

GESTÃO DA INFORMAÇÃO E CIÊNCIA DE DADOS EM PLATAFORMAS DE BIODIVERSIDADE MARINHA: DESAFIOS PARA INTEGRAÇÃO E ANÁLISE DE DADOS

INFORMATION MANAGEMENT AND DATA SCIENCE IN MARINE BIODIVERSITY PLATFORMS: CHALLENGES FOR DATA INTEGRATION AND ANALYSIS

Leandro Coqueiro Souza

ORCID: <https://orcid.org/0009-0009-2327-346X>

Mestre em Ciência, Tecnologia e Sociedade pelo Instituto Federal do Paraná (IFPR), Brasil.

Analista de Tecnologia da Informação na Universidade Federal do Paraná (UFPR), Brasil.

Doutorando em Gestão da Informação da Universidade Federal do Paraná (UFPR)

E-mail: leandrocoqueiro@ufpr.br

Denise Fukumi Tsunoda

ORCID: <https://orcid.org/0000-0002-5663-4534>

Doutora em Engenharia Elétrica e Informática Industrial pela Universidade Tecnológica Federal (UTFPR), Brasil.

Professora Titular do Departamento de Ciência e Gestão da Informação da Universidade Federal do Paraná (UFPR), Brasil.

E-mail: dtsunoda@ufpr.br

RESUMO: Objetivo: Analisar o papel das plataformas informacionais na gestão de dados ambientais marinhos, investigando como técnicas de ciência de dados podem apoiar a análise de grandes volumes de dados sobre biodiversidade. Método: O estudo adota abordagem exploratória e qualitativa, baseada na análise conceitual da literatura e na caracterização de plataformas de dados ecológicos disponibilizadas em portais públicos e iniciativas de compartilhamento científico. Resultados: Os resultados evidenciam que, embora as plataformas informacionais ampliem significativamente a disponibilidade de dados sobre biodiversidade marinha, sua efetiva integração e utilização científica são condicionadas à adoção de práticas estruturadas de gestão da informação, especialmente no que se refere à padronização, curadoria e interoperabilidade dos dados. Observa-se ainda que o uso de métodos de ciência de dados potencializa a análise de grandes volumes de dados ecológicos, porém permanece limitado por problemas relacionados à qualidade, heterogeneidade e fragmentação das bases informacionais. Conclusões: Conclui-se que o potencial analítico dos dados de biodiversidade marinha não depende exclusivamente de avanços tecnológicos, mas da articulação entre infraestruturas informacionais, práticas de gestão da informação e métodos de ciência de dados. O estudo contribui ao evidenciar essa interdependência, oferecendo uma análise integrada que reforça a importância de abordagens interdisciplinares para o fortalecimento da produção, organização e uso científico de dados ambientais.

PALAVRAS-CHAVE: gestão da informação; biodiversidade marinha; plataformas informacionais; ciência de dados; *Big Data*.

ABSTRACT: Objective: To analyze the role of information platforms in the management of marine environmental data, investigating how data science techniques can support the analysis of large volumes of biodiversity data. Method: This study adopts an exploratory and qualitative approach, based on a conceptual analysis of the literature and on the characterization of ecological data platforms available through public portals and scientific data-sharing initiatives. Results: The results show that, although information platforms significantly increase the availability of marine biodiversity data, their effective integration and scientific use depend on the adoption of structured information management practices, particularly regarding data standardization, curation, and interoperability. It is also observed that the use of data science methods enhances the analysis of large volumes of ecological data; however, it remains limited by issues related to data quality, heterogeneity, and fragmentation across information systems. Conclusions: It is concluded that the analytical potential of marine biodiversity data does not depend solely on technological advances, but on the articulation between information infrastructures, information management practices, and data science methods. This study contributes by highlighting this interdependence, offering an integrated analysis that reinforces the importance of interdisciplinary approaches to strengthen the production, organization, and scientific use of environmental data.

KEYWORDS: information management; marine biodiversity; information platforms; data science; big data.

1 INTRODUÇÃO

O avanço das tecnologias digitais e a ampliação da disponibilidade de dados ambientais têm reconfigurado as formas de produção, organização e uso da informação científica. No campo da fauna marinha, programas de monitoramento, registros de encalhes, observações de campo e iniciativas institucionais geram volumes crescentes de dados ecológicos, os quais são armazenados em diferentes bases de dados e sistemas informacionais.

A organização estruturada desses dados em ambientes computacionais específicos dá origem aos bancos de dados, compreendidos como sistemas destinados ao armazenamento, gerenciamento e recuperação eficiente de informações (Elmasri; Navathe, 2016). Conforme destaca Date (2004), tais sistemas possibilitam a organização estruturada de registros, viabilizando sua consulta, análise e recuperação ao longo do tempo, de acordo com as demandas dos usuários. Nesse cenário, os sistemas de gerenciamento de bancos de dados configuram-se como instrumentos centrais da gestão da informação, ao permitirem o tratamento de grandes volumes de dados, a otimização de processos e a extração de informações de forma eficiente (Ramakrishnan; Gehrke, 2008).

Embora as primeiras aplicações de bancos de dados remontem à década de 1960, voltadas principalmente a demandas comerciais e governamentais, observa-se, a partir dos anos 2000, sua expansão em diferentes campos científicos, incluindo os estudos sobre biodiversidade (Michener, 2015). Esse movimento está associado tanto ao avanço das tecnologias de coleta e armazenamento quanto à ampliação de iniciativas de monitoramento ambiental e de pesquisa colaborativa, que impulsionaram a geração e o compartilhamento de dados ecológicos em múltiplas plataformas digitais.

Como consequência, dados provenientes de registros de ocorrência de es-

pécies, monitoramentos sistemáticos e observações científicas passaram a ser disponibilizados em repositórios heterogêneos, ampliando o potencial de uso dessas informações em investigações científicas e na análise de fenômenos ambientais complexos. No entanto, a dispersão dessas informações, aliada à diversidade de formatos, padrões e níveis de acesso, constitui um desafio significativo para sua integração, recuperação e reutilização.

Sob essa perspectiva, a gestão da informação, no âmbito da Ciência da Informação, assume papel estratégico ao fornecer diretrizes para a organização, sistematização e compartilhamento de dados, favorecendo sua interoperabilidade e uso qualificado. Além disso, princípios contemporâneos como os *FAIR data* (*Findable, Accessible, Interoperable, Reusable*) têm sido amplamente discutidos como fundamentos para a melhoria da gestão de dados científicos, especialmente em contextos de grande volume e diversidade informacional (Wilkinson *et al.*, 2016; Mons *et al.*, 2017).

