

DATA PAPERS E DATA JOURNALS SOBRE BIODIVERSIDADE considerações acerca da preservação digital

Analysis of Data Papers Publications on Biodiversity considerations about digital preservation



Resumo

Introdução: A estruturação dos conjuntos de dados sobre Biodiversidade está sendo divulgada em uma linguagem reservada para a descrição do substrato da comunicação científica denominada Data Papers, isto é, os dados que sustentam pesquisas científicas nesse campo do conhecimento, independentemente do modelo tradicional de comunicação científica. **Objetivo:** Analisar as publicações em formato de Data Papers no campo da Biodiversidade em âmbito internacional. **Metodologia:** Pesquisa documental de abordagem qualitativa e aplica técnicas para coleta e exame das informações por meio de Análise de Conteúdo. Verifica a situação de 33 revistas apontados pela Global Biodiversity Information Facility (GBIF) que oferecem publicações em formato de Data Papers. Identifica-se: os temas correlatos à biodiversidade; os tipos de licenças, indexadores, a quantidade de Data Papers publicados, os títulos que possuem acesso aberto ou fechado, as revistas que mais publicam Data Papers sobre Biodiversidade e o idioma que foram publicados. **Resultados:** O número em Data Papers teve crescimento exponencial entre 2017 até maio de 2022 logo, os artigos sobre o campo da Biodiversidade também têm aumentado em diversos temas que envolvem todo o seu ecossistema. **Conclusão:** Os Data Papers analisados se caracterizam como documentos revisados por pares e representam conjuntos de dados indexados com padrões de metadados adequados para preservar digitalmente os dados registrados nas revistas que foram contempladas na presente análise.

Palavras-chave: Preservação digital. Comunicação científica. *Data Papers*. Biodiversidade.

Abstract

Introduction: The structuring of data sets on Biodiversity are being disseminated in a language reserved for describing the substrate of scientific communication called Data Papers, that is, the data that support scientific research in this field of knowledge, regardless of the traditional model of scientific communication. **Objective:** To analyze the publications in Data Papers format in the field of Biodiversity at the international level. **Methodology:** The study is characterized as documentary research with a qualitative approach and applies techniques for collecting and examining information through Content Analysis. It verifies the situation of 33 journals listed by the Global Biodiversity Information Facility (GBIF) that offer publications in Data Papers format. It identifies: the themes related to biodiversity, the types of licenses, indexers, the number of Data Papers published, the titles that have open or closed access, the journals that most publish Data Papers on Biodiversity and the language in which they were published. **Results:** The number of Data Papers in the field of Biodiversity has increased steadily in several topics involving its entire ecosystem, indicating that they are being accepted by more and more researchers. **Conclusion:** The Data Papers analyzed are characterized as peer-reviewed documents and represent indexed datasets with metadata standards suitable for digitally preserving the data recorded in the journals that were included in this analysis.

Keywords: Digital preservation. Scientific communication. *Data Papers*. Biodiversity.



Fabiano C. Corrêa da Silva

Universidade Federal do Rio Grande do Sul
E-mail: fabianocc@gmail.com
Porto Alegre – RS / Brasil



Ketlen Stueber

Universidade Federal do Rio Grande do Sul
E-mail: ketlstueber@hotmail.com
Porto Alegre – RS / Brasil



Amanda Santos Witt

Universidade Federal do Rio Grande do Sul
E-mail: amandawitt.asw@gmail.com
Porto Alegre – RS / Brasil

RBP

Revista Brasileira de
Preservação Digital

RBP

Brazilian Journal of
Digital PreservationCREDIT¹

• Conceituação	SILVA, F. C. C da.
• Curadoria de dados	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S.
• Análise formal	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S.
• Investigação	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S.
• Metodologia	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S.
• Administração de projetos	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S.
• Supervisão	SILVA, F. C. C da.; STUEBER, K.
• Visualização	SILVA, F. C. C da.; STUEBER, K.; WITT, A. S..
• Redação – rascunho original	SILVA, F. C. C da.; STUEBER, K.



LICENÇA DE USO

Os autores cedem à [Revista Brasileira de Preservação Digital](#) os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a Licença Creative Commons Attribution (CC BY) 4.0 International. Esta licença permite que terceiros remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

PUBLISHERS

Universidade Estadual de Campinas – Sistema de Bibliotecas / Instituto Brasileiro de Informação em Ciência e Tecnologia – Rede Brasileira de Serviços de Preservação Digital – Cariniana. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES

Gildenir Carolino Santos, Miguel Angel Márdero Arellano.

Submetido em: 01/06/2022 – Aceito em: 10/06/2022 – Publicado em: 12/07/2022

¹ Sobre o CRediT, consulte o site e conheça outros papéis: <https://casrai.org/credit/>

1 Introdução

Bio significa "vida" e *diversidade* significa "variedade". Esta área do conhecimento reúne todas as formas de vida, desde plantas, microorganismos, animais, insetos e toda a variedade de genes contidos em cada ser vivo, além dos ecossistemas em que uma espécie afeta a outra. A variedade de temas relacionados ao campo de estudos da Biodiversidade coincide com o crescente volume de dados científicos gerados em pesquisas e a sua preservação digital em repositórios institucionais e temáticos.

Notadamente, os estudos relacionados ao campo da Biodiversidade estão concentrados em modelos de comunicação científica canalizados por revistas que seguem o formato de artigos e *papers*. Esse modelo serve tanto para a publicação de resultados de pesquisas, quanto para a consulta de cientistas que dedicam seu tempo lendo artigos científicos em revistas especializadas.

Um artigo de dados é um tipo de publicação que surgiu como mecanismo para incentivar a publicação de dados sobre qualquer pesquisa. Normalmente, são estruturados de acordo com um padrão internacional de metadados (informações que dão contexto aos dados). A estruturação dos conjuntos de dados com base em padrões de descrição de metadados é utilizada para facilitar a comunidade de editores de conjuntos de dados em todo o mundo e facilitar a obtenção do reconhecimento apropriado para os autores.

