Biota Neotropica 22(spe): e20221394, 2022
www.scielo.br/bn

biotaneotropica
ISSN 1676-0611 (online edition)
Thematic Review

*species*Link: rich data and novel tools for digital assessments of biodiversity

***Dora Ann Lange Canhos*[1]*, *Eduardo A. B. Almeida*[2], *Ana Lucia Assad*[3],**

***Mercedes Maria da Cunha Bustamante*[4], *Vanderlei Perez Canhos*[1], *Arthur David Chapman*[5],**

***Renato De Giovanni*[1], *Vera Lúcia Imperatriz-Fonseca*[6], *Lúcia Garcez Lohmann*[6], *Leonor Costa Maia*[7],**

***Joseph T. Miller*[8], *Gil Nelson*[9], *A. Townsend Peterson*[10], *José Rubens Pirani*[6], *Sidnei de Souza*[1],**

***João Renato Stehmann*[11] & *Barbara Thiers*[12]**

[1]*Centro de Referência em Informação Ambiental, Av. Romeu Tórtima, 388, Barão Geraldo, 13084-79, Campinas, São Paulo, SP, Brasil.*
[2]*Universidade de São Paulo, Departamento de Biologia, Ribeirão Preto, São Paulo, SP, Brasil.*
[3]*Associação Brasileira de Estudos das Abelhas, São Paulo, SP, Brasil.*
[4]*Universidade de Brasília, Instituto de Ciências Biológicas, Departamento de Ecologia, Brasília, DF, Brasil.*
[5]*Australian Biodiversity Information Services, Melbourne, Australia.*
[6]*Universidade de São Paulo, Instituto de Biociências, São Paulo, SP, Brasil.*
[7]*Universidade Federal de Pernambuco, Recife, PE, Brasil.*
[8]*Global Biodiversity Information Facility, Copenhagen, Denmark.*
[9]*Integrated Digitized Biocollections, Gainesville, Florida, USA.*
[10]*University of Kansas, Lawrence, Kansas, USA.*
[11]*Universidade Federal de Minas Gerais, Departamento de Botânica, Belo Horizonte, MG, Brasil.*
[12]*New York Botanical Garden, Bronx, New York, USA.*
[*]*Corresponding author: dora@cria.org.br*

CANHOS, D.A.L., ALMEIDA, E.A.B., ASSAD, A.L., BUSTAMANTE, M.M.C., CANHOS, V.P., CHAPMAN, A.D., DE GIOVANNI, R., IMPERATRIZ-FONSECA, V.L., LOHMANN, L.G., MAIA, L.C., MILLER, J.T., NELSON, G., PETERSON, A.T., PIRANI, J.R., SOUZA, S., STEHMANN, J.R, THIERS, B. ***species*Link: rich data and novel tools for digital assessments of biodiversity**. Biota Neotropica 22(spe): e20221394. https://doi.org/10.1590/1676-0611-BN-2022-1394

***Abstract*:** *species*Link is a large-scale biodiversity information portal that exists thanks to a broad collaborative network of people and institutions. CRIA's involvement with the scientific community of Brazil and other countries is responsible for the significant results achieved, currently reaching more than 15 million primary biodiversity data records, 95% of which are associated with preserved specimens and about 25% with high-quality digital images. The network provides data on over 200,000 species, of which over 110,000 occur in Brazil. This article describes thematic networks within *species*Link, as well as some of the most useful tools developed. The importance and contributions of *species*Link are outlined, as are concerns about securing stable budgetary support for such biodiversity data e-infrastructures. Here we review the value of *species*Link as a major source of biodiversity information for research, education, informed decision-making, policy development, and bioeconomy.
Keywords: *speciesLink, data, biodiversity, information network, e-infrastructures, tools, biological collections, microorganisms, pollinators, bees, botany, plants, flora, fungi, algae, gaps, assessment, research, education, conservation, policy.*

*species*Link: dados valiosos e ferramentas inovadoras para avaliações digitais da biodiversidade