Paralelamente, o desenvolvimento de tecnologias emergentes, como a Ciência de Dados, amplia significativamente as possibilidades analíticas sobre grandes volumes de dados ambientais. Por meio da aplicação de métodos estatísticos, técnicas de aprendizado de máquina e abordagens computacionais avançadas, torna-se possível identificar padrões, realizar análises preditivas e extrair conhecimento a partir de dados complexos (Kitchin, 2014; Provost; Fawcett, 2013).

Apesar desses avanços, observa-se que ainda são limitadas as abordagens que articulam, de forma integrada, os fundamentos da gestão da informação com as potencialidades da ciência de dados no contexto de plataformas informacionais voltadas à biodiversidade marinha.

De modo geral, a literatura tende a tratar essas dimensões de forma dissociada: enquanto parte dos estudos enfatiza os aspectos técnicos e analíticos da ciência de dados, outros concentram-se na organização, curadoria e compartilhamento da informação, sem estabelecer uma articulação consistente entre esses campos.

Essa fragmentação limita a compreensão integrada dos desafios relacionados

à organização, integração e uso de dados científicos em ambientes informacionais complexos.

Tal lacuna evidencia a necessidade de investigações que não apenas identifiquem e descrevam essas plataformas, mas que também analisem criticamente suas estruturas, funcionalidades e potencial de uso para a produção de conhecimento científico e apoio à tomada de decisão.

Dessa forma, este estudo tem como propósito examinar a função das plataformas informacionais na gestão de dados ambientais marinhos, buscando compreender de que maneira abordagens da ciência de dados podem contribuir para a análise de grandes volumes de informações sobre biodiversidade, ampliando as possibilidades de geração de conhecimento nesse domínio.

2 PROCEDIMENTOS METODOLÓGICOS

Este estudo caracteriza-se como uma pesquisa de natureza exploratória, com abordagem qualitativa, orientada pela análise conceitual e pela investigação de bases de dados ambientais voltadas à fauna marinha. A escolha dessa abordagem justifica-se pela necessidade de compreender, de forma aprofundada, as características, estruturas e desafios associados à organização e ao uso de dados ecológicos em diferentes plataformas informacionais.

A pesquisa foi conduzida a partir de dois eixos metodológicos complementares: (i) revisão de literatura e (ii) levantamento e análise de bases de dados.

No primeiro eixo, realizou-se uma revisão narrativa com caráter estruturado, fundamentada em literatura das áreas de Ciência da Informação, gestão da informação e ciência de dados. A busca bibliográfica foi realizada em bases como Scopus, Web of Science e Google Scholar, utilizando combinações de termos como “*gestão da informação*”, “*biodiversity data*”, “*marine data*”, “*data science*” e “*environmental data management*”, empregadas por meio de operadores booleanos (AND/OR), de modo a abranger estudos relacionados à organização, integração e análise de

dados ambientais.

Foram priorizados trabalhos publicados nos últimos 10 anos, sem desconsiderar autores clássicos da área, com o objetivo de estabelecer um diálogo entre fundamentos teóricos consolidados e abordagens contemporâneas. A seleção dos estudos baseou-se em critérios de relevância temática, aderência aos objetivos da pesquisa e reconhecimento na literatura, contemplando tanto contribuições teóricas quanto aplicações no contexto de dados ambientais e biodiversidade.

Embora não se configure como uma revisão sistemática, o processo de seleção buscou contemplar estudos relevantes e amplamente reconhecidos na literatura, entendidos neste estudo como trabalhos com elevado número de citações e ampla circulação no campo científico, conforme orientações para revisões qualitativas (Snyder, 2019).

No segundo eixo, foi realizado um levantamento exploratório de bases de dados relacionadas à fauna marinha e à biodiversidade aquática, contemplando iniciativas nacionais e internacionais. A identificação das plataformas ocorreu por meio de múltiplas estratégias: consulta à literatura científica, análise de documentos institucionais, relatórios técnicos e acesso a portais eletrônicos mantidos por instituições de pesquisa, organismos governamentais e redes de compartilhamento de dados científicos.

Como critérios de inclusão, foram consideradas plataformas que: (a) disponibilizam dados sobre biodiversidade, com ênfase em fauna marinha; (b) possuem acesso público total ou parcial; e (c) apresentam relevância reconhecida na literatura ou em iniciativas institucionais. Foram excluídas bases de dados com acesso restrito sem descrição pública, bem como aquelas que não apresentavam informações suficientes para análise.

A partir desse processo, procedeu-se à análise qualitativa das plataformas identificadas, com base em um conjunto de categorias analíticas previamente definidas, inspiradas em princípios da gestão da informação e da literatura sobre dados científicos (Wilkinson *et al.*, 2016; Kitchin, 2014). As categorias consideradas

incluíram: (i) escopo temático, (ii) tipos de dados disponibilizados, (iii) estrutura e organização da informação, (iv) formas de acesso e compartilhamento, (v) interoperabilidade e padronização, e (vi) potencial de uso em pesquisas científicas e em processos de gestão ambiental.

A análise dos dados foi conduzida por meio de abordagem descritivo-analítica, buscando identificar padrões, convergências e limitações entre as plataformas investigadas. Para tanto, foi utilizada uma lógica de análise comparativa, permitindo não apenas descrever as características individuais das bases de dados, mas também evidenciar aspectos comuns e desafios recorrentes relacionados à gestão e ao uso de dados ecológicos.

Essa análise foi orientada pelas categorias previamente definidas, as quais subsidiaram a interpretação dos dados à luz dos referenciais da gestão da informação e da ciência de dados.

Por fim, os resultados foram interpretados à luz dos referenciais teóricos da gestão da informação e da ciência da informação, com base nas contribuições de Davenport e Prusak (1998), Kitchin (2014) e nos princípios *FAIR* (Wilkinson *et al.*, 2016), articulando-os com discussões contemporâneas sobre ciência de dados e *Big Data*. Essa etapa analítica buscou compreender de que forma as infraestruturas informacionais investigadas contribuem para a organização, integração e utilização de dados ambientais, bem como suas implicações para a produção de conhecimento científico e para a tomada de decisão em contextos de conservação marinha.

3 PLATAFORMAS INFORMACIONAIS PARA DADOS AMBIENTAIS MARINHOS

A partir do levantamento exploratório realizado, foram identificadas 13 plataformas de dados voltadas ao armazenamento e à disseminação de informações sobre fauna marinha e biodiversidade aquática, mantidas por instituições de pesquisa, programas de monitoramento ambiental e iniciativas internacionais de compartilhamento de dados científicos. Essas plataformas são amplamente utilizadas por pesquisadores, instituições governamentais e organizações não governamentais, configurando-se como infraestruturas relevantes para a produção e circulação de dados ecológicos.