No campo da Biodiversidade, onde os dados estão relativamente dispersos em plataformas mais ou menos conectadas, e alimentados por diversos autores, as publicações em formato de Data Papers solucionam a preservação digital adequada dos conjuntos de dados produzidos em larga escala. Um dos benefícios mais expressivos deste formato de publicação talvez seja o acolhimento dos dados brutos de uma pesquisa, disponibilizando a aproximação mais possível da origem de uma análise. Torna-se, assim, uma fonte orgânica, por meio de um formato permanentemente disponível para novas pesquisas.

2 Revisão de literatura

Os dados científicos, ou dados de pesquisa, são importantes fontes de estudo. Grácio e Arellano (2020) propõem um modelo processual para preservação digital de dados de pesquisa por meio de um plano de curadoria digital para conjunto de dados. Os elementos principais para a preservação dos dados de pesquisa envolvem aspectos organizacionais, legais e técnicos. Deve-se considerar: os objetivos da instituição que adota a política de preservação de dados; os recursos humanos e respectivas responsabilidades diante das etapas da curadoria digital; os recursos financeiros e os atos administrativos.

Diante dos aspectos legais e técnicos para seleção e descarte bem como desenvolvimento de padrões e modelos para metadados, considera-se os suportes (tipos e tempo de vida), questões envolvendo autenticidade, integridade e proveniência. A infraestrutura tecnológica e a construção de repositório digital tornam-se essenciais para o desenvolvimento e manutenção de estratégias de preservação digital dos dados de pesquisa (GRÁCIO; MÁRDERO ARELLANO, 2020). Ressalta-se a pertinência de difundir a importância, o uso e a gestão de preservação dos dados científicos por meio de data papers e data journals.

Ramos (2018) analisa a partilha de dados por meio dos data papers e data journals no campo da saúde. As reflexões da autora podem ser ampliadas com base na defesa do potencial de acesso e compartilhamento de data papers e data journals enquanto geradores de benefícios para o pesquisador dentro dos sistemas de avaliação científica. Tanto os autores de dados científicos quanto pesquisadores que buscam e acessam os mesmos ampliam as possibilidades de geração de conhecimento por conta da reprodutibilidade e reuso dos dados. As duas subseções a seguir ampliam as reflexões sobre os data journals e data papers em convergência às discussões sobre preservação digital.

2. 1 Data Journals, Data Papers

Data Journals são revistas especializadas em publicações de *Data Papers*. Todos os aspectos importantes para a futura reutilização dos dados são feitos por meio desse modelo de revista especializada. Estes garantem que os dados científicos publicados sejam documentados e arquivados a longo prazo (em repositórios e com identificadores persistentes), que mantenham-se interoperáveis e acessíveis (de acordo com os Princípios FAIR²). Com disseminação de metadados associados, documentação de acompanhamento e código de software os dados brutos são processados ou manipulados para reutilização e análise, para que possam estar acessíveis na web e referidos em forma única. As novas configurações da comunicação científica têm representado os conjuntos de dados levantados durante o processo de pesquisa por meio de *Data Papers*.

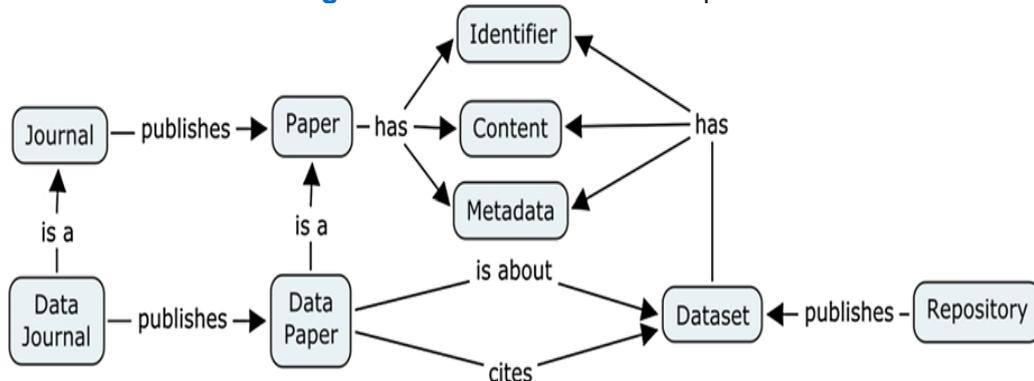
Os *Data Papers* são artigos de autoria, revisados por pares e citados em revistas acadêmicas, cujo conteúdo principal é uma descrição dos conjuntos de dados de pesquisa bem como informações básicas sobre a produção e aquisição de dados, com o objetivo de facilitar sua acessibilidade, disponibilidade e reutilização. Estão integrados na gestão de dados de pesquisa e relacionados à Repositórios de Dados (SCHÖPFEL; *et al*, 2019).

Geralmente são publicações que contém apresentação da pesquisa, matriz de dados e texto de metadados. Descrevem os conjuntos de dados em detalhe e não incluem interpretação ou discussão. Chavan e Penev (2011) definem o artigo de dados como “uma publicação em um periódico cujo objetivo é descrever dados em vez de relatar pesquisas ou suas conclusões”. Assim, *Data Papers* são formados por conjuntos de dados sem hipóteses e/ou os argumentos, resultados e discussões, elementos estes, pertencentes aos modelos tradicionais de publicação científica.

Os *Data Papers* incluem informações sobre autor(es), natureza e formato do conjunto de dados. Descrevem os objetivos, contexto, metodologia, processos, data de produção e condições de uso da pesquisa. São anexados ao artigo ou estão acessíveis no repositório onde foram depositados, conforme pode ser observado na figura 1:

² Os Princípios FAIR, um acrônimo para Findable, Accessible, Interoperable e Reusable, descrevem os recursos necessários para conseguir o arquivamento, preservação, manutenção, organização, recuperação e reutilização de dados científicos

Figura 1. Fluxo de um Data Paper



Fonte: Candela; et. al, 2015.

