

QUALIDADE DA INFORMAÇÃO EM HERBÁRIOS VIRTUAIS NO BRASIL: COLEÇÕES DE BRIÓFITAS COMO ESTUDO DE CASO²

Luciana Carvalho dos Reis^{3, 4},
Emília de Brito Valente²

Recebido 23.08.2021; Aceito 19.10.2021

ABSTRACT

Virtual collections are fundamental for scientific dissemination and democratizing access to knowledge produced in herbaria. However, the occurrence of errors and inconsistencies in virtual records can affect the quality of this data. In this study, we analyzed the bryophytes virtual collection from seven Brazilian herbaria in order to verify the quality of the available data. For each collection errors were identified, classified into four categories ('Taxonomic', 'Geographic', 'Nomenclatural' and 'Incomplete or absent') and quantified. We obtained the dataset from 'Herbário Virtual da Flora e dos Fungos' platform on August 26, 2020, using the searchform of the SpeciesLink tool. The results showed that the quantity of specimens in a collection is not related to higher or lower quality of data. The higher the number of specimens records containing wrong information in 'Taxonomic' and 'Geographic' categories, the lower is the Data Quality Index value. Considering that virtual herbaria are extremely important for scientific dissemination, which is guaranteed by the free and open access of their information. It is essential that the information contains as few errors as possible. Thus, it is necessary not only to continuous monitoring of data quality available but also training on the biological group and the importance of providing quality data in virtual databases for those responsible for data digitalization.

Keywords: Data quality, Botanical collections, Virtual database.

RESUMO

Coleções virtuais são fundamentais para a divulgação científica e para a democratização do acesso ao conhecimento produzido nos herbários. No entanto, a ocorrência de erros e inconsistências em registros virtuais podem afetar a qualidade destes dados. Neste estudo, analisamos as coleções virtuais de briófitas de sete herbários brasileiros, com o objetivo de verificar a qualidade dos dados disponibilizados. Para cada coleção os erros foram identificados, categorizados ('Taxonômico', 'Geográfico', 'Nomenclatural' e 'Incompleto ou ausente') e quantificados. Obtivemos o conjunto de dados a partir da plataforma Herbário Virtual da Flora e dos Fungos em 26 de agosto de 2020, utilizando o formulário de busca da ferramenta *SpeciesLink*. Os resultados mostraram que o tamanho da coleção não está relacionado a uma maior ou menor qualidade dos dados. Quanto maior foi a proporção de registros contendo erros 'Taxonômico' e 'Geográfico', menor foi o valor do Índice de Qualidade dos Dados. Considerando que os herbários virtuais são de grande importância para a divulgação científica, a qual é garantida pelo acesso aberto e gratuito das suas informações, é fundamental que as informações disponibilizadas contenham a menor quantidade de erros possível. Assim, se faz necessário não apenas o monitoramento contínuo da qualidade dos dados mas, também, a realização de treinamento sobre o grupo biológico e a importância da

² Baseado no Estágio Pós-Doutoral da primeira autora. BOLSISTA CAPES/BRASIL.

³ Doutora em Biologia Vegetal. Universidade Estadual de Feira de Santana, Departamento de Ciências Biológicas, Programa de Pós-Graduação em Botânica, Feira de Santana, 44036-900 Bahia, Brasil.

⁴ Autor para correspondência: luciana_cr@hotmail.com

disponibilização de dados de qualidade em bancos virtuais para aqueles responsáveis pela digitalização dos dados.

Palavras-chave: Qualidade de dados, Coleções botânicas, Banco de dados virtual.

INTRODUÇÃO

A criação de coleções botânicas virtuais tem sido fundamental para promover a democratização do acesso às informações produzidas e armazenadas em herbários ao redor do mundo (Soltis, 2017). Nas coleções virtuais são disponibilizadas informações de espécimes de plantas depositadas nos herbários, as quais podem ser acessadas de maneira *on-line*, remota e gratuita por todos os interessados (Page et al., 2015). Além de otimizar o tempo necessário para o acesso e estudo das coleções botânicas, a criação das coleções virtuais permite que dados e informações gerados pelos herbários permaneçam disponíveis aos usuários, mesmo em casos de perda ou destruição da coleção física. Isso é extremamente importante, especialmente para o caso dos herbários brasileiros, onde a ocorrência de problemas crônicos de falta de infraestrutura básica, dificuldade de manutenção e falta de apoio institucional colocam em risco a integridade das suas coleções físicas (Gasper et al., 2020).