No contexto brasileiro, destacam-se sistemas como o Sistema de Informação de Monitoramento de Mamíferos Marinhos (SIMMAM), o Sistema de Informação de Monitoramento da Biota Aquática (SIMBA), o Banco de Dados para Conservação das Tartarugas Marinha (BDC Tamar), o Banco Nacional de Amostras Biológicas de Albatrozes e Petréis (BAAP) e o Sistema de Informação de Monitoramento de Cetáceo (SISPMC). Essas plataformas concentram registros oriundos de programas específicos de monitoramento ambiental, reunindo informações sobre diferentes grupos taxonômicos, como mamíferos marinhos, tartarugas e aves oceânicas.

Em escala internacional, iniciativas como *Ocean Biodiversity Information System* (OBIS) e *Global Biodiversity Information Facility* (GBIF) ampliam significativamente o acesso a dados sobre biodiversidade, reunindo milhões de registros de ocorrência de espécies provenientes de diversas regiões do mundo. Outras plataformas, como Movebank e Seabird Tracking Database, apresentam foco específico no monitoramento de movimentação animal, disponibilizando dados de telemetria e rotas migratórias.

Adicionalmente, infraestruturas como *World Register of Marine Species* (WoRMS), *European Marine Observation and Data Network* (EMODnet), *Ocean Tracking*

Network (OTN) e (*Integrated Marine Observing System* (IMOS) ampliam as possibilidades de integração entre dados biológicos, taxonômicos e oceanográficos, configurando-se como importantes redes de dados científicos em escala global.

A síntese dessas plataformas, apresentado no Quadro 1, permite evidenciar não apenas suas características individuais, mas também padrões relevantes quando analisadas sob uma perspectiva comparativa.

Quadro 1 – Plataformas de bases de dados utilizadas para monitoramento da fauna marinha.

Plataforma	Escopo	Instituição responsável	Tipo de dados	Acesso	Estudos científicos
SIMMAM	Brasil	Universidades e programas de monitoramento	Monitoramento e enalhes	Parcial	Koether (2024); Barreto <i>et al.</i> (2012)
BDC Tamar	Brasil	Projeto Tamar / ICM-Bio	Reprodução e monitoramento	Institucional	Koether (2024); Marcovaldi <i>et al.</i> (2007)
BAAP	Brasil	Projeto Albatroz e instituições de pesquisa	Amostras biológicas	Institucional	Koether (2024); Sgarioni <i>et al.</i> (2023)
SIMBA	Brasil	Universidades e programas ambientais	Dados ecológicos	Parcial	Koether (2024); Caetano (2022)
SISPMC	Brasil	Programas de monitoramento e instituições de pesquisa	Ocorrência e monitoramento	Institucional	Koether (2024); Maricato (2023)
OBIS	Global	IOC-UNESCO	Ocorrência de espécies	Aberto	De Pooter <i>et al.</i> (2017); Costello <i>et al.</i> (2013)
GBIF	Global	Rede internacional de instituições científicas	Biodiversidade	Aberto	Beck <i>et al.</i> (2014); Meyer <i>et al.</i> (2016); Ivanova (2021)
Movebank	Global	Max Planck Institute e parceiros	Telemetria animal	Parcial	Kranstauber <i>et al.</i> (2011); Kays <i>et al.</i> (2022)

Seabird Tracking	Global	BirdLife International	Rastreamento de aves	Parcial	Bernard <i>et al.</i> (2021); Davies <i>et al.</i> (2021)
WoRMS	Global	Flanders Marine Institute	Taxonomia	Aberto	Costello <i>et al.</i> (2013); Horton <i>et al.</i> (2021)
EMODnet	Regional	Comissão Europeia	Dados oceanográficos	Aberto	Martín Míguez <i>et al.</i> (2019); Schmitt <i>et al.</i> (2019)
OTN	Global	Dalhousie University	Telemetria acústica	Parcial	Hussey <i>et al.</i> (2015); Cooke <i>et al.</i> (2004)
IMOS	Austrália	Governo australiano e instituições científicas	Dados ambientais	Aberto	Lynch <i>et al.</i> (2014); Hill <i>et al.</i> (2010)

Fonte: elaboração própria (2026).

A análise das plataformas, à luz das categorias metodológicas definidas, evidencia diferenças estruturais significativas relacionadas ao escopo, organização da informação, formas de acesso e potencial de uso dos dados.

Em termos de escopo, observa-se a coexistência de plataformas especializadas, vinculadas a programas de monitoramento específicos (como as bases nacionais), e plataformas integradoras, que operam como agregadores globais de dados (como OBIS e GBIF). Essa distinção indica diferentes modelos de produção e governança da informação, nos quais sistemas locais tendem a apresentar maior detalhamento e controle sobre os dados, enquanto plataformas globais priorizam a integração e a ampla disseminação.

De forma complementar, essa diferenciação permite identificar distintos níveis de maturidade informacional entre as plataformas analisadas, especialmente no que se refere à padronização, interoperabilidade e capacidade de integração dos dados, evidenciando que, embora sistemas globais ampliem o alcance e a visibilidade das informações, ainda persistem desafios relacionados à harmonização de dados provenientes de múltiplas fontes.

No que se refere à organização e estruturação da informação, verifica-se que plataformas internacionais apresentam maior grau de padronização e interoperabilidade, frequentemente alinhadas a padrões internacionais de dados e princípios

de compartilhamento científico, como os *FAIR data* (Wilkinson *et al.*, 2016). Por outro lado, bases nacionais, embora ricas em dados primários, tendem a apresentar heterogeneidade em formatos e estruturas, o que pode dificultar sua integração com outras infraestruturas informacionais.

Em relação às formas de acesso, identifica-se variação entre plataformas com dados abertos e amplamente acessíveis, como OBIS e GBIF, e sistemas com acesso parcial ou restrito, frequentemente vinculados a instituições ou programas específicos. Essa diferença impacta diretamente o potencial de reutilização dos dados, evidenciando a importância de políticas de ciência aberta e de gestão da informação orientadas à ampliação do acesso e da transparência (Mons *et al.*, 2017).

Importa esclarecer que os sistemas informacionais podem apresentar diferentes níveis de acesso aos dados. Nos sistemas abertos, as informações são públicas e disponibilizadas sem restrições. Nos sistemas fechados, o acesso é limitado a usuários autorizados, geralmente por razões institucionais, legais ou de segurança. Já os sistemas com acesso parcial ou restrito combinam essas abordagens, ao disponibilizar parte dos dados ao público, enquanto mantêm informações sensíveis acessíveis apenas a usuários autorizados, conforme critérios de controle e gestão.