Os *Data Papers* têm sido incorporados gradativamente em diversas áreas do conhecimento como forma de divulgação científica. De acordo com levantamento realizado na plataforma Web of Science (2022), em 2017 haviam 1.083 *Data Papers* registrados. Já em 2022, a WOS apresenta 10.266 documentos, ou seja, nove vezes mais do que a quantidade produzida há cinco anos.

Ainda que os conjuntos de dados divulgados nos *Data Papers* sejam compostos pela junção de dados brutos, o seu conceito é heterogêneo considerando a variedade de diretórios, bases de dados e revistas que definem este substrato da comunicação científica. Alguns exemplos:

Quadro 1. Termos utilizados por revistas internacionais em referência aos documentos de dados.

Denominação	Revista
<i>Data Paper</i>	Annals of Forest Science, Zookeys
<i>Data notes</i>	F1000 Research, Gigascience
<i>Data article</i>	Data in Brief
<i>Data descriptor</i>	Scientific Data
<i>Dataset brief</i>	Proteomics
<i>Data in brief</i>	Genomics Data
<i>Database article – Software article</i>	BMC, Chemistry Central Journal
<i>Data reports</i>	Frontiers in Plant Science
<i>Resource article</i>	Plant Journal
<i>Database paper</i>	Plant & Cell Physiology
<i>Methods, software, databases, and tools</i>	Plos One

Fonte: Termos coletados nos sites das revistas que constam no quadro, 2022.

Os *Data Papers* reúnem todos os tipos de dados, brutos ou derivados (observações, medições, levantamentos, sequências, mapas, imagens, simulações computacionais, etc.), desde que o conjunto de dados constituído tenha valor científico, ou impacto ambiental, social, econômico, cultural. Este tipo de artigo informa a comunidade científica sobre a existência, originalidade, qualidade e disponibilidade de conteúdos que podem contribuir com o desenvolvimento de estudos das mais diversas áreas do conhecimento. Promove o trabalho de autores, explicita a importância dos dados produzidos e seu potencial de reutilização para pesquisas futuras. Os *Data Papers* geralmente são estruturados em 3 pontos:

- 1 - descrição do conjunto de dados e seu contexto,
- 2 - descrição dos métodos de obtenção,
- 3 - explicação do potencial de reutilização de dados.

As instruções para os autores geralmente propõem *Data Papers* que constam nas Políticas Editoriais, ou nas seções de instruções para os autores. O *Data Paper* deve conter todos os elementos que descrevem o conjunto de dados para que possam ser compreendidos e reutilizados por pesquisadores fora do estudo: tipo de dados, local e contexto de coleta, protocolos experimentais, equipamentos, variáveis estudadas, formato de arquivos, metadados, normas, bem como todas as informações explicativas das tabelas de dados (conteúdo das linhas e colunas, significado das abreviaturas, unidades de medida, *dentre outras*).

Se alguma destas informações já estiver publicada, por exemplo, os protocolos, basta mencionar o nome do protocolo com a referência que lhe dá acesso. O *Data Paper* também pode referir-se a um protocolo acessível no repositório em que o conjunto de dados está depositado. No repositório é possível associar um (ou mais) arquivo(s) de dados a arquivo(s) de texto, tais como, um arquivo “*Leia-me*” contendo várias informações, o arquivo de questionário de pesquisa, o arquivo de dicionário de variáveis, o protocolo de obtenção, etc. Assim, o *Data Paper* deve explicar por que o conjunto de dados é importante, como pode ser útil para outros pesquisadores ou outros atores (sociedade civil, tomadores de decisão, organizações não governamentais, *dentre outros*) e como pode ser reutilizado. O Quadro 2 apresenta as vantagens e desvantagens dos *Data Papers* em relação a outros tipos de documentos:

Quadro 2. Veículos de divulgação de dados, pontos fortes e fracos

Canais de divulgação de dados	Vantagens	Desvantagens
<i>Repositórios</i>	<ul style="list-style-type: none"> - São desenhados e reconhecidos pela comunidade científica; - Cotação direta de dados possível; - Sem (pouco) limite de tamanho (à critério da instituição mantenedora) 	<ul style="list-style-type: none"> Metadados mais ou menos ricos (reutilização mais ou menos fácil; conforme a equipe gestora opta por configurar os metadados) -Divulgação orientada para comunidades específicas de interesse.
<i>Artigo tradicional</i>	Máxima integração de dados e artigo: citável, pesquisável.	Dados difíceis de encontrar e de forma pouco reutilizável.

<i>Material complementar de artigo tradicional</i>	Formato dos dados livre de restrições de redação do artigo (formato, volume, natureza dos dados, etc.).	-Tamanho (peso do arquivo) geralmente limitado; -Apresentação heterogênea.
<i>Dados associados a documento de dados</i>	-Maior visibilidade: indexação, citação dupla (do <i>Dataset</i> e do conjunto de dados); -Pesquisa e reutilização de dados facilitada pela riqueza de metadados; -Sem restrição de tamanho; - <i>Peer Review</i> , crédito aos autores.	-Escolha do artigo para o conjunto de dados: disciplinar (biologia, ciências da vida, ciências do solo, química) institucional, genérico (por exemplo, Zenodo e figshare); -Tempo de redação; -Custo de publicação.

Fonte: Elaborado pelos autores, 2022.

Com base no Quadro 2, percebe-se que os *Data Papers* ampliam as possibilidades de divulgação científica. Recomenda-se que os pesquisadores e seus respectivos grupos utilizem este recurso aliado aos demais meios de publicação para aumentar o alcance e profundidade dos resultados e contribuições de seus estudos. Os *Data Papers* servem de guarda para manutenção, acesso e preservação digital dos dados de pesquisa.