A utilização de dados virtuais em pesquisa, divulgação científica e estudos integrados na área de biodiversidade cresceu significativamente nas últimas duas décadas (Nelson & Ellis, 2018; Soltis et al., 2018), devido ao elevado potencial que este tipo de dados tem para elucidação de uma gama de questões evolutivas e ecológicas (Pyke & Ehrlich, 2010). Nas coleções virtuais, além das informações taxonômicas e informações referentes ao local de coleta daquele espécime, os registros podem incluir imagens do espécime contido na exsicata ao qual se refere aquele registro virtual (Kovtonyuk et al., 2018; Kislov et al., 2017). Ou seja, coleções virtuais disponibilizam um conjunto de informações que podem ser utilizadas para a realização de diferentes estudos. As imagens digitalizadas dos espécimes nas coleções virtuais podem subsidiar estudos sobre a fenologia de plantas, bem como o desenvolvimento de ferramentas com esta perspectiva (Pearson et al., 2020; Weaver et al., 2020). Dados de ocorrência das espécies, disponíveis em coleções virtuais, subsidiam inventários florísticos (Simões et al., 2020a; Colli-Silva et al., 2016), construção de modelos de nicho espacial de espécies (Simões et al., 2020b) e a construção de modelos de distribuição potencial de espécies nativas (Sciamarelli & Torgeski, 2019; Gladson et al., 2018) e de espécies invasoras (Bradley, 2013). Dados geográficos (longitude e latitude) foram utilizados, por exemplo, para modelar a distribuição potencial de uma espécie de angiosperma e os efeitos das mudanças climáticas sobre ela (Zhang et al., 2020). As coleções botânicas virtuais também são bastante úteis em metodologias não-formais de ensino, com aplicabilidade nos diferentes níveis de ensino, para incentivar o aprendizado ativo sobre as plantas (Flannery, 2013).

É notório o papel dos herbários virtuais para a divulgação científica, cujo principal objetivo é permitir que pessoas leigas tomem conhecimento e assimilem o progresso científico, compreendendo o mundo em que vivem (Silva, 2006). Desse modo, é imprescindível que os dados disponibilizados pelas coleções virtuais apresentem boa qualidade, a qual está intrinsecamente relacionada à confiabilidade dos dados, ou seja, ao fornecimento da informação completa, correta e padronizada, de acordo com normas internacionalmente aceitas (Chapman, 2005). Em coleções de herbários virtuais, a maioria das inconsistências encontradas está relacionada a erros de digitação dos dados, principalmente devido ao fato desse processo ser realizado por pessoas não especializadas na área botânica (Dalcin et al., 2012).

No Brasil, uma das principais plataformas para disponibilização de coleções botânicas virtuais é o INCT – Herbário Virtual da Flora e dos Fungos (<http://inct.florabrasil.net/>).

Criada em 2011, esta plataforma reúne os acervos virtuais da maioria dos herbários brasileiros, disponibilizando mais 5 milhões de registros de espécimes de plantas e fungos, incluindo mais de 40 coleções virtuais de briófitas pertencentes a herbários localizados no Brasil e no exterior.

Briófitas constituem o segundo maior grupo de plantas terrestres, com aproximadamente 18.150 espécies (Goffinet & Shaw, 2009) em todo o mundo, das quais 1.524 ocorrem no Brasil (Costa & Peralta, 2015). As briófitas são plantas avasculares de tamanhos diminutos e que não apresentam flores ou frutos, cujas espécies estão distribuídas nos filos Bryophyta (musgos), Marchantiophyta (hepáticas) e Anthocerotophyta (antóceros) (Glime, 2017). Em relação às coleções virtuais, diferentemente do que ocorre para os demais grupos vegetais, a inclusão de imagens nos registros de briófitas ainda é um pouco menos usual e isso, provavelmente, decorre do fato de que as exsicatas de briófitas, via de regra, contém mais de uma espécie 'misturadas' na amostra (Gadelha-Neto et al., 2013). Contudo, isso não diminui o potencial elucidativo desses dados.

Informações obtidas em coleções virtuais sobre locais de ocorrência de espécimes de briófitas, combinados com dados de literatura e de amostragem, foram utilizadas para avaliar a eficácia e a representatividade das Unidades de Conservação na Floresta Atlântica brasileira, através da estimativa de distribuição potencial de dez espécies-chave pertencentes a este grupo vegetal (Silva et al., 2014). Um levantamento das espécies dos gêneros *Calymperes* e *Syrrhopodon* ocorrentes na porção nordeste da Floresta Atlântica brasileira, também utilizou uma combinação de dados virtuais, de literatura e de amostragem (Maciel-Silva et al., 2015). Dados virtuais georreferenciados foram utilizados para avaliar a distribuição da riqueza de espécies e o endemismo em briófitas, em uma escala continental (Stevenson et al., 2012). Informações sobre locais de ocorrência, obtidos em coleções virtuais, foram utilizados para verificar a correlação entre *hotspots* de samambaias e briófitas e, se a riqueza de samambaias pode ser usada como substituta para a riqueza de briófitas (Nagalingum et al., 2014).

Considerando a representatividade das briófitas entre as plantas terrestres e considerando a existência de um número bastante representativo de acervos virtuais para este grupo vegetal, as coleções virtuais de briófitas foram definidas como o foco deste trabalho. Nosso principal objetivo é verificar a qualidade dos dados nas coleções virtuais de briófitas dos seguintes herbários brasileiros: Herbário da Universidade Estadual de Feira de Santana, Herbário Alexandre Leal Costa, Herbário Geraldo Mariz, Herbário da Universidade Federal de Sergipe, Herbário Professor Aluizio Bittencourt, Herbário Rosa Mochel e Herbário Lauro Pires Xavier. Para isso, estas coleções foram analisadas para identificação e classificação dos principais tipos de erros.