Quanto ao potencial de uso, observa-se que essas plataformas têm sido amplamente empregadas em estudos científicos voltados à análise de padrões de distribuição de espécies, monitoramento da biodiversidade e avaliação de mudanças ambientais em diferentes escalas. Koether (2024) destaca o uso integrado de bases como SIMMAM, BAAP, SIMBA, BDC Tamar e SISPMC para a consolidação de informações sobre megafauna marinha no Brasil, evidenciando seu papel na organização e no compartilhamento de dados oriundos de múltiplas iniciativas de monitoramento.

No âmbito internacional, bases como OBIS e GBIF têm sido amplamente utilizadas em pesquisas sobre padrões biogeográficos e lacunas de conhecimento em biodiversidade (De Pooter *et al.*, 2017; Beck *et al.* 2014; Ivanova, 2021). Essas plataformas permitem a realização de análises espaciais e temporais em larga escala,

ampliando as possibilidades de investigação científica. Estudos como os de Appeltans *et al.* (2012) e Costello *et al.* (2013) reforçam o papel dessas infraestruturas na integração e padronização de grandes volumes de dados ecológicos, enquanto Beck *et al.* (2014) e Meyer *et al.* (2016) destacam sua relevância para estudos em biodiversidade e conservação.

Apesar dessas contribuições, a literatura também aponta desafios importantes associados ao uso dessas bases. Moura *et al.* (2020) destacam limitações relacionadas à qualidade dos dados, incluindo inconsistências taxonômicas, falhas de georreferenciamento e lacunas de registro. Esses aspectos evidenciam que, além da disponibilidade de dados, a efetividade dessas plataformas depende diretamente de práticas robustas de curadoria, padronização e gestão da informação, reforçando a necessidade de integração entre os princípios da Ciência da Informação e as abordagens da ciência de dados.

De forma geral, os resultados indicam que, embora exista uma infraestrutura significativa de dados sobre biodiversidade marinha, persistem desafios relacionados à interoperabilidade, padronização e acesso, os quais impactam diretamente o potencial de uso desses dados em pesquisas científicas e na formulação de políticas públicas voltadas à conservação ambiental.

Observa-se ainda que tais limitações estão associadas, em grande medida, à heterogeneidade dos dados e à ausência de mecanismos consistentes de integração entre plataformas, evidenciando diferentes níveis de maturidade informacional e restringindo a reutilização ampliada dessas informações.

4 GESTÃO DA INFORMAÇÃO EM DADOS DE BIODIVERSIDADE

A compreensão da gestão da informação em bases de dados ambientais requer, inicialmente, situar o debate no campo da Ciência da Informação. Essa área do conhecimento dedica-se ao estudo dos processos de produção, organização, armazenamento, recuperação, disseminação e uso da informação em diferentes

contextos sociais, científicos e institucionais. Para Saracevic (1999), a Ciência da Informação investiga as propriedades e o comportamento da informação, bem como os fluxos e processos que envolvem sua comunicação e utilização, constituindo-se como um campo interdisciplinar que articula contribuições da biblioteconomia, da ciência da computação e de outras áreas voltadas à compreensão dos fenômenos informacionais.

No interior desse campo, a Gestão da Informação (GI) emerge como um conjunto de práticas e estratégias orientadas à organização e ao uso eficiente dos recursos informacionais. De acordo com Choo (2006), a GI envolve processos estruturados destinados a identificar necessidades informacionais, adquirir dados relevantes, organizá-los de maneira adequada e disponibilizá-los para apoiar a tomada de decisão e a produção de conhecimento. Nessa perspectiva, a informação passa a ser compreendida como um recurso estratégico que, quando devidamente estruturado e contextualizado, contribui para a geração de valor científico e institucional.

Diversos autores destacam que a GI pode ser compreendida como um ciclo informacional composto por etapas interdependentes, envolvendo atividades como coleta, criação ou aquisição, organização, armazenamento, disseminação e uso estratégico da informação (Davenport; Prusak, 1998; Wang; Zhang, 2021). Esse ciclo permite transformar dados dispersos em informações significativas, favorecendo a circulação do conhecimento e sua aplicação em diferentes contextos organizacionais e científicos.

No contexto dos dados científicos, especialmente aqueles associados à biodiversidade, aspectos como qualidade, padronização e interoperabilidade tornam-se elementos centrais da GI. Conforme discutido por Michener (2015), a organização adequada de dados científicos é fundamental para garantir sua preservação, compartilhamento e reutilização. Nesse sentido, a adoção de padrões, metadados estruturados e práticas de curadoria de dados constitui um requisito essencial para assegurar a confiabilidade e a integridade das informações produzidas.

Avanços recentes no campo da gestão de dados científicos têm enfatizado

a importância de princípios como os *FAIR data*, que orientam a organização e o compartilhamento de dados de forma a maximizar sua reutilização em diferentes contextos (Wilkinson *et al.*, 2016; Mons *et al.*, 2017). Esses princípios reforçam a necessidade de estruturas informacionais que promovam não apenas o armazenamento de dados, mas também sua integração e interoperabilidade em ambientes digitais complexos.

Quando aplicada ao contexto da biodiversidade, a gestão da informação assume um papel ainda mais estratégico. A produção de dados ecológicos provenientes de programas de monitoramento ambiental, inventários biológicos e pesquisas científicas gera grandes volumes de informações heterogêneas, que demandam processos estruturados de organização e integração. Nesse caso, sistemas de informação e plataformas digitais configuram-se como infraestruturas essenciais para reunir registros de ocorrência de espécies, dados ambientais e metadados associados, possibilitando análises em diferentes escalas espaciais e temporais.

Entretanto, conforme evidenciado na análise das plataformas apresentada na seção anterior, a existência dessas infraestruturas não garante, por si só, a efetiva integração e reutilização dos dados. Diferenças nos padrões de estruturação, nos modelos de governança e nas políticas de acesso podem limitar o potencial de uso dessas informações, evidenciando a necessidade de práticas mais robustas de gestão da informação que considerem aspectos técnicos, organizacionais e institucionais.

Dessa forma, a gestão da informação em dados de biodiversidade não se restringe à organização de bases de dados, mas envolve a articulação entre padrões, tecnologias, políticas de acesso e práticas de curadoria, com vistas a ampliar o acesso, a integração e a reutilização de informações científicas. Esse processo favorece a realização de estudos ecológicos, análises de padrões biogeográficos e a formulação de estratégias de conservação ambiental, ao mesmo tempo em que fortalece a transparência científica e a colaboração entre instituições de pesquisa, organismos governamentais e iniciativas internacionais.