2.2 Preservação digital de *Data Papers*

A publicação de um documento de dados faz parte da dinâmica global de abertura da ciência (Ciência Aberta) e responde aos princípios *FAIR*, que recomendam que os dados sejam fáceis de encontrar, acessíveis, interoperáveis e reutilizáveis. O desafio do *Data Paper* é a reutilização dos dados tornados públicos. Trata-se de implementar boas práticas para que os dados sejam descritos e documentados corretamente, ou seja, de acordo com os padrões disciplinares reconhecidos e praticados pelas diversas comunidades científicas.

A preservação digital de *Data Papers* consiste em garantir tanto a integridade dos dados quanto a acessibilidade, compreensibilidade e usabilidade dos dados para seus usuários (preservação "semântica"). Os formatos que os conjuntos de dados são indexados é condição necessária, senão suficiente, para a transformação dos dados em informações acessíveis, compreensíveis e utilizáveis.

Além da diversidade de formatos mais comuns, a variedade de formatos usados pelos produtores (pesquisadores) de dados é considerável, principalmente em comunidades científicas especializadas como é o caso do campo da Biodiversidade. A variedade de formatos aumenta os riscos que pesam sobre a capacidade dos produtores de dados e instituições reproduzirem corretamente os *Data Papers*. Esses múltiplos riscos incluem:

- A incapacidade de identificar o formato e, portanto, a ferramenta de reprodutibilidade adequada;
- A indisponibilidade de software de leitura devido ao seu custo, à política de uso da empresa que os desenvolve ou ao seu método de compra;

- A utilização de ferramentas inadequadas ou mal utilizadas que dão origem a uma reprodutibilidade equivocada, ou à perda involuntária de dados na sequência de uma migração mal controlada.

Para reduzir estes riscos (compra e manutenção de *hardware*, *software*, padrões, desenvolvimento e manutenção de competências, etc.), é desejável limitar a uma lista de formatos padronizados, preferencialmente abertos e não proprietários. Nesse ponto, as políticas das revistas que recebem *Data Papers* devem ter o cuidado de escolher e recomendar claramente as regras para submissão de conjuntos de dados. Por outro lado, quando revistas recebem dados em formatos não contemplados em sua política editorial, devem convertê-los em um formato adequado.

Para desenvolver sua própria política de formatos para preservação digital, cada revista deve ponderar a avaliação dos critérios objetivos em relação às suas prioridades. Desenvolver uma política de formato consiste em contemplar os riscos acima mencionados. Deste modo, as revistas que já possuem formatos de arquivos em sua política evidenciam a importância dos autores adequarem suas publicações, ao passo que reduzem as possibilidades de complexidade sobre os formatos aceitos. A política de preservação digital torna-se, portanto, um documento que não é exaustivo ou universal, e sim um regulamento que atende tanto às demandas da revista quanto dos autores.

4 Materiais e métodos

Diante da importância dos dados científicos para o avanço dos estudos no campo da Biodiversidade, analisa-se como a configuração dos *Data Papers* contribuem para a comunidade de pesquisadores, cujo principal objetivo é descrever um ou mais conjuntos de dados, de forma a permitir a sua reutilização. A categorização do estudo compreende em abordagem qualitativa e pesquisa documental através da aplicação de técnicas de análise de conteúdo para interpretação e descrição dos resultados. A categorização do estudo compreende em abordagem qualitativa, quantitativa e pesquisa documental através da aplicação de técnicas de análise de conteúdo para interpretação e descrição dos resultados.

A abordagem qualitativa descreve, compreende e interpreta questões levantadas pelos pesquisadores. A validação de informações perpassa pela consistência adquirida pela avaliação dos elementos teóricos e os achados da pesquisa (MARTINS ;THEÓPHILO, 2007). Enquanto que a abordagem quantitativa é aplicada para a estruturação dos instrumentos formais da coleta através de instrumentos de controle e ênfase nos dados numéricos e sua respectiva objetividade (GERHARDT; SILVEIRA, 2009).

A pesquisa documental, segundo Gil (2008; 2002), caracteriza-se por meio da natureza das fontes. Dentre as vantagens deste tipo de estudo está a variabilidade, estabilidade e temporalidade dos documentos.

A análise de conteúdo (AC) é, indiscutivelmente, um dos instrumentos metodológicos mais eficientes para desenvolver o que Bardin (2016) denomina de *análise de comunicações* devido sua multiplicidade de aplicações. É adaptável a todas as formas de comunicação, independente da natureza ou do suporte. Segundo Bardin (2016), a análise de conteúdo é formada por um conjunto de técnicas de análise, procedimentos objetivos e sistemáticos de descrição de informações para desvelar sentidos no plano de análise. É utilizada para a identificação e sistematização na coleta e análise dos dados organizada em três etapas principais.

a) *A pré-análise*: para seleção, delimitação do estudo; aplicação de regras de recorte e delineamento do objeto. O corpus de análise está formado por 33 revistas científicas sobre Biodiversidade em âmbito internacional editadas por instituições públicas e privadas. A seleção dos revistas parte de títulos indicados pela *Global Biodiversity Information Facility* (GBIF, 2022) sobre publicações em formato de *Data Papers* para a área da Biodiversidade.