MATERIAL E MÉTODOS

Conjunto de dados

O conjunto de dados corresponde aos registros de briófitas pertencentes às coleções virtuais de sete herbários, localizados em cinco dos sete estados da região nordeste do Brasil:

Bahia:

- HUEFS – Herbário da Universidade Estadual de Feira de Santana;
- ALCB – Herbário Alexandre Leal Costa, da Universidade Federal da Bahia;

Maranhão:

- HABIT – Professor Aluizio Bittencourt, da Universidade Estadual do Maranhão;
- SLUI – Herbário Rosa Mochel, da Universidade Estadual do Maranhão;

Paraíba:

- JPB – Herbário Lauro Pires Xavier, da Universidade Federal da Paraíba.

Pernambuco:

- UFP – Herbário Geraldo Mariz, da Universidade Federal de Pernambuco;

Sergipe:

- ASE – Herbário da Universidade Federal de Sergipe;

Os dados foram obtidos na plataforma Herbário Virtual de Flora e Fungos em 26 de agosto de 2020, através do formulário de busca da ferramenta *SpeciesLink*. Para cada coleção selecionada foi aplicado o filtro *Filo*, onde eram selecionadas apenas as espécies referentes ao grupo de interesse para download dos dados. Considerando a possibilidade de existirem registros de briófitas cujo este campo não estivesse preenchido, para todas as coleções foram obtidos aqueles registros com o campo *Filo* indicado como “Em branco”, para posterior seleção dos registros referentes ao grupo das briófitas. A Tabela 1 mostra o número de registros obtidos para cada banco de dados virtual de acordo com o filtro selecionado.

Apenas para a coleção do Herbário Geraldo Mariz, da Universidade Federal de Pernambuco, os dados utilizados foram obtidos diretamente através da Curadoria que disponibilizou uma planilha com banco de dados exportado. De posse dos dados, o primeiro passo consistiu na conferência manual de todos os registros que apresentavam o campo ‘Filo’ sem informação (ou seja, estava em branco). Esse procedimento foi necessário para verificar a ocorrência de registros de espécimes de briófitas nesse conjunto de dados. Em seguida, foi gerada uma planilha para cada coleção virtual contendo apenas os registros de briófitas.

Caracterização dos tipos de erros

Em virtude da impossibilidade de conferência das exsicatas de briófitas depositadas nas coleções físicas em cada herbário, todo o trabalho foi realizado de maneira remota. Desse modo, foram verificados apenas aspectos relacionados à grafia de informações, ausência ou erro em coordenadas geográficas e inconsistências na inclusão de espécies em *Filo*, *Ordem*, *Classe* e *Família*. Relacionamos quatro categorias principais de erros, adaptadas para coleções *on-line* a partir do proposto por Dalcin et al. (2012) e descritas abaixo:

Taxonômico: São os erros causados principalmente por digitação errada ou pela transcrição de um erro constante na própria etiqueta da exsicata, ou pela inclusão equivocada de uma espécie ou subespécie em um dado gênero, família, classe, ordem ou filo.

Nomenclatural: Ocorrem principalmente por falta de padronização nos nomes e das formas de escrita do mesmo valor. Um exemplo são os nomes de autores de espécies, ou nomenclatura utilizada para indicar o filo ao qual pertence a espécie (ex. Moss, Briófitas, Briophyta);

Geográfico: Consiste em coordenadas geográficas imprecisas, indicando local diferente daquele informado na exsicata. Nesta categoria também foram incluídos erros de grafia em nomes de Estados, Municípios e localidade da coleta do espécime;

Incompleto ou ausente: Quando algumas informações estão totalmente ausentes ou são apenas parcialmente preenchidas, como por exemplo as coordenadas geográficas. Aqui também estão incluídas ausência de informações taxonômicas como filo, ordem, classe e família, além de nome dos autores das espécies.

Identificação de erros

Considerando que a maioria dos erros encontrados em bancos de dados virtuais estão relacionados a problemas de digitação, falta de padronização de dados e coordenadas geográficas imprecisas (Chapman, 2005), cada registro dentro de cada coleção foi cuidadosamente verificado para a identificação de erros.

Para identificação dos registros contendo inconsistências em informações de georreferenciamento, utilizamos a ferramenta *DataCleaning* da plataforma *SpeciesLink* onde as seguintes classes de erros foram consultadas: Coordenada fora do limite mundial, Coordenada no mar, Lat/Long com zero e Análise da coordenada no Brasil.

Para cada coleção analisada foi gerada uma planilha de dados contendo informações para localização do registro (Código da Coleção, Código da exsicata, Coletor e Número de coletor), a identificação da informação errada e a informação corrigida. Posteriormente, as planilhas foram enviadas para a curadoria de cada uma das coleções.

Estimativa da qualidade dos dados

Para avaliar a qualidade dos dados, pesos foram atribuídos a cada tipo de erro identificado nas sete coleções analisadas de acordo com a ordem simples dos valores reais: quanto maiores são as porcentagens reais, menores são os pesos (devido à relação negativa entre a quantidade de erros e a qualidade dos dados). Tendo em vista que as coleções apresentaram quantidades díspares de registros, valores percentuais para cada tipo de erro foram utilizados para ordenar as coleções. Os pesos atribuídos variam de 1 a 7, onde 7 representa a maior qualidade e 1 a menor qualidade dos dados. A soma dos valores foi considerada como um índice de qualidade dos dados.