5 CIÊNCIA DE DADOS E ANÁLISE DE GRANDES VOLUMES DE DADOS

A Ciência de Dados (CD) tem se consolidado, nas últimas décadas, como um campo central para a análise de grandes volumes de informação produzidos em diferentes áreas do conhecimento. De modo geral, esse campo pode ser compreendido como um conjunto de abordagens analíticas, computacionais e estatísticas voltadas à extração de conhecimento a partir de dados (Macgillivray, 2021).

Dhar (2013) destaca que a Ciência de Dados configura-se como um campo interdisciplinar que articula áreas como Estatística, Ciência da Computação, Ciência da Informação e aprendizado de máquina, permitindo a análise de fenômenos complexos por meio da integração de diferentes métodos e perspectivas analíticas.

No plano conceitual, a CD não se limita ao uso de ferramentas computacionais, mas envolve um processo estruturado de investigação orientado por dados. Esse processo inclui a formulação de problemas analíticos, a seleção de métodos adequados, o tratamento e a modelagem dos dados, bem como a interpretação dos resultados obtidos (Yu; Kumbier, 2020). Dessa forma, o trabalho do cientista de dados ultrapassa a dimensão técnica, exigindo competências analíticas e epistemológicas para a construção de conhecimento a partir de dados.

Entre as habilidades associadas a esse campo destacam-se a capacidade de formular problemas, aplicar métodos estatísticos, identificar correlações entre variáveis e empregar técnicas de aprendizado de máquina e mineração de dados (Dhar, 2013). Tais competências permitem a identificação de padrões, tendências e relações em grandes volumes de dados, contribuindo para a produção de conhecimento em diferentes domínios científicos.

O avanço da Ciência de Dados está diretamente relacionado ao fenômeno do *Big Data*, caracterizado pelo crescimento exponencial da produção de dados e pelo desenvolvimento de tecnologias capazes de armazenar e processar grandes volumes de informação. Esse fenômeno é frequentemente descrito a partir de di-

mensões como volume, variedade e velocidade, às quais se somam aspectos como veracidade e valor, ampliando os desafios associados ao tratamento e à análise de dados complexos (Kitchin, 2014; Marr, 2016).

No contexto da ciência contemporânea, observa-se um processo crescente de transformação de diferentes dimensões da realidade em dados. Floridi (2014) descreve esse fenômeno como “datificação”, no qual aspectos sociais, científicos e ambientais passam a ser convertidos em dados passíveis de análise computacional. De modo convergente, Frické (2015) caracteriza esse contexto como um “dilúvio de dados”, que redefine os modos de produção do conhecimento científico ao ampliar significativamente a capacidade de geração, armazenamento e processamento de informações.

No âmbito dos estudos sobre biodiversidade, essas transformações assumem particular relevância. A disponibilidade de grandes volumes de dados provenientes de plataformas informacionais, como as analisadas neste estudo, cria condições favoráveis para a aplicação de métodos da CD na análise de padrões ecológicos, distribuição de espécies e mudanças ambientais. Técnicas como modelagem estatística, aprendizado de máquina e análise espacial têm sido empregadas para explorar dados provenientes de diferentes fontes, possibilitando análises em múltiplas escalas e contribuindo para o avanço do conhecimento sobre ecossistemas marinhos (Reichstein *et al.*, 2019).

Entretanto, a aplicação da CD nesse contexto está diretamente condicionada à qualidade, organização e integração dos dados disponíveis. Problemas relacionados à heterogeneidade, incompletude e inconsistência das bases de dados podem comprometer a confiabilidade das análises e a interpretação dos resultados. Assim, reforça-se a interdependência entre CD e GI, na medida em que o potencial analítico dos dados depende de práticas adequadas de curadoria, padronização e organização informacional.

Dessa forma, a CD não deve ser compreendida de maneira isolada, mas como parte de um ecossistema mais amplo de produção e gestão da informação. No caso

dos dados de biodiversidade, sua efetividade está diretamente associada à existência de infraestruturas informacionais robustas e a práticas consolidadas de gestão da informação, capazes de garantir a qualidade, a interoperabilidade e a reutilização dos dados em diferentes contextos científicos.

6 DESAFIOS INFORMACIONAIS E PERSPECTIVAS

O crescimento das plataformas digitais voltadas à biodiversidade marinha tem ampliado significativamente a disponibilidade de dados ambientais provenientes de programas de monitoramento, pesquisas científicas e iniciativas de ciência aberta. Essas infraestruturas informacionais desempenham papel estratégico ao permitir a agregação e o compartilhamento de registros oriundos de diferentes instituições, métodos de coleta e escalas geográficas, ampliando o potencial de análise e reutilização dos dados científicos.

Iniciativas globais de compartilhamento de dados de biodiversidade têm contribuído para expandir o acesso a informações ecológicas em larga escala, favorecendo estudos sobre distribuição de espécies e padrões biogeográficos (Costello *et al.*, 2013; Appeltans *et al.*, 2012). Entretanto, o aumento do volume, da diversidade e da complexidade dessas informações evidencia desafios persistentes relacionados à organização, padronização e interoperabilidade dos dados, especialmente quando considerados contextos institucionais e tecnológicos heterogêneos.

A gestão da informação assume, nesse cenário, papel central ao possibilitar a estruturação de processos voltados à coleta, organização, armazenamento, disseminação e uso de dados científicos. Conforme discutem Davenport e Prusak (1998), a gestão da informação envolve práticas destinadas a transformar dados dispersos em informações úteis para processos analíticos e decisórios. No entanto, conforme evidenciado na análise das plataformas apresentada anteriormente, a efetividade desses processos depende não apenas da existência de bases de dados, mas da adoção de padrões, políticas de acesso e estratégias de governança informacional.

No campo da biodiversidade, tais desafios tornam-se mais complexos em função da heterogeneidade das fontes de dados, da diversidade de formatos e da necessidade de integração entre diferentes sistemas e plataformas informacionais. Estudos indicam que lacunas, inconsistências taxonômicas e vieses espaciais ainda representam limitações relevantes para a utilização dessas bases em pesquisas científicas (Meyer *et al.*, 2016; Moura *et al.*, 2020), evidenciando que a disponibilidade de dados não implica, necessariamente, sua qualidade ou adequação para análises avançadas.

Outro desafio relevante refere-se à articulação entre infraestruturas de dados, políticas de compartilhamento e práticas de governança da informação científica. A consolidação de padrões internacionais de metadados e protocolos de interoperabilidade tem sido apontada como elemento fundamental para ampliar a integração entre diferentes bases de dados (Wilkinson *et al.*, 2016; Mons *et al.*, 2017). A adoção de princípios de ciência aberta e de estratégias de curadoria de dados contribui para promover maior transparência, rastreabilidade e reutilização da informação científica (Michener, 2015).

