métodos que serão empregados posteriormente, as variáveis com grande quantidade de valores, como as quantitativas, são transformadas em qualitativas por meio da criação de categorias de dados. Esse também é o momento de atender a LGPD (BRASIL, 2018) para realização de anonimização ou pseudonimização de dados, conforme a necessidade avaliada. No contexto

das funções arquivísticas, essa etapa se enquadra nas fases da classificação e descrição de documentos.

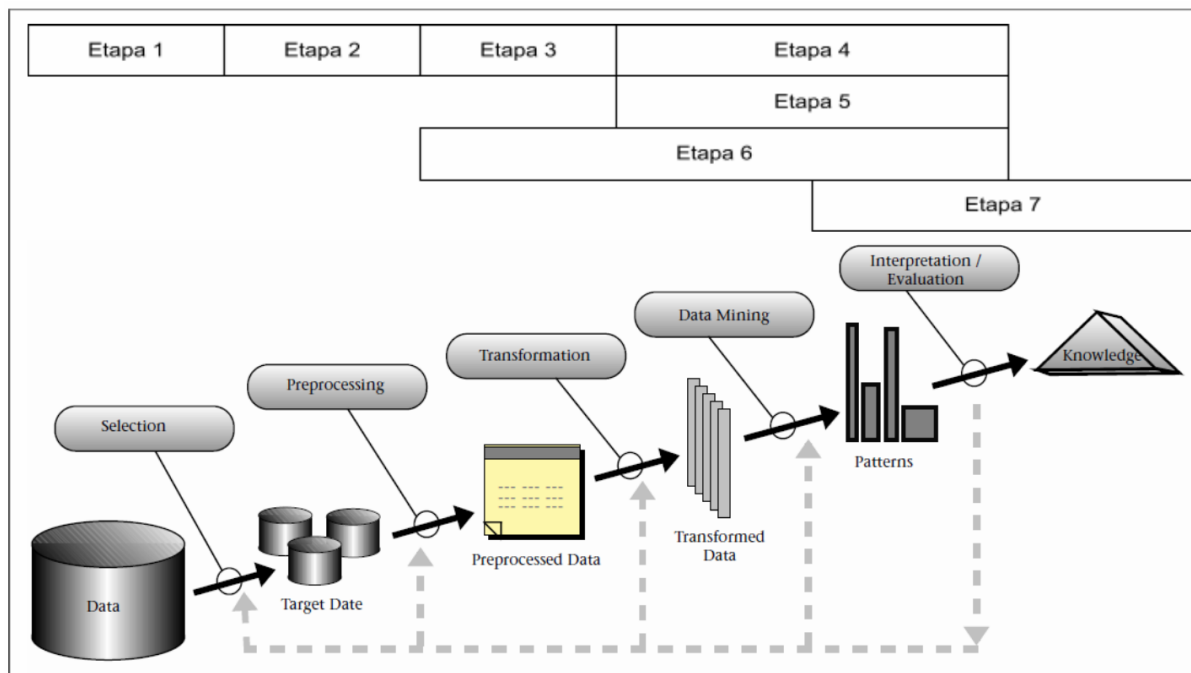
Etapa 4 - Análise exploratória na base de dados por meio de consultas SQL/SGBD. Com apoio dos *softwares OpenRefines* e *MySQL Workbench*, são feitas diversas consultas, escritas em SQL, para extrair informações que, posteriormente, contribuem para o *storytelling* da etapa 7.

Etapa 5 - Visualização de dados com *dashboard*. Usando os *softwares Google Data Studio* ou *Power BI Desktop* são feitas representações gráficas dos dados da base com o intuito de criar um *dashboard* para contribuir para o *storytelling* da etapa 7.

Etapa 6 - Análise exploratória por meio de técnicas de análise de redes por inspeção visual. Suportada pelos *softwares OpenRefine* e *Gephi*, essa etapa é a mais complexa para os estudantes e visa mapear os dados da base para o formato de rede e executar uma mineração de dados utilizando-se do método de análise de redes complexas por inspeção visual. São feitas interpretações e avaliações analíticas sobre os resultados obtidos na inspeção visual.