RESULTADOS

No total, considerando as sete coleções selecionadas, foram analisados 37.340 registros virtuais de briófitas, oriundos de coletas realizadas em diferentes países ao redor do mundo. A Tabela 2 mostra o número de registros obtidos para cada coleção virtual de briófitas, nos quais foram procedidas a verificação e correção de informações. Em todas as coleções analisadas verificou-se a ocorrência de registros com mais de um tipo e/ou categoria de erro. Em relação ao tamanho, as coleções virtuais de briófitas analisadas podem ser organizadas na seguinte ordem, de acordo com o número de registros disponibilizados: UFP (16.525), ALCB (13.024), HABIT (3.491), HUEFS (3.433), JPB (612), ASE (239) e SLUI (16).

A tabela 3 mostra as proporções de erros por categoria, considerando cada coleção analisada. A coleção ASE foi a única a não apresentar ocorrência de erros Taxonômicos ou Geográficos, além uma baixa proporção de registros contendo erros da categoria Nomenclatural. A coleção ALCB apresentou baixo número de erros tanto nas categorias 'Taxonômico' e 'Geográfico', mas houve elevada proporção de erros na categoria 'Nomenclatural'. A coleção UFP apresentou menos de 10% dos registros contendo erros Taxonômicos e baixa incidência de erros Geográficos, ocorrendo maior incidência de erros na categoria 'Nomenclatural'. Na coleção HABIT houve baixa incidência de erro nas categorias 'Taxonômico' e 'Nomenclatural', mas elevada proporção de registros apresentando algum erro na categoria 'Geográfico'. A coleção SLUI apresentou valores intermediários de erro nas categorias 'Taxonômico' e 'Geográfico'. Não houve registro com erro Nomenclatural na coleção SLUI. A coleção HUEFS apresentou elevada proporção de erro Taxonômico e menos de 2% dos registros apresentavam erros na categoria Geográfico. A coleção JPB apresentou elevada proporção de erros nas categorias Taxonômico e Nomenclatural. Para a categoria 'Incompleto ou Ausente', todas as coleções apresentaram proporção bastante elevada, com valores chegando a 100% na maioria das coleções analisadas (Tabela 3).

Além dos erros analisados nas quatro categorias definidas neste trabalho, em algumas coleções foi observada a ausência de informação acerca do número de tombamento da exsicata (ou código da exsicata dentro da coleção). Esta é uma informação extremamente importante e sua disponibilização é que vai permitir a localização daquele registro dentro da coleção física. Registros sem essa informação foram observados nas coleções ALCB (353 registros/2,71%) e UFP (190 registros/1,15%). Também foi observada a ocorrência de duplicidade de registros na coleção UFP.

Em relação à qualidade dos dados analisados, as coleções ASE, HABIT e SLUI foram as que apresentaram maiores valores do índice de qualidade, respectivamente. A coleção HUEFS apresentou a menor qualidade dos dados (Tabela 4).

DISCUSSÃO

Plataformas de dados digitais são fundamentais para o processo de globalização do conhecimento (Page et al., 2015). Na plataforma INCT - Herbário Virtual da Flora e dos Fungos, mais de 5 milhões de registros de espécimes de plantas e fungos depositados em herbários brasileiros e internacionais são disponibilizados de maneira *on-line* aberta e gratuita para toda a sociedade. Quando se trata da digitalização das exsicatas de briófitas para os bancos virtuais, alguns cuidados devem ser tomados para que as informações necessárias e inerentes ao grupo sejam incluídas. Desse modo, é extremamente necessário que a pessoa responsável pela realização desta digitalização tenha um mínimo conhecimento sobre o grupo, ou pelo menos, saiba onde buscar as informações de maneira correta e eficiente para a garantia da melhor qualidade possível desses dados. Contudo, uma vez que a digitalização das informações para as coleções virtuais é realizada pelo componente humano, a menos que esforços extraordinários tenham sido feitos, a ocorrência de erros em bancos de dados virtuais é comum e, até mesmo, esperada (Redman, 1996). Além disso, a própria natureza das coleções botânicas, como os nomes científicos e dados históricos, pode facilitar a ocorrência de erros, a qual pode ser ainda maior se o processo de adição dos dados for realizado por usuários sem conhecimento específico da área (Dalcin et al., 2012).

No geral, as coleções analisadas neste estudo apresentaram boa qualidade dos dados, mas é notória a ocorrência da diminuição da qualidade dos dados acompanhando um aumento do número de registros contendo erros das Categorias 'Taxonômico' e 'Geográfico' (Tabela 3). Esse achado está de acordo com o postulado por Dalcin et al. (2012), os quais destacam a maior importância dos erros Taxonômicos, seguido dos erros nas categorias 'Geográfico', 'Nomenclatural' e 'Incompleto ou Ausente'. Além disso, uma análise dos resultados deixa claro que não necessariamente as coleções maiores tendem a apresentar uma maior proporção de registros contendo erros.