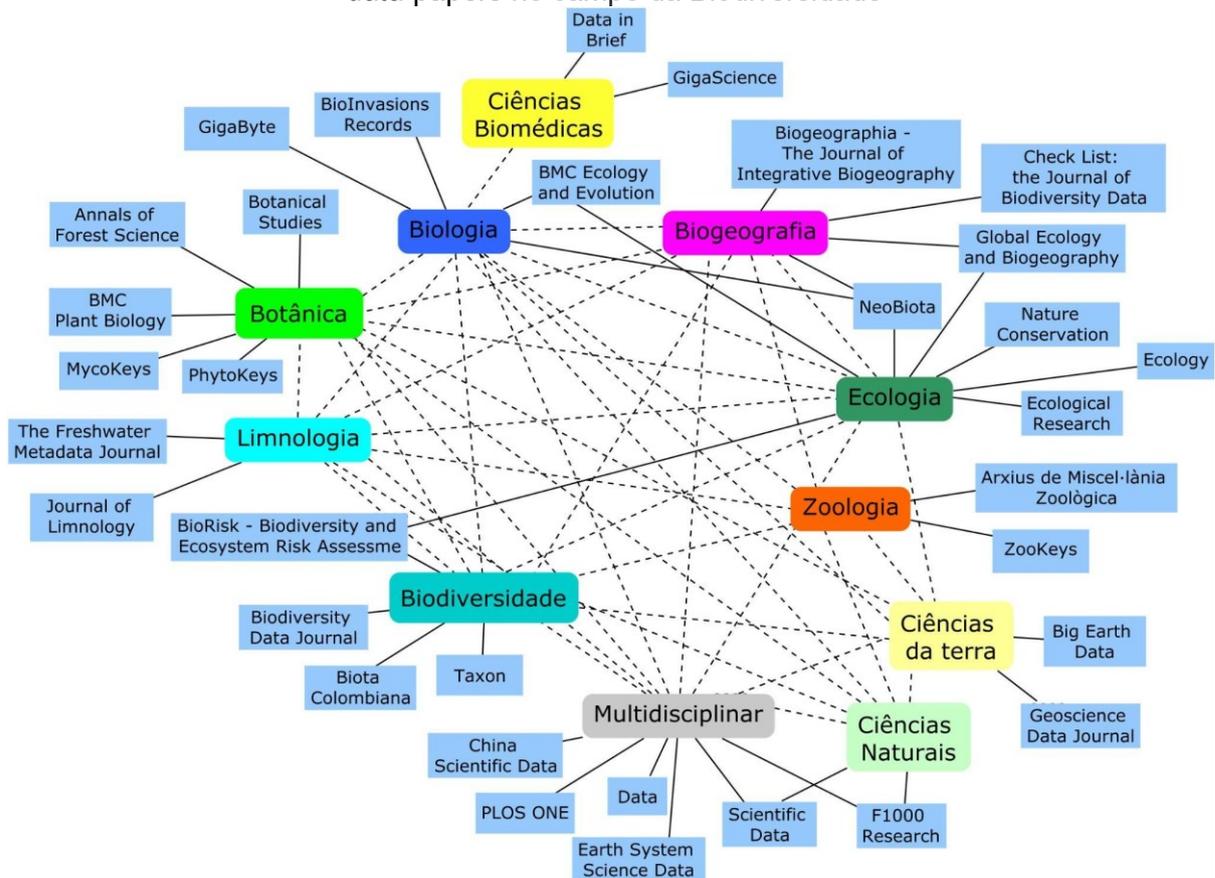
b) *A exploração do material*: que consiste na administração das técnicas sobre o *corpus* e; nos estudos desenvolvidos para esta pesquisa, a análise de conteúdo contribui para organização dos dados coletados e para a visualização dos resultados. Neste sentido, prioriza-se que os dados e documentos “revelem os achados” sem a aplicação de categorias pré-estabelecidas nos *corpus* pesquisados.

c) *O tratamento dos resultados e interpretações*: que podem ser desenvolvidos através de operações estatísticas (frequência do uso dos termos), sínteses e seleção dos resultados (inferências; interpretações, orientações para uma nova análise) para fins teóricos ou pragmáticos. Identifica-se em cada periódico: os temas correlatos à biodiversidade; os tipos de licenças e recursos de preservação digital, a quantidade de *Data Papers* publicados. A apresentação e análise dos resultados encontra-se na seção a seguir.

5 Discussão e Resultados

Dentre as 33 revistas indicadas pelo *Global Biodiversity Information Facility* (GBIF) que aceitam e publicam estudos da área da Biodiversidade, 23 são oriundos das Ciências Biológicas (estudo da vida). As dez revistas restantes pertencem aos campos de Ciências Naturais (estudo da natureza que envolvem Biologia, Física, Química e Matemática), Ciências da Terra (Geociências que abrangem conhecimentos sobre meio ambiente, cartografia, climatologia, entre outros), Ciências Biomédicas (Biomedicina que envolve pesquisas relacionadas às doenças humanas) e área multidisciplinar. A figura 2 apresenta nas linhas pontilhadas as relações de interação entre os temas gerais das revistas analisados, enquanto as representações em “linha-cheia” conectam as revistas com as respectivas áreas de foco e escopo segundo suas políticas editoriais.

Figura 2. Revistas e suas respectivas áreas principais que publicam *data papers* no campo da Biodiversidade



Fonte: Elaborado pelos autores, 2022.

Com base no corpus de análise, a área de Ciências Biomédicas possui dois revistas que publicam *Data Papers* sobre Biodiversidade, as Ciências da Terra e as Ciências Naturais também publicam duas respectivamente, no entanto, as duas revistas das Ciências Naturais também produzem conteúdos diversificados. Logo, se interligam com o campo Multidisciplinar, formado por seis revistas. Os *Data Papers* sobre Biodiversidade dentro da área de Ciências Biológicas distribuem-se em:

- a) Quatro revistas da Biologia, sendo que dois deles convergem com as áreas de Ecologia e Biogeografia;
- b) Cinco revistas de Botânica;
- c) Dois sobre Limnologia (biologia das águas);
- d) Quatro sobre Biodiversidade (em que um está interligado com a Ecologia);
- e) Na área da Zoologia encontram-se dois revistas;
- f) A Ecologia possui sete revistas, sendo que quatro deles possuem conexões com a Biogeografia e Biodiversidade (um de cada, respectivamente) enquanto que duas revistas estão interligadas com a área da Biologia e;
- g) A Biogeografia conta com três revistas em que uma também é pertencente ao campo da Ecologia.