Etapa 7 - *Storytelling* de um cenário do PBL desenvolvido. Utilizando-se dos resultados obtidos na análise exploratória de dados com as consultas SQL, com a visualização de dados por *dashboard* e com a inspeção visual realizada sobre as redes de informação, os estudantes criam um *storytelling* sobre um cenário de uma parte do PBL e o apresentam para a turma. Nesse *storytelling* são abordadas as ações executadas em cada etapa do PBL e a sua importância para obtenção dos resultados. Além disso, os estudantes são estimulados a especular o quanto os resultados obtidos poderiam beneficiar a sociedade ou a instituição detentora da base de dados.

Figura 2 - Fases do processo KDD com as etapas do PBL.



Fonte: Adaptado de Lyon et al. (2015)

Na taxonomia proposta por Barrows (1986), conforme apresentado na subseção 4.3, o trabalho do PBL proposto na disciplina é enquadrado como nível cinco denominado de *problem-based*, entre os seis níveis existentes. Nesse nível os estudantes fazem livre investigação para encontrar a solução de um problema. O sexto e último nível exigiria reavaliar o resultado após a solução proposta. Isso não foi possível, uma vez que os resultados obtidos não foram apresentados às instituições que geraram a base de dados.

## 5 CONSIDERAÇÕES FINAIS

Neste relato de experiência discutiu-se a importância da disciplina CD para a formação do arquivista e apresentou-se uma proposta de programa de disciplina para o curso de graduação em Arquivologia, com ênfase no conteúdo e na metodologia empregada. Relatou-se também alguns procedimentos pedagógicos adotados em experiências de ofertas de turmas da disciplina, como, por exemplo, estudos de casos e problemas desenvolvidos sob a metodologia PBL, onde conseguiu-se atingir níveis elevados na taxonomia proposta por Barrows (1986) sobre métodos de ensino e o quanto eles se aproximam do uso de problemas reais na aprendizagem. Os *feedbacks* dos alunos que cursaram a disciplina foram positivos, tanto na motivação e engajamento nas técnicas e conteúdos trabalhados, quanto na percepção de melhoria na sua formação.



Quanto à atuação do arquivista na Ciência de Dados, há demanda relatada em algumas pesquisas, e percebeu-se que esse profissional pode contribuir de forma bastante significativa em equipes interdisciplinares e com papéis relevantes na gestão de dados, na curadoria de dados de pesquisa e na conscientização de pesquisadores sobre o cuidado que se deve ter com os dados. Considerando-se o contexto das funções arquivísticas, ele deve atuar desde a produção/criação de documentos, passando pela avaliação, descrição, difusão, preservação e aquisição, obedecendo os princípios arquivísticos como proveniência e organicidade.

Apesar de direcionada ao curso de graduação em Arquivologia, a oferta da disciplina CD é aberta a alunos de outros cursos. Nesse caso, durante a Unidade 1, direcionada à atuação do arquivista na CD, os alunos de outros cursos são estimulados a investigar como e o quanto a CD pode colaborar nas atribuições do profissional de sua área de atuação e também na sua formação durante o curso. Além disso, o contexto do problema para o PBL, abordado na subseção 4.4, é escolhido conforme motivação ou aproximação da área de conhecimento do estudante. Em todos os casos os contextos escolhidos, bem como os resultados obtidos, são compartilhados e debatidos com a turma. Para o estudante de Arquivologia, essa composição heterogênea é benéfica para a sua formação, uma vez que ele tem oportunidade de conhecer outros contextos de aplicação dos métodos de CD.

Conclui-se que é importante haver uma disciplina de CD na formação do aluno de graduação em Arquivologia, pois a tendência é que essas habilidades nos campos de conhecimento híbridos entre tecnologias e Arquivologia sejam cada vez mais demandadas pelo mercado. Contudo, o ensino de CD deve ser contextualizado na área de aplicação, e deve ser adequado às habilidades prévias apresentadas pelos alunos, para efetivação do aprendizado e para que o aluno possa aplicar as ferramentas e conhecimentos nas demandas futuras. Como continuidade da pesquisa, pretende-se acompanhar os alunos que cursaram a disciplina CD para verificação de demandas quanto ao desenvolvimento de TCC em temáticas relacionadas; e pretende-se observar se o conteúdo abordado na disciplina de CD será utilizado pelos egressos no mercado de trabalho.