Para os erros da categoria 'Nomenclatural', o principal problema encontrado nas coleções analisadas corresponde à falta de padronização da informação, especialmente para o dado referente ao nome do autor das espécies. Esta falta de padronização parece ser resultado da utilização de diferentes bases de dados para conferência de autores (ex.: Flora do Brasil 2020, *Tropicos* e *The Plant List*, muito embora esta última plataforma não esteja sendo atualizada). Embora as três plataformas sejam igualmente confiáveis, em uma quantidade considerável de nomes de espécies há diferenças entre os autores atribuídos a estas. Esta situação é exemplificada a seguir: Enquanto na plataforma Flora do Brasil 2020 encontramos a espécie *Odontolejeunea lunulata* (Weber) Schiffn., na plataforma The Plant List encontramos *Odontolejeunea lunulata* (F. Weber) Schiffner. Embora ambas as formas para o nome do autor estejam corretas, a plataforma INCT – Herbário Virtual da Flora e dos Fungos entende como dois autores diferentes e então, indica estes registros na ferramenta *DataCleaning* como "registros suspeitos para o nome do autor". Nestes

casos, onde há duplicidade em nome de autores recomenda-se fazer a escolha por um único formato de escrita (Brumitt & Powell, 1992). Também a ausência ou não do ponto no final do nome faz com que este seja entendido como um nome distinto. É possível que a falta de padronização dos nomes de autores seja potencializada pelo fato de diferentes pessoas estarem responsáveis pela inserção dos dados, ao longo do tempo. É comum que, ao ser iniciado o trabalho de um novo responsável pela digitalização, este não tenha se comunicado com o anterior para verificar qual a base de dados que estava sendo utilizada.

A ocorrência de erros na categoria 'Nomenclatural' referentes ao Filo das espécies, bem como de erros na categoria 'Taxonômico' (ex.: Filo, Ordem e Classe), pode ser evitada com a adoção dos sistemas de classificação propostos Renzaglia et al. (2009), Crandall-Stotler et al. (2009) e Goffinet et al. (2009) para os filios Anthocerotophyta, Marchantiophyta e Bryophyta, respectivamente. Além disso, recomendamos fortemente a seguinte padronização das informações no campo Filo através da utilização dos seguintes termos: 1) *Anthocerotophyta* para as espécies de antóceros, 2) *Marchantiophyta* para as espécies de hepáticas, e 3) *Bryophyta* para as espécies de musgos. A plataforma digital *Tropicos* também é muito útil para conferência dos níveis taxonômicos citados acima.

Informações ausentes foram extremamente comuns em todas as coleções analisadas, especialmente para os campos referentes a filo, ordem e classe. A ausência dessas informações, em muitos casos pode ocorrer por estas serem tidas como "óbvias" para o grupo das briófitas, e por isso não há uma preocupação com a sua digitalização. Contudo, a não inclusão de informações pode acabar por comprometer os resultados de buscas nas plataformas digitais, gerando dados subestimados e não condizentes com a realidade. Em suma, tanto a disponibilização de registros com informações inconsistentes quanto registros com informações ausentes ou incompletas, em coleções virtuais, pode prejudicar as buscas realizadas implicando na redução da quantidade de informação disponível para responder determinada pergunta, na obtenção de conclusões errôneas e em um maior gasto de tempo para realização da pesquisa, devido à necessidade de tempo para organização e limpeza dos dados antes de sua utilização (Chapman, 2005).

Para garantir a melhor qualidade possível dos dados que são disponibilizados virtualmente, é extremamente importante que as coleções incluam a qualificação taxonômica dos acervos, a validação dos dados referentes aos registros de cada lote/espécime, bem como possuam processos de análise de qualidade, limpeza/correções das informações e georreferenciamento (Peixoto et al., 2006). Além disso, embora seja fundamental para garantir uma maior "adequação ao uso" para os usuários, o processo de detecção e correção de erros pode ser demorado e tedioso, incluindo a eliminação de registros inválidos, contudo ele não deve ser ignorado (Williams et al., 2002). A necessidade de limpeza de dados está centrada na melhoria da sua qualidade, reduzindo a quantidade de erros e melhorando a documentação e apresentação desses dados (Chapman, 2005).

CONCLUSÃO

É fundamental que a qualidade dos dados nos bancos de dados virtuais seja continuamente monitorada por especialistas como forma de estimular a produção segura de pesquisas a partir desses dados. Também, faz-se necessário que a pessoa responsável pela digitalização das exsiccatas de briófitas receba um rápido treinamento sobre o grupo, sobre a importância da padronização das informações, da atenção à qualidade dos dados que serão disponibilizados e sobre as plataformas utilizadas para encontrar informações sobre as espécies. Uma sugestão é que um banco de dados digital secundário seja mantido pelo herbário, o qual deverá servir como base para consulta pelos responsáveis