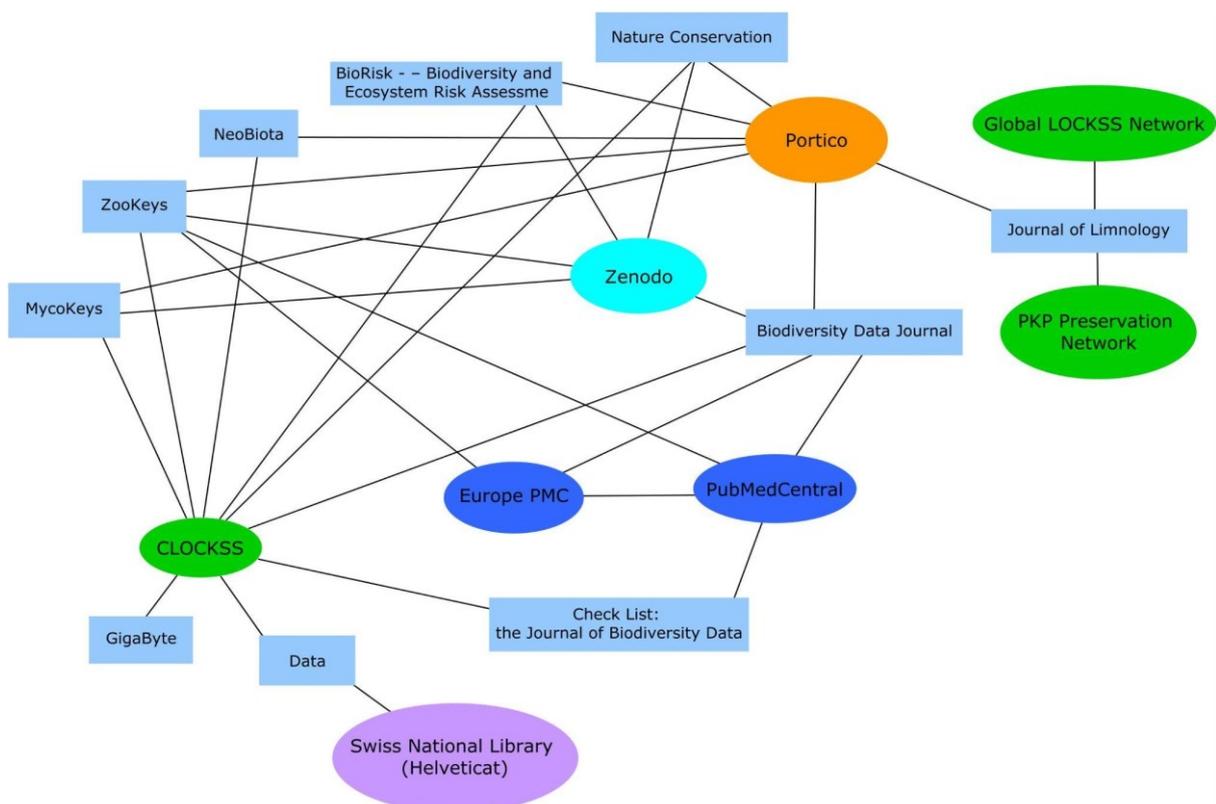
Todas as revistas são de acesso aberto, com exceção de duas revistas que oferecem o acesso aberto de modo opcional. Quatro revistas não definem o tipo de licença utilizado, mas disponibilizam suas publicações em acesso aberto. O periódico

Biota Colombiana utiliza a licença CC BY-NC-ND e o BMC Plant Biology possui a licença CC BY + CC0. A maioria das revistas, 25 delas, possuem licença CC BY.

Numa escala decrescente, a revista ZooKeys apresentou o maior número de indexadores com o total de 68. Seguido da Biodiversity Data Journal com 66 indexadores. A Ecological Research possui 46 e a Global Ecology and Biogeography apresentou 42 indexadores. A BMC Ecology and Evolution conta com 36 indexadores e a Annals of Forest Science com 34. A Biota Colombiana conta com 19 indexadores. As revistas Data; BMC Plant Biology e Earth System Science Data possuem 13 indexadores, o periódico F1000 Research, por sua vez, é contemplado por 11 plataformas de indexação.

As revistas Botanical Studies e GigaScience possuem oito indexadores; Big Earth Data e Geoscience Data Journal possuem seis; Data in Brief e Check List: the Journal of Biodiversity Data têm cinco; BioInvasions Records e Scientific Data possuem quatro, Arxius de Miscel·lània Zoològica têm três e o periódico China Scientific Data possui vínculo com apenas uma plataforma indexação. As revistas Biogeographia - The Journal of Integrative Biogeography; BioRisk - Biodiversity and Ecosystem Risk Assessme; The Freshwater Metadata Journal; GigaByte, Journal of Limnology; MycoKeys; Nature Conservation; NeoBiota; PhytoKeys; PLOS ONE e Taxon não especificaram a existência e/ou quantidade de indexadores.

Figura 3. Revistas internacionais que arquivam *data papers* no campo da Biodiversidade



Fonte: Elaborado pelos autores, 2022.

Tendo em vista os riscos envolvidos na perda de dados, algumas revistas estão tomando medidas adicionais de segurança na manutenção e gerenciamento dos seus arquivos digitais, aplicando um rigoroso controle de preservação digital distribuída. Dez revistas arquivam seus conjuntos de dados em repositórios gerais e de dados, via serviços de preservação digital e formato eletrônico com formatos e esquemas de

metadados e sistemas de backup individual (por meio do software livre LOCKSS e a rede CLOCKS que se estabelece através de redes e parcerias entre bibliotecas e editoras). Nove revistas utilizam o CLOCKS. O Global LOCKSS Network e o PKP Preservation Network pertencem ao mesmo periódico. Seis arquivam seus Data Papers e demais publicações no Pórtico. Cinco fazem uso do Zenodo. Três revistas operam com o PubMedCentral e duas com o Europe PMC. A Swiss National Library (Helvética) foi citada apenas uma vez.