## REFERÊNCIAS

AALST, Wil Van der. Data Science in Action. *Em*: AALST, Wil Van der (org.). **Process Mining**. Berlin, Heidelberg: Springer, 2016. p. 3–23. DOI: [10.1007/978-3-662-49851-4\\_1](https://doi.org/10.1007/978-3-662-49851-4_1). Disponível em: [https://doi.org/10.1007/978-3-662-49851-4\\_1](https://doi.org/10.1007/978-3-662-49851-4_1). Acesso em: 11 jul. 2022.

BARROWS, H. S. A taxonomy of problem-based learning methods. **Medical Education**, [S. l.], v. 20, n. 6, p. 481–486, 1986. DOI: [10.1111/j.1365-2923.1986.tb01386.x](https://doi.org/10.1111/j.1365-2923.1986.tb01386.x). Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2923.1986.tb01386.x>. Acesso em: 11 jul. 2022.

BELLOTTO, Heloísa Liberalli. **O arquivista na sociedade contemporânea**. 2004. Disponível em: <https://www.marilia.unesp.br/Home/Extensao/CEDHUM/texto01.pdf>. Acesso em: 11 jul. 2022.

BRASIL. Lei nº 6.546, DE 4 de julho de 1978. Dispõe sobre a regulamentação das profissões de Arquivista e de Técnico de Arquivo, e dá outras providências. **Diário Oficial da União**, Brasília, DF, 4 de julho de 1978.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados (LGPD). **Diário Oficial da União**, Brasília, DF, 14 de agosto de 2018.

BROOKS, David. The Philosophy of Data. **The New York Times**, [S. l.], 2013. Disponível em: <https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html>. Acesso em: 11 jul. 2022.

CETIC.BR. **Pesquisa sobre o uso das tecnologias de informação e comunicação nos equipamentos culturais brasileiros**. São Paulo: Comitê Gestor da Internet no Brasil, 2019. Disponível em: [https://www.cetic.br/media/docs/publicacoes/1/tic\\_cultura\\_2018\\_livro\\_eletronico.pdf](https://www.cetic.br/media/docs/publicacoes/1/tic_cultura_2018_livro_eletronico.pdf). Acesso em: 11 jul. 2022.

CHAGAS, Cintia Aparecida; NEGREIROS, Leandro Ribeiro; SILVA, Welder Antônio. **Formação em Arquivologia no Brasil: análise e perspectivas acerca de uma possível harmonização curricular**. Florianópolis, SC: Rocha Gráfica e Editora, PPGCI/UFMG, 2021. Disponível em: [https://www.nyota.com.br/files/ugd/c3c80a\\_3b32e497b10b431baf43bc522c4a344f.pdf](https://www.nyota.com.br/files/ugd/c3c80a_3b32e497b10b431baf43bc522c4a344f.pdf). Acesso em: 11 jul. 2022.

COOK, Terry. Arquivologia e pós-modernismo: novas formulações para velhos conceitos. **Informação Arquivística**, [S. l.], v. 1, n. 1, p. 26, 2012. Disponível em: [https://www.brapci.inf.br/repositorio/2015/12/pdf\\_deb3461ca4\\_0000018241.pdf](https://www.brapci.inf.br/repositorio/2015/12/pdf_deb3461ca4_0000018241.pdf). Acesso em: 11 jul. 2022.

CRISTOVÃO, Henrique Monteiro. **Ciência de Dados e Análise de Redes Complexas**. web page. 2022a. Disponível em: <https://bit.ly/henrique-cristovao-arccd>. Acesso em: 11 jul. 2022.

CRISTOVÃO, Henrique Monteiro. **Estudos de casos sobre análise exploratória por meio de análise de redes**. web page. 2022b. Disponível em: <https://bit.ly/henrique-cristovao-ecae>. Acesso em: 11 jul. 2022.