pela digitalização, onde deverão constar as seguintes informações sobre cada espécie presente na coleção, entre outras: filo, ordem, classe e família à qual pertencem, bem como a forma de escrita adotada para o autor de cada espécie. Este banco de dados evitará que cada novo responsável pela digitalização busque informações sobre as espécies em diferentes plataformas, evitando as diferentes formas de escrita em nomes de autores, por exemplo. Também é fundamental que seja definida e estabelecida uma metodologia para conferência das informações contidas em planilhas geradas por diferentes pessoas antes de proceder a sua importação para o banco de dados virtual da instituição.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela bolsa de estágio Pós-Doutoral vinculado ao Projeto nº 23038.000776/2017-54 (INCT - Herbário Virtual da Flora e Dos Fungos), concedida à primeira autora; ao Dr. Geadelande Carolino Delgado Júnior e à doutoranda Wanessa Vieira Menezes Batista pela leitura e sugestões no manuscrito.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRADLEY, B.A. 2013. Distribution models of invasive plants over-estimate potential impact. *Biological Invasions* 15: 1417-1429.
- BRUMMITT, R.K.; POWELL, C.E. 1992. *Authors of Plant Names Standard. International Working Group on Taxonomic Databases (TDWG)* <http://www.tdwg.org/standards/101>.
- CHAPMAN, A.D. 2005. *Principles and Methods of Data Cleaning – Primary Species and Species, version 1.0*. Report for the Global Biodiversity Information Facility Copenhagen. Disponível em: <https://www.gbif.org/document/80545/uses-of-primary-species-occurrence-data>. Acesso em 20 de Dezembro de 2020.
- COLLI-SILVA, M.; BEZERRA, T.L.; FRANCO, G.A.D.C.; IVANAUSKAS, N.M.; SOUZA, F.M. 2016. Registros de espécies vasculares em unidades de conservação e implicações para a lista da flora ameaçada de extinção no estado de São Paulo. *Rodriguésia* 67(2): 405-425.
- COSTA, D.P.; PERALTA, D.F. 2015. Bryophytes diversity in Brazil. *Rodriguésia* 66: 1063-1071.
- CRANDALL-STOTLER, B.; STOTLER, R.E.; LONG, D.G. 2009. Morphology and classification of the Marchantiophyta. In: GOFFINET, B.; SHAW, A.J. (Eds.), *Bryophyte Biology*. Cambridge University Press, p. 1-54.
- DALCIN, E.C.; SILVA, L.A.E.; CABANILLAS, C.C.; LOURES, M.G.S.M.; MONTEIRO, V.F.; SILVA, G.Z.; SOUZA, J.M. 2012. Data Quality Assessment at the Rio de Janeiro Botanical Garden Herbarium Database and Considerations for Data Quality Improvement 8th International Conference on Ecological Informatics (ISEI). Brasília, 3-7 December.
- FLANNERY MC. 2013. Plant Collections Online: Using Digital Herbaria in Biology Teaching. *Bio-science* 39(1): 3-9.
- FUNK, V. 2003. The importance of herbaria. *Plant Science Bulletin* 49: 94-95.
- GADELHA-NETO, P.C.; LIMA, J.R.; BARBOSA, M.R.V.; BARBOSA, M.A.; MENEZES, M.; PÔRTO, K.C.; WARTCHOW, F.; GIBERTONI, T.B. 2013. *Manual de Procedimentos para Herbários*. Peixoto AI, Maia LC (Orgs.). Editora Universitária UFPE, Recife, Pernambuco.
- GASPER, A.L.; STEHMANN, J.R.; ROQUE, N.; BIGIO, N.C.; SARTORI, A.L.B.; GRITZ, G.S. 2020. Brazilian herbaria: an overview. *Acta Botanica Brasílica* 34(2): 352-359.
- GLADSON, W.A.; FARNEZI, M.M.M.; PEREIRA, I.M.; MACHADO, E.L.M.; SILVA, M.A.P. 2018. Modelagem de distribuição de Candeia para Minas Gerais. *Enciclopédia Biosfera* 15(28): 677-686.