***Resumo*:** *species*Link é um portal de informações em larga escala sobre biodiversidade, que existe graças a uma ampla rede colaborativa de pessoas e instituições. O envolvimento do CRIA com a comunidade científica do Brasil e de outros países é responsável pelos resultados expressivos alcançados, atingindo atualmente mais de 15 milhões de registros de dados primários de biodiversidade, sendo 95% associados a espécimes preservados e cerca de 25% a imagens digitais de alta qualidade. A rede fornece dados sobre mais de 200.000 espécies, das quais mais de 110.000 ocorrem no Brasil. Este artigo descreve as redes temáticas do *species*Link, bem como algumas das ferramentas mais úteis desenvolvidas. A importância e as contribuições do *species*Link são destacadas, assim

como as preocupações em garantir um apoio financeiro estável para e-infraestruturas de dados sobre biodiversidade. Aqui revisamos o valor do *species*Link como uma das principais fontes de informação sobre biodiversidade para pesquisa, educação, tomada de decisão, desenvolvimento de políticas e bioeconomia.

Palavras-chave: *speciesLink, dados, biodiversidade, rede de informação, e-infraestruturas, ferramentas, coleções biológicas, microrganismos, polinizadores, abelhas, botânica, plantas, flora, fungos, algas, lacunas, avaliação, pesquisa, educação, conservação, política.*

A Bit of History

In June 1992, Brazil hosted the United Nations Conference on Environment and Development, which included the goal of "establishing an equitable global partnership through the creation of new levels of cooperation among States, key sectors of societies, and people …" (U.N. Conferences, Rio 1992). Chapter 40 of the meeting report, known as Agenda 21, included the following statement:

> *There is a general lack of capacity, particularly in developing countries, and in many areas at the international level, for the collection and assessment of data, for their transformation into useful information, and their dissemination. There is also need for improved coordination among environmental, demographic, social, and developmental data, and information activities.*" (Agenda 21, 1992).

At the meeting, the Convention on Biological Diversity (CBD) was opened for signature until June 4, 1993, at which time it had received 168 signatures. CBD entered into force on December 29, 1993. Article 17. Exchange of Information indicates:

> *"The Contracting Parties shall facilitate the exchange of information, from all publicly available sources, relevant to the conservation and sustainable use of biological diversity, taking into account the special needs of developing countries".*

CBD catalyzed considerable new activity, some of it built on national initiatives from different parts of the world (Chapman 2017). In 1993, a meeting was organized by the *Base de Dados Tropical* (BDT), which later became the *Centro de Referência em Informação Ambiental* (CRIA), in Campinas, bringing together people working on biodiversity informatics initiatives in Australia (ERIN – Environmental Resources Information Network and the Australian National Botanic Gardens), Mexico (CONABIO), Costa Rica (InBio), Finland (FinBIN), and Ecuador (BioBanco), among others. This meeting held in-depth discussions on the exchange of information and ideas on technology, software, and methodologies to further the aims of Agenda 21. A second, meeting held in 1994, established the Biodiversity Information Network – Agenda 21 (BIN21), an informal collaborative network of like-minded initiatives.

This period saw many relevant developments in the emerging field of biodiversity informatics, including a meeting of the Taxonomic Databases Working Group (TDWG; now Biodiversity Information Standards) held in 1992 in Xalapa, Mexico, to discuss methods for database development and data exchange. These were the initial discussions that produced the Darwin Core biodiversity information standard, today accepted internationally (Wieczorek et al. 2012). This period also witnessed a broad rollout of digitization efforts aimed at capturing the full content of biological collections.

Between 1996 and 1999, members of BDT/CRIA were involved in formulating a program on biodiversity for the *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP) to promote concrete actions to implement the information-sharing terms of the Convention on Biological Diversity. The Biota-FAPESP program emerged from this initiative in 1999. Other information systems were being developed at the same time, including the North American and the Inter-American biodiversity information networks (NABIN and IABIN), Species Analyst, Red Mexicana de Información de la Biodiversidad (REMIB), and the Global Biodiversity Information Facility, GBIF (Edwards et al. 2000), among others. GBIF resulted from a mega-science initiative established by the Organization for Economic Co-operation and Development (OECD) that aimed to provide a global-scale perspective, including multiple collaborative efforts to support local initiatives such as *species*Link.