DURANTI, Luciana. Diplomats: New Uses for an Old Science, Part I. **Archivaria**, [S. l.], p. 7–27, 1989. Disponível em: <https://archivaria.ca/index.php/archivaria/article/view/11567>. Acesso em: 11 jul. 2022.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI Magazine**, [S. l.], v. 17, n. 3, p. 37–54, 1996.

FURNER, Jonathan. “Data”: The data. *Em*: KELLY, Matthew; BIELBY, Jared (org.). **Information cultures in the digital age: a festschrift in honor of Rafael Capurro**. Wiesbaden: Springer

Fachmedien, 2016. p. 287–306. DOI: [10.1007/978-3-658-14681-8\\_17](https://doi.org/10.1007/978-3-658-14681-8_17). Disponível em: [https://doi.org/10.1007/978-3-658-14681-8\\_17](https://doi.org/10.1007/978-3-658-14681-8_17). Acesso em: 11 jul. 2022.

GAVA, Tânia Barbosa Salles; FERRARI, Luciana Itida; MORAES, Margarete Farias De. O ensino em Arquivologia na perspectiva da era digital: O caso do curso de Arquivologia da UFES. *Em*: BARROS, Thiago Henrique Bragato; SANTOS JR, Robero Lopes Dos; CÂNDIDO, Gilberto Gomes (org.). **A pesquisa e o ensino em Arquivologia: perspectivas na era digital**. Belém, PA: REPARQ,[s.d.]. p. 261–269. Disponível em: <https://feparg.org/vi-reparq>. Acesso em: 11 jul. 2022.

GRANT, Rebecca. Recordkeeping and research data management: a review of perspectives. **Records Management Journal**, [S. l.], v. 27, n. 2, p. 159–174, 2017. DOI: [10.1108/RMJ-10-2016-0036](https://doi.org/10.1108/RMJ-10-2016-0036). Disponível em: <https://doi.org/10.1108/RMJ-10-2016-0036>. Acesso em: 11 jul. 2022.

HAND, David J.; MANNILA, Heikki; SMYTH, Padhraic. **Principles of data mining**. Cambridge, MA: MIT Press, 2001.

HARARI, Yuval Noah. **Homo Deus**. 1. ed. São Paulo: Companhia das Letras, 2016.

HJØRLAND, Birger. Data (with big data and database semantics). **Knowledge Organization**, [S. l.], v. 45, n. 8, p. 643–652, 2018. DOI: [10.5771/0943-7444-2018-8-643](https://doi.org/10.5771/0943-7444-2018-8-643). Disponível em: <https://www.isko.org/cyclo/data>. Acesso em: 11 jul. 2022.

HOOLAND, Seth Van; VERBORGH, Ruben; WILDE, Max De. Cleaning Data with OpenRefine. **Programming Historian**, [S. l.], 2013. Disponível em: <https://programminghistorian.org/en/lessons/cleaning-data-with-openrefine>. Acesso em: 11 jul. 2022.

HUMPHREY, Charles K.; ESTABROOKS, Carole A.; NORRIS, Judy R.; SMITH, Jane E.; HESKETH, Kathryn L. Archivist on board: contributions to the research team. **Forum Qualitative Sozialforschung / Forum: Qualitative Social Research**, [S. l.], v. 1, n. 3, 2000. DOI: [10.17169/fqs-1.3.1022](https://doi.org/10.17169/fqs-1.3.1022). Disponível em: <https://www.qualitative-research.net/index.php/fqs/article/view/1022>. Acesso em: 11 jul. 2022.

HUVILA, Isto. “If we just knew who should do it”, or the social organization of the archiving of archaeology in Sweden. **IR Information Research**, [S. l.], v. 21, n. 2, 2016. Disponível em: <http://informationr.net/ir/21-2/paper713.html#.YKx846hKjYY>. Acesso em: 11 jul. 2022.

LYON, L.; MATTERN, Eleanor; ACKER, Amelia; LANGMEAD, Alison. Applying translational principles to data science curriculum development. *Em*: IPRES 2015 PROCEEDINGS 2015, Chapel Hill, North Carolina. **Anais [...]**. . *Em*: IPRES 2015. Chapel Hill, North Carolina: University of Pittsburgh, 2015. Disponível em: <http://d-scholarship.pitt.edu/27159/>. Acesso em: 11 jul. 2022.