O Periódico Zookeys utiliza seis sistemas de preservação digital, o Biodiversity Data Journal de cinco. BioRisk - Biodiversity and Ecosystem Risk e Journal of Limnology apropriam-se de três sistemas de preservação cada um. As revistas MycoKeys, Nature Conservation, NeoBiota, especificam o uso de três sistemas mas afirmam fazer uso de outros arquivos internacionais, sem denominar quais são para manter a preservação de seus dados. O periódico Data possui dois sistemas e apenas a revista Gigabyte adota um único sistema de arquivamento e preservação digital.

Acerca dos Princípios FAIR, verifica-se no Quadro 4 as 18 revistas que citam e aplicam os Princípios FAIR. A coluna ao lado dos respectivos títulos detalha os códigos de dados encontrados em cada uma.

Quadro 4. Princípios FAIR e códigos de dados nas revistas de Biodiversidade

Revistas	Códigos de dados
Annals of Forest Science	XML; CSV; XLS/XLSX; RDF
Big Earth Data	Jpeg, TIFF, PNG, script; html5; custom dcm
Biodiversity Data Journal	Textos: RTF ; PDF; HTML; XML_ Animações: SWF; DHTML (HTML/HTML5)_ Imagens: SVG; GIF; JPEG/JFIF; PNG; TIFF_ Filmes: MOV; MPG (MPEG); OGG; WebM _ Data sets: CSV; TSV Padrão para dados: DwC - Darwin Core
BioRisk	Textos: RTF ; PDF; HTML; XML_ Animações: SWF; DHTML (HTML/HTML5)_ Imagens: SVG; GIF; JPEG/JFIF; PNG; TIFF_ Filmes: MOV; MPG (MPEG); OGG; WebM _ Data sets: CSV; TSV Padrão para dados: DwC - Darwin Core
Data	TIFF, JPEG, EPS; PDF; PNG; JPEG; SVG
Ecological Research	CSV; Não detalha os demais códigos
Ecology	EPS; PDF; TIFF; PNG; EPS; GIF; JPG; TIF/TIFF; PNG; WMF; DOC; PPT; PSD; AI; PS
F1000Research	TIFFs; EPS; DOC
GigaByte	TIF; archive; TEXT; RAW; IBD; zip; DOC; GZIP; CSV
GigaScience	TIF; archive; TEXT; RAW; IBD; zip; DOC; GZIP; CSV
MycoKeys	xls; ods; csv
Nature Conservation	DOC; DOCX; RTF, ou ODF; CSV; TSV; EPS; TIFF ; PNG; JPEG; GIF; BMP; SVG
NeoBiota	CSV; TSV; PDF ; SWF; MOV; MPG; XLS; ODS
PhytoKeys	XLS; ODS; CSV; TSV; PDF; SWF; MOV; MPG; XLS
PLOS ONE	TIFF; EPS; SVG
Scientific Data	DOC; DOCX; TEX; PDF; EPS; TIFF; JPG; XLS; XLS
Taxon	EPS; PDF; TIFF; PNG; EPS; GIF; JPG; TIF/TIFF; PNG; WMF; DOC; PPT; PSD; AI; PS
ZooKeys	CSV; EPS; TIFF; PNG; JPEG; GIF; BMP; SVG; PDF

Fonte: Elaborado pelos autores (2022)

Dentre os títulos elencados no quadro acima, verifica-se que seis delas combinam mais de uma política de dados. A Big Earth Data, aplica a política de dados da Taylor & Francis Open. BioRisk; Ecology; GigaScience e Taxon seguem as recomendações do FORCE11 Software Citation. Cinco delas adotam os Princípios de Panton: MycoKeys; Nature Conservation; NeoBiota; PhytoKeys e ZooKeys.

Três revistas que atuam e desenvolvem seus planos de gestão de dados em acesso aberto não especificaram quais princípios seguem: Biota Colombiana, BMC Plant Biology e China Scientific Data. Enquanto que a revista Biogeographia – The Journal of Integrative Biogeography incentiva mas não exige que os dados sejam disponibilizados em livre acesso pelos autores. Afirma também estar desenvolvendo um plano de gestão de dados em parceria com a LifeWatch Itália.

Seguir as políticas da editora conveniada são as opções escolhidas por três revistas: Botanical Studies (SpringerOpen); Check List (Princípios de Panton e Pensoft); Data in Brief - (Princípios Elsevier). Cinco revistas não especificam quais princípios de preservação utilizam: Arxius de Miscel·lània Zoològica; The Freshwater Metadata Journal; Geoscience Data Journal; Journal of Limnology e Global Ecology and Biogeography.

Para chegar a quantidade de *Data Papers* publicados, diversos fatores influíram durante os processos de busca e coleta dos dados. Registra-se a redução da quantidade de materiais encontrados com base nas seguintes questões:

- A não especificação do tipo de documento nos sistemas avançados de busca, tampouco a identificação dos mesmos nas listagens de materiais disponíveis em alguns catálogos e;
- A impossibilidade de distinção do tema sobre biodiversidade nos resultados disponíveis nas revistas multidisciplinares.

No universo das 33 revistas analisados, encontrou-se 17 revistas com publicações em formato de *Data Papers*, gerando um total de 882 documentos, publicados entre 2003 até maio de 2022.

Quadro 3. Ranking entre os 17 revistas que publicaram *Data Papers*

Total de <i>Data Papers</i>	Periódico(s)
315	Biodiversity Data Journal
131	Ecology
93	ZooKeys
76	Geoscience Data Journal
68	Ecological Research
59	BMC Plant Biology
36	Arxius de Miscel·lània Zoològica

24	Annals of Forest Science. PhytoKeys
15	Biota Colombiana
13	Big Earth Data
7	BMC Ecology and Evolution. Nature Conservation
5	BioInvasions Records. NeoBiota
3	MycoKeys
1	BioRisk - Biodiversity and Ecosystem Risk Assessment

Fonte: Elaborado pelos autores, 2022.