*species*Link's Development

1. First steps

Organizing and making data and information available online was central when first discussing a biodiversity research program for FAPESP. Ideas such as "*not losing data that are born digital*" and "*sharing data openly online*" guided all discussions, inspired by existing biodiversity data systems. The first information system developed for the Biota-FAPESP program was SinBiota (FAPESP Grant # 98/05117-1, 1999–2002). This information system aimed at storing and integrating data from Biota-Fapesp's first projects that carried out biodiversity surveys across the state of São Paulo, with the State's cartographic base updated by the *Instituto Florestal* of São Paulo. CRIA was responsible for the development and maintenance of SinBiota up to 2010 when the system was transferred to UNICAMP, under the responsibility, supervision, and coordination of Biota/FAPESP.

In 2001, the second information system approved by FAPESP was *species*Link (FAPESP Grant # 01/02175-5, 2001–2004). This project aimed at integrating SinBiota with legacy data associated with specimens deposited in 12 biological collections housed at universities and research institutes within the state of São Paulo. Besides developing a distributed system to integrate specimen occurrence data, the project also aimed at developing mathematical models to predict species' geographic distribution, in collaboration with the University of Kansas (KU). KU was also responsible for CRIA's involvement with the communication protocol *Distributed Generic Information Retrieval* (DiGIR), one of the first developed to exchange biodiversity data.

Some key premises guided *species*Link's development. First, it was clear that each biological collection should maintain full control and responsibility for its data. Second, the network would accept all data available, meaning that the system would not filter "bad" data.

Within this context, CRIA could help data providers in finding errors or incomplete data. Data providers could correct possible errors in their system and return corrected and updated information to the network. By maintaining a system that allowed for regular updates, CRIA established an important bond with active collections that transcended the initial funding period. As each collection had different levels of expertise and infrastructure, these collections were able to use the software of their choice to manage their data. Flexibility in various aspects of the project ensured broad participation, even of collections with low or unstable internet connections.

At the time, most e-infrastructures were new and the global infrastructure to integrate all datasets (i.e. GBIF) was just beginning. This was an excellent opportunity for networks like CRIA to build partnerships, share experiences with other infrastructures, and participate in the development of internationally accepted data standards and communication protocols. What was initially thought of as a distributed network very quickly developed into a collaborative effort, led by the needs of data providers and users that presented new ideas, ultimately resulting in new tools and products. Data providers and data users, together with *species*Link's developers, Brazilian colleagues, and international collaborators, all contributed to the development of *species*Link, therefore constituting its innovation center.

An important milestone in the establishment of *species*Link was an international event sponsored by FAPESP, CNPq, and Petrobras, and organized by CRIA in October 2002 in Indaiatuba, Brazil (Zorzetto 2002). The event referred to as the "*Indaiatuba meeting*", assembled different initiatives, organized in several working groups to define strategies and next steps in their fields.

Meetings with the botanical community included discussions on the Botanical Society of Brazil's (SBB) strategic plan for the development of Brazil's Flora. Important references included *Flora Brasiliensis*, at the time Brazil's only flora, published between 1840 and 1906, as well as regional and state floras carried out as part of SBB's strategy. Other important initiatives presented at the meeting included the digitization strategy of New York´s Botanical Garden herbarium and Australia's Virtual Herbarium. The meeting also included working groups on pollination biology that discussed, among other topics, the compilation of a list of bees, including Moure's Bee Catalogue and the North American Integrated Taxonomic Information System (ITIS). Other working groups included Species 2000 and ITIS, to discuss the Catalogue of Life, an index of all species. In parallel, TDWG worked on standard data fields. It was also at this event that *species*Link's prototype was demonstrated online with data from one biological collection. The Indaiatuba meeting was key to *species*Link's development, promoting new partnerships and projects, including:

- CRIA´s partnership with Species 2000 (Catalogue of Life);
- CRIA's participation in TDWG – Taxonomic Databases Working Group;
- A cooperative project funded by FAPESP, Natura, and Vitae Foundation to digitize and publish *Flora Brasiliensis* online through a partnership between UNICAMP, Missouri Botanical Garden, and CRIA;
- A project funded by FINEP to digitize, organize, and launch *Moure's Bee Catalogue* online through a partnership between the Federal University of Paraná and CRIA;
- A data repatriation program initiated in 2006 with New York Botanical Garden; and,
- The organization of a working group in 2005 to produce guidelines and strategies for the modernization of Brazilian biological collections and consolidation of biodiversity information systems.