Além disso, a crescente incorporação de métodos de CD impõe novos desafios relacionados à preparação, integração e qualidade dos dados. A aplicação de técnicas avançadas de análise, como aprendizado de máquina e modelagem preditiva, depende diretamente da consistência e da padronização das bases de dados, reforçando a interdependência entre gestão da informação e capacidade analítica (Reichstein *et al.*, 2019).

Dessa forma, observa-se que a integração entre plataformas informacionais, práticas de gestão da informação e métodos de ciência de dados constitui um elemento estratégico para o avanço das pesquisas sobre biodiversidade marinha. No entanto, essa integração ainda se apresenta como um desafio, exigindo não apenas soluções tecnológicas, mas também alinhamento institucional, definição de políticas de dados e fortalecimento de práticas colaborativas entre diferentes atores envolvidos na produção e no uso da informação científica.

O Quadro 2 sistematiza os principais desafios informacionais identificados na literatura, relacionando-os a seus impactos e às possíveis perspectivas de enfrentamento, com base nas dimensões discutidas ao longo deste estudo.

Quadro 2 - Desafios informacionais e perspectivas na gestão e análise de dados

Tema	Desafio	Impactos	Perspectivas
Crescimento da produção de dados	Crescimento exponencial da geração de dados científicos e institucionais	Sobrecarga informacional, aumento da complexidade analítica e dificuldades de processamento	Uso de infraestruturas de <i>Big Data</i> e métodos de CD
Qualidade e confiabilidade dos dados	Presença de lacunas, inconsistências e vieses nos conjuntos de dados	Comprometimento da confiabilidade das análises científicas e dos processos decisórios	Desenvolvimento de práticas de curadoria e validação e padronização de dados
Interoperabilidade e integração	Dificuldade de integração entre diferentes bases de dados e plataformas	Fragmentação da informação e limitação da reutilização de dados	Adoção de padrões de metadados e protocolos de interoperabilidade
Governança e gestão da informação	Fragilidade ou ausência de políticas institucionais para gestão e compartilhamento de dados	Riscos à integridade da informação, à conformidade e ao uso ético dos dados	Implementação de políticas de governança e gestão da informação
Competências profissionais	Necessidade de profissionais com habilidades interdisciplinares	Limitações na análise, interpretação e uso estratégico de grandes volumes de dados	Formação de cientistas de dados e especialistas em gestão da informação
Inovação tecnológica	Rápida evolução de ferramentas de análise e processamento de dados	Descompasso entre capacidades tecnológicas e práticas institucionais	Integração de inteligência artificial e aprendizado de máquina em processos informacionais
Cultura de dados	Resistência institucional ao compartilhamento e à reutilização de dados	Redução da colaboração científica e limitação do uso estratégico da informação	Fortalecimento da ciência aberta e da cultura orientada a dados

Fonte: elaboração própria (2026).

A análise desses desafios evidencia que a gestão, integração e análise de dados científicos demandam abordagens cada vez mais articuladas entre diferentes áreas do conhecimento. A consolidação de práticas de gestão da informação, aliada ao avanço de métodos de ciência de dados, tende a ampliar as possibilidades de organização, interpretação e reutilização de grandes volumes de dados ambientais.

Nesse sentido, torna-se relevante avançar para modelos mais integrados de governança da informação científica, capazes de articular padrões técnicos, políticas de acesso e práticas colaborativas. Tais perspectivas contribuem para o fortalecimento das infraestruturas informacionais e para o desenvolvimento de pesquisas mais robustas no campo da biodiversidade marinha, especialmente no que se refere à produção de conhecimento aplicável à conservação e à gestão dos ecossistemas.

6 CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo discutir as relações entre plataformas informacionais, gestão da informação e ciência de dados no contexto dos dados ambientais marinhos, com ênfase nos desafios associados à organização, integração e análise de grandes volumes de dados científicos. A partir de uma abordagem exploratória e qualitativa, fundamentada na análise conceitual da literatura e na caracterização de plataformas de dados, buscou-se compreender como esses campos se articulam na produção, organização e uso de informações sobre biodiversidade marinha. De modo geral, os achados indicam que o objetivo proposto foi atingido, ao evidenciar as relações entre essas dimensões.

Os resultados mostram que a crescente produção de dados ambientais tem ampliado a relevância das infraestruturas informacionais na organização e disseminação dessas informações. As plataformas analisadas configuram-se como componentes fundamentais para a agregação e disponibilização de dados de múltiplas fontes, possibilitando a construção de bases mais amplas e potencialmente

integradas. Entretanto, observam-se níveis distintos de maturidade informacional, especialmente quanto à padronização, interoperabilidade e capacidade de integração dos dados.

A análise indica que a disponibilidade de dados e a existência de plataformas não garantem sua efetiva integração e utilização científica. Sua efetividade depende da adoção de práticas consistentes de gestão da informação, relacionadas à coleta, organização, padronização, curadoria e disseminação dos dados. Persistem limitações estruturais que restringem o pleno aproveitamento dessas infraestruturas, evidenciando a centralidade da gestão da informação para assegurar qualidade, interoperabilidade e reutilização.

No que se refere à ciência de dados, verifica-se que o avanço de métodos analíticos amplia as possibilidades de exploração de grandes volumes de dados ambientais, permitindo a identificação de padrões e relações complexas. Contudo, esse potencial depende diretamente da qualidade e organização das bases de dados, evidenciando a interdependência entre capacidade analítica e gestão da informação.

A principal contribuição do estudo reside na análise integrada entre plataformas informacionais, gestão da informação e ciência de dados, demonstrando que o potencial científico dos dados de biodiversidade marinha não depende apenas de avanços tecnológicos, mas da articulação entre infraestruturas, práticas de gestão e métodos analíticos. Ao explicitar essa interdependência, o estudo contribui para o campo da Ciência da Informação ao aproximar, de forma estruturada, essas dimensões no contexto dos dados ambientais.

Como limitações, destaca-se a abordagem exploratória e qualitativa, baseada na literatura e na caracterização de plataformas, sem aplicação empírica de métodos de ciência de dados ou avaliação quantitativa dessas infraestruturas. A ausência de métricas para avaliação da qualidade, interoperabilidade e uso dos dados também representa uma limitação relevante.

Em termos de perspectivas futuras, recomenda-se o desenvolvimento de estudos empíricos voltados à aplicação de técnicas de ciência de dados em bases

de biodiversidade marinha, bem como a realização de análises comparativas mais aprofundadas sobre interoperabilidade, qualidade e governança de dados. Destaca-se, ainda, a necessidade de construção e validação de indicadores operacionais e mensuráveis de maturidade informacional, capazes de quantificar, de forma objetiva, aspectos como padronização, completude, consistência e capacidade de integração entre plataformas. Tais indicadores poderão subsidiar avaliações empíricas mais robustas, contribuindo para o aprimoramento das infraestruturas informacionais e para o uso mais efetivo dos dados no contexto científico e ambiental.