- GLIME, J.M. 2017. Introduction. Chapt. 1. In: GLIME JM. *Bryophyte Ecology*. Volume 1. Physiological Ecology. Ebook sponsored by Michigan Technological University and the International Association of Bryologists. Disponível em: <https://digitalcommons.mtu.edu/bryophyte-ecology/>. Acesso em 20 de Dezembro de 2020.
- GOFFINET, B.; SHAW, A.J. 2009. *Bryophyte Biology*. 2 ed. Cambridge University Press.
- GOFFINET, B.; BUCK, W.R.; SHAW, A.J. 2009. Morphology and classification of the Bryophyta. In: GOFFINET, B.; SHAW, A.J. (Eds.) *Bryophyte Biology*. Cambridge University Press, p. 56-138.
- KISLOV, D.E.; BAKALIN, V.A.; PIMENOVA, E.A.; VERKHOLAT, V.P.; KRESTOV, P.V. 2017. An electronic management system for a digital herbarium: development and future prospects. *Botanica Pacifica* 6(2): 59-68.
- KOVTONYUK, N.K.; HAN, I.V.; GATILOVA, E.A. 2018. Digitization of vascular plant herbarium collections at the Central Siberian Botanical Garden, Novosibirsk, Russia. *Skvortsovia* 4(3): 100-111.
- MACIEL-SILVA, A.S.; FARIAS, R.S.; PÔRTO, K.C. 2015. Os gêneros Calymperes e Syrrhopodon (Calymperaceae, Bryophyta) em Floresta Atlântica do Nordeste brasileiro: aspectos reprodutivos e distribuição. *Pesquisas, Botânica* 67: 273-286.
- NAGALINGUM, N.S.; KNERR, N.; MISHLER, B.D.; CARGILL, D.C. 2014. Overlapping fern and bryophyte hotspots: assessing ferns as a predictor of bryophyte diversity. *Telopea* 17: 383-392.
- NELSON, G.; ELLIS, S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions Royal Society B* 374: 20170391.
- PAGE, L.M.; MACFADDEN, B.J.; FORTES, J.A.; SOLTIS, P.S.; RICCARDI, G. 2015. Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience* 65: 841-842.
- PEARSON, K.D.; NELSON, G.; ARONSON, M.F.J.; BONNET, P.; BRENSKELLE, L.; DAVIS, C.C.; DENNY, E.G.; ELLWOOD, E.R.; GOËAU, H.; HEBERLING, J.M.; JOLY, A.; LORIEUL, T.; MAZER, S.J.; MEINEKE, E.K.; STUCKY, B.J.; SWEENEY, P.; WHITE, A.E.; SOLTIS, P.S. 2020. Machine Learning Using Digitized Herbarium Specimens to Advance Phenological Research. *BioScience* 70: 610-620.
- PEIXOTO, A.L.; BARBOSA, M.R.V.; MENEZES, M.; MAIA, L.C. 2006. Diretrizes e estratégias para a modernização de coleções botânicas brasileiras tendo como foco a formação de taxonomistas e a consolidação de sistemas integrados de informação sobre biodiversidade. Disponível em <http://www.cria.org.br/cgee/documentos/botanica.doc>. Acesso em 28 de agosto de 2020.
- PEIXOTO, A.L.; BARBOSA, M.R.V.; CANHOS, D.A.L.; MAIA, L.C. 2009. Coleções botânicas: objetos e dados para a ciência. In: Granato M, Rangel M (orgs.). *Cultura material e patrimônio da Ciência e Tecnologia*. Museu da Astronomia e Ciências Afins, Rio de Janeiro.
- PYKE, G.H.; EHRLICH, P.R. 2010. Biological collections and ecological/environmental research: A review, some observations and a look to the future. *Biological Reviews* 5: 247-266.
- REDMAN, T.C. 1996. *Data Quality for the Information Age*. Artech House Inc.
- RENZAGLIA, K.S.; VILLAREAL, J.C.; DUFF, R.J. 2009. New insights into morphology, anatomy, and systematic of hornworts. In: GOFFINET, B.; SHAW, A.J. (Eds.) *Bryophyte Biology*. Cambridge University Press, p. 138-171.
- SCIAMARELLI, A.; TORGESKI, M.R. 2019. Evaluation of the distribution models of “buriti” and “paratudo”, arboreal species of the pantanal, with data of the quaternary and the present climate. *R. Ra’e Ga* 46(3): 101–112.
- SILVA, M.P.P.; KAMINO, L.H.Y.; PÔRTO, K.C. 2014. Is the current network system of protected areas in the Atlantic Forest effective in conserving key species of bryophytes? *Tropical Conservation Science* 7(1): 61-74.

- SIMÕES, S.S.; ZAPPI, D.C.; AONA, L.Y.S. 2020a. A família Cactaceae no Parque Nacional de Boa Nova, Estado da Bahia, Brasil. *Hoehnea* 47: e1102019.
- SIMÕES, S.S.; ZAPPI, D.C.; COSTA, G.M.; OLIVEIRA, G.; AONA, L.Y.S. 2020b. Spatial niche modelling of five endemic cacti from the Brazilian Caatinga: Past, present and future. *Austral Ecology* 45(1): 35-47.
- SOLTIS, P.S. 2017. Digitization of herbaria enables novel research. *American Journal of Botany* 104: 1281-1284.
- SOLTIS, P.S.; NELSON, G.; JAMES, S.A. 2018 Green digitization: online botanical collections data answering real-world questions. *Applications in Plant Sciences* 6(2): e01028.
- STEVENSON, L.A.; GONZÁLEZ-OROZCO, C.E.; KNERR, N.; CARGILL, D.C.; MILLER, J.T. 2012. Species richness and endemism of Australian bryophytes. *Journal of Bryology* 34(2): 101-107.
- ZHANG, K.; SUN, J.; TAO, J. 2020. Impact of Climate Change on the Distribution of *Euscaphis japonica* (Staphyleaceae) Trees. *Forests* 11(5): 525.
- WEAVER, W.N.; NG, J.; LAPORT, R.G. 2020. Leaf Machine: Using machine learning to automate leaf trait extraction from digitized herbarium specimens. *Applications in Plant Sciences* 8(6): e11367.
- WILLIAMS, P.H.; MARGUILES, C.R.; HILBERT, D.W. 2002. Data requirements and data sources for biodiversity priority area selection. *Journal of Biosciences* 27(4): 327-338.

Tabela 1. Número de registros obtidos para cada coleção virtual de briófitas, de acordo com as informações do filtro selecionado. O número entre parênteses na coluna Registros disponíveis no SpeciesLink se refere ao número total de registros em cada herbário virtual.