MADEIRO, Andre França; DIAS, Guilherme Ataíde. Arquivista de dados: análise do perfil profissional em sites de bancos internacionais de trabalho. **ÁGORA: Arquivologia em debate**, [S. l.], v. 30, n. 61, p. 649–664, 2020. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/924>. Acesso em: 11 jul. 2022.

MARTINS, Dalton Lopes. Data science teaching and learning models: focus on the Information Science area. *Em*: RODRIGUES DIAS, Thiago Magela (org.). **Advanced Notes in Information Science**. [s.l.] : ColNes Publishing, 2022. v. 2. DOI: [10.47909/anis.978-9916-9760-3-6.100](https://doi.org/10.47909/anis.978-9916-9760-3-6.100). Disponível em: <https://pub.colnes.org/index.php/anis/article/view/100>. Acesso em: 11 jul. 2022.

MASON, Hilary; PATIL, Dj. **Data Driven**. Sebastopol, CA, USA: O'Reilly Media, Inc., 2015.

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. **Big Data: the essential guide to work, life and learning in the Age of Insight**. [s.l.] : Hachette UK, 2013.

NATIONAL SCIENCE FOUNDATION. **Long-lived digital data collections: enabling research and education in the 21st century**. Alexandria, VA, USA: National Science Board, 2005. Disponível em: <https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>. Acesso em: 11 jul. 2022.

OLIVEIRA, Carolina de. **Um estudo de caso sobre datasets do Ministério da Justiça: dados brutos ou documentos arquivísticos?** 2015. Dissertação - (Mestrado Profissional em Gestão de Documentos e Arquivos) - Centro de Ciências Humanas e Sociais, Universidade Federal do Estado do Rio de Janeiro - UNIRIO, Rio de Janeiro, RJ, 2015. Disponível em: <http://www.repositorio-bc.unirio.br:8080/xmlui/handle/unirio/11754>. Acesso em: 11 jul. 2022.

OLIVEIRA, Flávia Helena de. **A formação em Arquivologia nas universidades brasileiras: objetivos comuns e realidades particulares**. 2014. (Doutorado em Ciência da Informação) - Faculdade de Ciência da Informação da Universidade de Brasília, Brasília, DF, 2014. Disponível em: <https://repositorio.unb.br/handle/10482/17966>. Acesso em: 11 jul. 2022.

PENNA, Giuseppe Della; MAGAZZENI, Daniele; OREFICE, Sergio. A spatial relation-based framework to perform visual information extraction. **Knowledge and Information Systems**, [S. l.], v. 30, n. 3, p. 667–692, 2012. DOI: [10.1007/s10115-011-0394-4](https://doi.org/10.1007/s10115-011-0394-4). Disponível em: <https://link.springer.com/article/10.1007/s10115-011-0394-4>. Acesso em: 11 jul. 2022.

PORTO, Fábio; ZIVIANI, Arthur. Ciência de Dados. *Em*: 2014, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: SBC, 2014. Disponível em: <https://www.lncc.br/~ziviani/papers/III-Desafios-SBC2014-CiD.pdf>. Acesso em: 03 mar. 2020.

RIBEIRO, Luis Roberto de Camargo. **Aprendizagem baseada em problemas (PBL): uma experiência no ensino superior**. [s.l.] : EdUFSCar, 2008. Disponível em: <https://books.scielo.org/id/w57z2>. Acesso em: 11 jul. 2022.

SANT'ANA, Ricardo César Gonçalves. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Informação & Informação**, [S. l.], v. 21, n. 2, p. 116–142, 2016. DOI: [10.5433/1981-8920.2016v21n2p116](https://doi.org/10.5433/1981-8920.2016v21n2p116). Disponível em: <https://www.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 11 jul. 2022.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria digital e dados de pesquisa. **AtoZ: novas práticas em informação e conhecimento**, [S. l.], v. 5, n. 2, p. 67–71, 2016. DOI: [10.5380/atoz.v5i2.49708](https://doi.org/10.5380/atoz.v5i2.49708). Disponível em: <https://revistas.ufpr.br/atoz/article/view/49708>. Acesso em: 11 jul. 2022.