Por fim, conclui-se que a articulação entre plataformas informacionais, práticas de gestão da informação e métodos de ciência de dados constitui elemento estratégico para o avanço das pesquisas em biodiversidade marinha, ao fortalecer as infraestruturas informacionais e ampliar a capacidade analítica dos dados, com implicações para a conservação e gestão dos ecossistemas marinhos, indicando a necessidade de abordagens integradas que favoreçam o aprimoramento das plataformas de dados e o uso científico qualificado das informações ambientais.

REFERÊNCIAS

- APPELTANS, W. *et al.* The magnitude of global marine species diversity. **Current Biology**, Cambridge, v. 22, n. 23, p. 2189-2202, 2012. Disponível em: <https://doi.org/10.1016/j.cub.2012.09.036>. Acesso em: 03 mar. 2026.
- BARRETO, A. S. *et al.* SIMMAM – Sistema de Apoio ao Monitoramento de Mamíferos Marinhos: uma nova ferramenta para a gestão ambiental. *In*: RIO OIL & GAS EXPO AND CONFERENCE, 212, Rio de Janeiro. **Anais [...]**. Rio de Janeiro de Janeiro: Instituto Brasileiro de Petróleo, Gás e Biocombustíveis, 2012. Disponível em: https://www.researchgate.net/publication/289523408_SIMMAM_-_SISTEMA_DE_APOIO_AO_MONITORAMENTO_DE_MAMIFEROS_MARINHOS_UMA_NOVA_FERRAMENTA_PARA_A_GESTAO_AMBIENTAL. Acesso em: 15 mar. 2026.
- BECK, J. *et al.* Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. **Ecological Informatics**, Amsterdam, v. 19, p. 10-15, 2014. Disponível em: <https://doi.org/10.1016/j.ecoinf.2013.11.002>. Acesso em: 10 mar. 2026.
- BERNARD, A. *et al.* Toward a global strategy for seabird tracking. **Conservation Letters**, [s. l.], v. 14, n. 3, p. e12804, 2021. Disponível em: <https://doi.org/10.1111/conl.12804>. Acesso em: 04 mar. 2026.
- CAETANO, P. B. **Ingestão plástica em aves marinhas necropsiadas pelo Projeto de Monitoramento de Praias da Baía de Santos (PMP-BS)**. 2022. Dissertação (Mestrado em Ciências Ambientais) – Universidade Federal de São Carlos, São Carlos, 2022. Disponível em: <https://repositorio.ufscar.br/server/api/core/bitstreams/c92dc639-0d7f-46de-a402-63d368f7a8be/content>. Acesso em: 15 mar. 2026.
- CHOO, C. W. **The knowing organization**: how organizations use information to construct meaning, create knowledge and make decisions. New York: Oxford University Press, 2006. DOI: [https://doi.org/10.1016/0268-4012\(96\)00020-5](https://doi.org/10.1016/0268-4012(96)00020-5). Acesso em 25 fev. 2026.

COOKE, S. J. *et al.* Biotelemetry: a mechanistic approach to ecology. **Trends in Ecology & Evolution**, Amsterdã, v. 19, n. 6, p. 334-343, 2004. Disponível em: <https://doi.org/10.1016/j.tree.2004.04.003>. Acesso em: 19 mar. 2026.

COSTELLO, M. J. *et al.* Global coordination and standardisation in marine biodiversity through the Ocean Biogeographic Information System (OBIS). **Marine Ecology Progress Series**, San Francisco, v. 487, p. 203-213, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0051629>. Acesso em 12 mar. 2026.

DATE, C. J. **Introdução a sistemas de bancos de dados**. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier Brasil: Campus, 2004.

DAVENPORT, T. H.; PRUSAK, L. **Working knowledge**: how organizations manage what they know. Boston: Harvard Business School Press, 1998.

DAVIES, T. E. *et al.* Multispecies tracking reveals a major seabird hotspot in the North Atlantic. **Conservation letters**, [s. l.], v. 14, n. 5, p. e12824, 2021. Disponível em <https://doi.org/10.1111/conl.12824>. Acesso em: 01 mar. 2026.

DE POOTER, D. *et al.* Toward a new data standard for combined marine biological and environmental datasets: expanding OBIS beyond species occurrences. **Biodiversity Data Journal**, Sofia, v. 5, e10989, 2017. Disponível em: <https://doi.org/10.3897/BDJ.5.e10989>. Acesso em: 03 mar. 2026.

DHAR, V. Data science and prediction. **Communications of the ACM**, New York, v. 56, n. 12, p. 64–73, dec. 2013. Disponível em: <https://doi.org/10.1145/2500499>. Acesso em: 03 mar. 2026.

ELMASRI, R.; NAVATHE, S. B. **Fundamentals of database systems**. 7th ed. Boston: Pearson, 2016.

FLORIDI, L. **The fourth revolution**: how the infosphere is reshaping human reality. Oxford: Oxford University Press, 2014.

FRICKE, M. Big Data and its epistemology. **Journal of the Association for Information Science and Technology**, [s. l.], v. 66, n. 4, p. 651-661, 2015. Disponível em: <https://doi.org/10.1002/asi.23212>. Acesso em: 10 mar. 2026.

HILL, K. L. *et al.* Observing systems and oceanographic data integration. **Annual Review of Marine Science**, Palo Alto, v. 2, p. 305-335, 2010. Disponível em: <https://doi.org/10.1146/annurev-marine-120308-081135>. Acesso em: 19 mar. 2026.

HORTON, T. *et al.* World Register of Marine Species (WoRMS): an authoritative global database for marine biodiversity. **Nucleic Acids Research**, Oxford, v. 49, n. D1, p. D984-D988, 2021. Disponível em: <https://doi.org/10.1093/nar/gkaa1081>. Acesso em: 19 mar. 2026.

HUSSEY, N. E. *et al.* Aquatic animal telemetry: a panoramic window into the underwater world. **Science**, Washington, v. 348, n. 6240, p. 1255642, 2015. Disponível em: <https://doi.org/10.1126/science.1255642>. Acesso em: 19 mar. 2026.

IVANOVA, N. V.; SHASHKOV, M. P. The possibilities of GBIF data use in ecological research. **Russian Journal of Ecology**, [s. l.], v. 52, n. 1, p. 1-8, 2021. Disponível em: <https://doi.org/10.1134/S1067413621010069>. Acesso em: 09 mar. 2026.

KAYS, R. *et al.* The Movebank system for studying global animal movement and demography. **Methods in Ecology and Evolution**, [s. l.], v. 13, n. 2, p. 419-431, 2022. Disponível em: <https://doi.org/10.1111/2041-210X.13767>. Acesso em: 18 mar. 2026.