Coleção virtual	Filtro de busca	Registros disponíveis no SpeciesLink
ALCB	Filo "Briófitas"	5.454
	Filo "Marchantiophyta"	7.215
	Filo "Bryophyta"	79
	Filo "Em branco"	276 (1.233)
ASE	Filo "Bryophyta"	198
	Filo "Marchantiophyta"	41
	Filo "Em branco"	0 (316)
HABIT	Filo "Em branco"	3.491 (3.518)
HUEFS	Filo "Bryophyta"	2.950
	Filo "Em branco"	474 (567)
	Filo "Ascomycota"	12
JPB	Filo "Moss"	5
	Filo Briophyta"	556
	Filo "Em branco"	134 (3129)
UFP	Filo "Em branco"	16.525 (65.546)
SLUI	Filo "Bryophyta"	16
	Filo "Em branco"	0 (272)

Tabela 2. Número de registros contendo erros para cada coleção virtual, de acordo com as categorias de erros. O número entre parênteses se refere ao tamanho da coleção de briófitas analisada. As informações referentes à Coordenadas Geográficas foram obtidas na plataforma SpeciesLink.

Coleção Virtual	Categorias de erros			
	Taxonômico	Nomenclatural	Geográfico	Incompleto ou ausente
ALCB (13.024 registros)	Filo - 331 Família - 119 Gênero - 33 Espécie - 175 Subespécie - 11	Filo - 5.454 Autor - 1.481	Coordenadas Geográficas - 386 Estado - 77 Município - 03	Filo - 276 Ordem - 13.020 Classe - 13.020 Autor - 1.304
ASE (239 registros)	-	Autor - 16	-	Ordem - 239 Classe - 239
HABIT (3.491 registros)	Família - 160 Gênero - 28 Espécie - 04	Autor - 39	Coordenadas Geográficas - 2.632 Estado - 10 Município - 22 Localidade - 37	Filo - 3.491 Ordem - 3.491 Classe - 3.491 Autor - 01
HUEFS (3.433 registros)	Filo - 901 Ordem - 3.059 Classe - 12 Família - 592 Gênero - 65 Espécie - 93 Subespécie - 19	Autor - 136	Coordenadas Geográficas - 66 Município - 01 Localidade - 44	Filo - 471 Ordem - 374 Classe - 3.421 Autor - 126
JPB (612 registros)	Classe - 556 Família - 145 Gênero - 84 Espécie - 89	Filo - 561 Autor - 189	Coordenadas Geográficas - 06 Localidade - 13	Filo - 51 Ordem - 612 Classe - 56
UFP (16.525 registros)	Família - 1.389 Gênero - 183 Espécie - 326 Subespécie - 14	Autor - 4.026	Coordenadas Geográficas - 258 Município - 14 Localidade - 07	Filo - 16.525 Ordem - 16.523 Classe - 16.523 Família - 02 Autor - 2.022
SLUI (16 registros)	Filo - 05	-	Coordenadas Geográficas - 06 Município - 04 Localidade - 03	Ordem - 16 Classe - 16

Tabela 3. Valores proporcionais de erros por categoria em cada coleção analisada.

Coleção Virtual	Erros por categoria (%)			
	Taxonômico	Nomenclatural	Geográfico	Incompleto ou ausente
ALCB	5,1	53,2	3,6	99,9
ASE	0	6,7	0	100
HABIT	5,5	1,1	75,4	100
HUEFS	89	3,96	1,5	99,6
JPB	90,8	91,6	3,1	100
UFP	9,6	24,4	1,5	100
SLUI	31,5	0	25	100

Tabela 4. Proporção de registros para cada tipo de erro em cada categoria identificada para as sete coleções analisadas. As coleções foram enumeradas de acordo com a ordem decrescente do Índice de Qualidade dos dados.

Coleção Virtual	Categoria/Erros																			Índice de qualidade
	Taxonômico							Nomenclatural		Geográfico				Incompleto ou ausente						
	Filo	Ordem	Classe	Família	Gênero	Sp.	Subsp	Filo	Autor	Coord Geo.	Estado	Município	Localidade	Filo	Ordem	Classe	Família	Autor		
ASE	0	0	0	0	0	0	0	0	6,7	0	0	0	0	0	100	100	0	0	104	
HABIT	0	0	0	4,6	0,8	0,1	0	0	1,1	75	0,3	1	1,1	100	100	100	0	0,02	86	
SLUI	31,25	0	0	0	0	0	0	0	0	37,5	0	25	18,75	0	100	100	0	0	85	
ALCB	2,5	0	0	0,9	0,25	1,3	0,1	41,9	11,4	3	0	0,6	0,02	2,1	99,95	99,95	0	10	77	
UFP	0	0	0	8,4	1,1	2	0,1	0	24,4	1,6	0	0,1	0,04	100	100	100	0,01	12,2	75	
JPB	0	0	90,8	23,7	13,7	14,5	0	91,7	30,9	1	0	0	2,1	8,3	100	9,15	0	0	71	
HUEFS	26,2	89,1	0,3	17,2	1,9	2,7	0,55	0	4	1,9	0	0,02	1,3	13,7	10,9	99,65	0	3,7	61	