Das seis revistas com maior número de publicações em formato de artigos de dados, dois são editados no Reino Unido, dois na Bulgária, um nos Estados Unidos e um no Japão. O único periódico da América Latina entre as revistas analisados é o Biota Colombiana que possui 15 *Data Papers*.

Todas as revistas publicam em Inglês. Arxius de Miscel·lània Zoològica, aceita também submissões em Catalão e Espanhol, mas opta por publicar os estudos somente na língua inglesa. O Biota Colombiana aceita o envio de materiais em Espanhol, inglês ou português, mas tem preferência pelo idioma inglês. O China Scientific Data, por sua vez, é um periódico bilíngue da Chinese Academy of Sciences. Cintra *et al.* (2020) assinalam que o inglês pode ser considerado como uma espécie de “língua franca” da ciência, pois possibilita a comunicação e a troca de informações entre pesquisadores de distintos países. Meneghini e Parker em seu estudo de 2007 já apontavam para o crescimento da tendência de se publicar artigos em inglês por países não anglófonos, o que reflete os esforços das instituições de ensino e pesquisa, dos pesquisadores e das próprias revistas em se adaptarem nesse sentido, visando à internacionalização da produção científica (MENEGHINI; PARKER, 2007; CINTRA *et al.*, 2020).

6 Conclusão

O número em *Data Papers* teve crescimento exponencial entre 2017 até maio de 2022. Logo, os artigos sobre o campo da Biodiversidade também têm aumentado em diversos temas que envolvem todo o seu ecossistema. A partir dos pontos elencados para análise especificados na seção metodológica do estudo, conclui-se que cinco grandes áreas contemplam temas correlatos à Biodiversidade: Ciências Biológicas; Ciências Naturais; Ciências da Terra; Ciências Biomédicas e o campo Multidisciplinar. Apenas dois títulos oferecem o acesso aberto como opção, os demais 31 revistas são de acesso aberto, sendo que 27 especificam o tipo de licença utilizada.

Cabe observar que apenas dez revistas arquivam seus *Data Papers* (e demais publicações). Nove deles se apropriam de dois ou mais sistemas. O uso de indexadores é mais frequente e é utilizado em 22 revistas. Apenas 11 não especificaram o uso de serviços de indexação. Encontrou-se um total de 882 *Data Papers* em 17 revistas analisados.

As seis revistas que mais publicam *Data Papers* sobre Biodiversidade são: Biodiversity Data Journal; Ecology; ZooKeys; Geoscience Data Journal; Ecological Research; BMC Plant Biology. Inglês é o idioma adotado em todas as revistas analisadas. Os *Data Papers* analisados se caracterizam como documentos revisados por pares e representam conjuntos de dados indexados com padrões de metadados adequados para preservar digitalmente os dados registrados nas revistas que foram contempladas na presente análise.

Conforme o levantamento realizado, fica evidente o crescimento de revistas que recebem conjuntos de dados como insumo da comunicação científica. Ao contrário de um repositório, os *Data papers* garantem um processo de revisão explícito e descritivo, focado na consistência dos dados e no método de coleta.

Referências

BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2016.

CANDELA, L. *et al.* Data journals: a survey. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1747–1762, 2015. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23358>. Acesso em: 21 maio 2022.

CHAVAN, V.; PENEV, L. The Data Paper: a mechanism to incentivize data publishing in biodiversity science. **BMC Bioinformatics**, n. 12, 2011. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-S15-S2>. Acesso em: 21 maio 2022.

CINTRA, P. R.; SILVA, M. D. P. da; FURNIVAL, A. C. Uso do inglês como estratégia de internacionalização da produção científica em Ciências Sociais Aplicadas: estudo de caso na SciELO Brasil. **Em Questão**, Porto Alegre, v. 26, n. 1, p. 17–41, 2020. DOI: 10.19132/1808-5245261.17-41. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/88528>. Acesso em: 26 maio 2022.

GERHARDT, T. E. SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GRÁCIO, J. C. A.; MÁRDERO ARELLANO, M. Á. A gestão da preservação digital de dados de pesquisa: proposta de um modelo processual. **Revista Brasileira de Preservação Digital**, Campinas, SP, v. 1, 2020. DOI: 10.20396/rebpred.v1i00.13223. Disponível em: <https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/13223>. Acesso em: 19 jun. 2022.

GLOBAL BIODIVERSITY INFORMATION FACILITY. **Artigos de dados**: receber reconhecimento acadêmico para seus conjuntos de dados. Copenhagen, 2022. Disponível em: <https://www.gbif.org/pt/data-papers>. Acesso em: 20 maio 2022.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

MARTINS, G. de A.; THEÓPHILO, C. R. **Metodologia da investigação científica para ciências sociais aplicadas**. São Paulo: Atlas, 2007.

MENEGHINI, R.; PACKER, A. Is there science beyond English? **EMBO Reports**, Heidelberg, v. 8, n. 2, p. 112-116, 2007.

RAMOS, J. F. dos S. L. **Data papers e o sistema de recompensas na avaliação da ciência**. 2018. 24 f. Projeto de Pesquisa (Especialização) – Curso de Especialização em Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, 2018.

SCHÖPFEL, J. *et al.* Data papers as a new form of knowledge organization in the field of research data. *In*: COLLOQUE INTERNACIONAL D'ISKO-FRANCE, 12., 2019, Montpellier. **Anais [...]**.

Disponível em: <https://halshs.archives-ouvertes.fr/ISKOFRANCE2019/halshs-02284548>. Acesso em: 21 maio 2022.