KITCHIN, R. **The data revolution: big data, open data, data infrastructures and their consequences.** Londres: SAGE Publications, 2014.

KOETHER, B. **Análise e integração de bancos de dados nacionais para a megafauna marinha: contribuições para as políticas públicas ambientais**. 2024. Dissertação (Mestrado em Sistemas Costeiros e Oceânicos) – Universidade Federal do Paraná, Centro de Estudos do Mar, Campus Pontal do Paraná, Pontal do Paraná, 2024. Disponível em: <https://acervodigital.ufpr.br/xmlui/handle/1884/94526>. Acesso em: 08 mar. 2026.

- KRANSTAUER, B. *et al.* The Movebank data model for animal tracking. **Environmental Modelling & Software**, [s. l.], v. 26, n. 6, p. 834-835, 2011. Disponível em: <https://doi.org/10.1016/j.envsoft.2010.12.005>. Acesso em: 19 mar. 2026.
- LYNCH, T. P. *et al.* The Australian Integrated Marine Observing System (IMOS): delivering data streams to address national and international research priorities. **Marine and Freshwater Research**, Melbourne, v. 65, n. 7, p. 569-576, 2014. Disponível em: <https://doi.org/10.1071/MF13361>. Acesso em: 19 mar. 2026.
- MACGILLIVRAY, H. Statistics and data science must speak together. **Teaching Statistics**, [s. l.], v. 43, n. S1, p. S5-S10, 2021. Disponível em: <https://doi.org/10.1111/test.12281>. Acesso em: 02 mar. 2026.
- MARCOVALDI, M. A. *et al.* Fifteen Years of Hawksbill Sea Turtle (*Eretmochelys imbricata*) Nesting in Northern Brazil. **Chelonian Conservation and Biology**, Massachusetts, v. 6, n. 2, p. 223-228, 2007. Disponível em: https://www.tamar.org.br/publicacoes_html/pdf/2007/2007_Marcovaldi_et_al_Eretmochelys_Brasil_CCB_2007.pdf. Acesso em: 19 mar. 2026.
- MARICATO, G. A. B. **Baleias e golfinhos em águas costeiras e oceânicas**: uma abordagem integrada para distribuição, exposição a atividades humanas e conservação das espécies no Brasil e no mundo. 2023. Tese (Doutorado em Oceanografia) – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2023. Disponível em: <http://www.bdtd.uerj.br/handle/1/20806>. Acesso em: 10 mar. 2026.
- MARR, B. **Big data in practice**: how 45 successful companies used big data analytics to deliver extraordinary results. Chichester: Wiley, 2016.
- MARTÍN MÍGUEZ, B. *et al.* The European marine observation and data network (EMODnet): visions and roles of the gateway to marine data in Europe. **Frontiers in Marine Science**, [s. l.], v. 6, p. 313, 2019. Disponível em: <https://doi.org/10.3389/fmars.2019.00313>. Acesso em 01 mar. 2026.
- MEYER, C. *et al.* Multidimensional biases, gaps and uncertainties in global plant occurrence information. **Ecology Letters**, [s. l.], v. 19, n. 8, p. 992-1006, 2016. Disponível em: <https://doi.org/10.1111/ele.12624>. Acesso em: 15 mar. 2026.
- MICHENER, W. K. **Ecological data sharing. Ecological Informatics**, [s. l.], v. 29, p. 33-44, 2015. Disponível em: <https://doi.org/10.1016/j.ecoinf.2015.06.010>. Acesso em: 01 mar. 2026.
- MONS, B. *et al.* Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. **Information Services & Use**, Amsterdã, v. 37, n. 1, p. 49-56, 2017. Disponível em: <https://doi.org/10.3233/ISU-170824>. Acesso em: 19 mar. 2026.
- MOURA, M. R. *et al.* Evaluating the suitability of biodiversity databases for macroecological studies: a conceptual framework for data quality. **Ecography**, [s. l.], v. 43, n. 4, p. 1-13, 2020.
- PROVOST, F.; FAWCETT, T. **Data science for business**: what you need to know about data mining and data-analytic thinking. Sebastopol: O'Reilly Media, 2013.
- RAMAKRISHNAN, R.; GEHRKE, J. **Sistemas de gerenciamento de banco de dados**. 3. ed. São Paulo: AMGH, 2008.
- REICHSTEIN, M. *et al.* Deep learning and process understanding for data-driven Earth system science. **Nature**, Londres, v. 566, n. 7743, p. 195-204, 2019. Disponível em: <https://doi.org/10.1038/s41586-019-0912-1>. Acesso em: 19 mar. 2026.
- SARACEVIC, T. Information science. **Journal of the American Society for Information Science**, [s. l.], v. 50, n. 12, p. 1051-1063, 1999.
- SCHMITT, T. *et al.* EMODnet Bathymetry: a compilation of bathymetric data in the European waters. In: OCEANS 2019 - MARSEILLE. 219, Marseille. **Proceedings** [...]. Piscataway: IEEE, 2019. p. 1-7. Disponível em: <https://doi.org/10.1109/OCEANSE.2019.8867250>. Acesso em: 20 mar. 2026.
- SGARIONI, A. Z. *et al.* A novel variant of Babesia sp. (Piroplasmida) as a hemoparasite in procellariiform seabirds. **Parasitology Research**, [s. l.], v. 122, n. 8, p. 1935-1941, 2023. Disponível em: <http://doi.org/10.1007/s00436-023-07894-4>. Acesso em 19 mar. 2026.
- SNYDER, H. Literature review as a research methodology: an overview and guidelines. **Journal of**

Business Research, Amsterdã, v. 104, p. 333-339, 2019. Disponível em: <https://doi.org/10.1016/j.jbusres.2019.07.039>. Acesso em: 19 mar. 2026.

WANG, Y.; ZHANG, Y. Fifty years of information management research: a conceptual structure analysis using structural topic modeling. **International Journal of Information Management**, [s. l.], v. 58, 102316, 2021. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2021.102316>. Acesso em: 08 fev. 2026.

WILKINSON, M. D. *et al.* The FAIR Guiding principles for scientific data management and stewardship. **Scientific Data**, Londres, v. 3, n. 1, p. 1-9, 2016. Disponível em: <https://doi.org/10.1038/sdata.2016.18>. Acesso em: 09 mar. 2026.

YU, B.; KUMBIER, K. Veridical data science. **Proceedings of the National Academy of Sciences**, [s. l.], v. 117, n. 8, p. 3920–3929, 2020. Disponível em: <https://doi.org/10.1073/pnas.1901326117>. Acesso em: 07 mar. 2026.