UFES. **Projeto pedagógico do curso de graduação em Arquivologia da Universidade Federal do Espírito Santo**, 2017. Disponível em: <https://arquivologia.ufes.br/grade-curricular>. Acesso em: 11 jul. 2022.

WASSERMAN, Stanley; FAUST, Katherine. **Social network analysis: methods and applications**. Cambridge, England; New York: Cambridge University Press, 1994.

WILKINSON, Mark D. et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, [S. l.], v. 3, n. 1, p. 160018, 2016. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18). Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 11 jul. 2022.

YEO, Geoffrey; LOWRY, James. Data, information and records: exploring definitions and relationships. Em: THURSTON, Anne (org.). **A Matter of Trust**. Building Integrity into Data, Statistics and Records to Support the Achievement of the Sustainable Development Goals. London: University of London Press, 2020. p. 49–66. Disponível em: <https://www.jstor.org/stable/j.ctv14t45v5.8>. Acesso em: 11 jul. 2022.

## NOTAS DE AUTORIA

### Henrique Monteiro Cristovão

Doutor em Ciência da Informação na Universidade de Brasília (UnB) com estágio de pesquisa (Doutorado Sanduíche) no Institute for Human & Machine Cognition (IHMC/EUA). Mestre em Informática na Universidade Federal do Espírito Santo (UFES). Bacharel em Matemática Aplicada e Computacional na UFES. Professor Adjunto na UFES no Programa de Pós Graduação em Ciência da Informação (PPGCI) e no Departamento de Arquivologia. Líder do grupo de pesquisa Organização e Recuperação de Conhecimento em Rede (NetKOR), registrado no CNPq. Participa do grupo de pesquisa Observatório da Informação Arquivística Digital. Temas de pesquisa atuais: Organização, representação e recuperação de informação e conhecimento em rede; Interoperabilidade, metadados, ontologias e dados ligados na Web semântica; Ciência de Dados pelo viés da Ciência da Informação com ênfase em processos de análise de redes, descoberta e visualização de conhecimento..

Link Currículo Lattes - <http://lattes.cnpq.br/5035919384923489>

### Luciana Itida Ferrari

Doutora em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro (COPPE/UFRJ) com estágio de pesquisa (Doutorado Sanduíche) no Instituto Gulbenkian de Ciência, em Portugal. Mestre em Informática pela Universidade Federal do Rio de Janeiro (UFRJ) e bacharel em Ciência da Computação pela Universidade Católica de Petrópolis (UCP). Professora Adjunta na Universidade Federal do Espírito Santo (UFES), no Departamento de Arquivologia. Coordena o grupo de pesquisa Observatório da Informação Arquivística Digital e é integrante do grupo de pesquisa Ged/A - Documentos Digitais: Gestão, Curadoria Digital, Preservação, Acesso e Transparência Ativa registrado no CNPq. Temas de pesquisa atuais: Tecnologia da Informação aplicada à Arquivologia, incluindo SIGAD e RDC-Arq; Ciência de Dados, incluindo mineração de dados, inteligência artificial, business intelligence e visualização de resultados; Tecnologias Educacionais.

Link Currículo Lattes - <http://lattes.cnpq.br/8738517383983971>

### Margarete Farias de Moraes

Doutora em Educação pela UFES. Mestre em História das Ciências da Saúde pela COC/FIOCRUZ. Bacharel em Arquivologia pela UNIRIO, Licenciatura e Bacharel em História pela UERJ. Estágio pós doutoral pelo Programa de Pós Graduação em Difusão do Conhecimento na UFBA. Professora e pesquisadora do Departamento de Arquivologia da UFES e do Programa de Pós Graduação em Ciência da Informação da UFES. Experiência em docência e organização de centros de documentação e gestão de documentos e informações em instituições de saúde, com ênfase no planejamento e gestão arquivísticos. Membro do grupo de pesquisa Observatório da Informação Arquivística Digital. Tenho como campos de investigação a história, a educação e o gerenciamento e tecnologias da informação na área de saúde..

Link Currículo Lattes - <http://lattes.cnpq.br/6346587909398922>