The definition of a strategy for Brazilian biological collections associated with a biodiversity information system was key to *species*Link's evolution as it involved the Brazilian Societies of Botany, Microbiology, and Zoology, RNP (Rede Nacional de Ensino e Pesquisa), responsible for Brazil's academic network, and CRIA. The result of this work (Peixoto et al. 2006) was presented by the Brazilian Government at COP8 held in Curitiba in 2006, as Brazil`s national strategy for biological collections. This work guided CRIA's strategies and work plans in subsequent years, strengthening its partnerships with biological collections and RNP.

While working to share *Flora Brasiliensis* online, also launched during COP8 by Fapesp, CRIA developed the system *Flora Brasiliensis revisited* in partnership with UNICAMP, which allowed specialists to update the scientific names online. This work was more complex than expected, and only a few families were dealt with, such as Bignoniaceae, Clusiaceae, and Cyperaceae, among others. However, this experience established the basis for the development of the online system to coordinate and prepare the *Catálogo de Plantas e Fungos do Brasil* (Forzza et al. 2010a, 2010b). This project was coordinated by the *Jardim Botânico do Rio de Janeiro* (JBRJ) with the participation of hundreds of botanists from all over the world. CRIA was responsible for integrating data from existing lists and from *Flora Brasiliensis revisited* as a baseline for the Catalog and for developing an online system that would allow for new online data entries and corrections, together with an administrative system to coordinate and evaluate data input and output. The Catalog's printed copy and public interface were launched in 2010, meeting Target 1 of the Global Strategy for Plant Conservation (GSPC-CBD). In 2015, CRIA transferred the full system to JBRJ.

In 2005, FAPESP approved the project *OpenModeller – A framework for species distribution modelling* (FAPESP Grant *# 04/11012-0,* 2005–2008). This project aimed to develop tools for data cleaning and ecological niche modelling, while continuing to support the integration of data from biological collections. The project's main product was the development of a cross-platform environment to carry out ecological niche modelling experiments (Muñoz et al. 2009).

Tools and Auxiliary Systems

1. spLinker

Since many biological collections did not meet the requirements for serving data directly to the network, CRIA developed a Desktop application called *spLinker*, allowing the upload of data to the server whenever necessary. This way, unlike other biodiversity networks, even collections without a stable and fast Internet connection, appropriate equipment, and software to serve data to the network 24 hours a day, as well as a qualified team, responsible for maintaining the system, were able to join the network. For the network, any software can be used to manage collection data, as long as it meets the collection's needs. Data fields are mapped following the Darwin Core data model and

spLinker can communicate directly with the database or spreadsheet. Once *spLinker* is configured and fields are mapped, the curator or person responsible for the collection may send non-sensitive data to the network's cache node, allowing the data to be harvested by *species*Link. As all data served to *species*Link are openly accessible, any sensitive data, whether a specific field or record, must be marked as such, so *spLinker* recognizes it and does not send this specific data to the network.

2. *Data quality*

Data quality is a major concern since the beginning of *species*Link, Arthur Chapman spent a year in Campinas (Grant # 02/10039-7, 2003–2004) working with CRIA's team, looking at ways in which data quality management could be incorporated (Chapman 2004). His work at CRIA was built upon his earlier work at ERIN, focused on developing methods for testing and improving data quality, ultimately leading to the publication of several documents by GBIF (Chapman 2005a, 2005b). The first online product at *species*Link produced for collection curators and managers was the ***data-cleaning report***, where a set of tools highlight possible errors or incomplete data and produce an online report to help collection curators and managers identify errors and correct their data. The system verifies scientific names, collection dates, and geographic data and offers suggestions to fill out specific blank fields.

3. *Species names*

The first tool developed to find errors in the specimens' names could not depend on taxonomic authority lists, as, at the time, this information was scattered and incomplete. The solution was to compare scientific names within each collection, highlighting those that were phonetically equal, but written differently, marking them as suspect records. This tool helped users in finding spelling mistakes but did not help in validating scientific names. Today, the search interface uses this phonetic algorithm as a tool for users to expand the amount of data retrieved by being able to include names with minor spelling errors when searching.

Today, when data are harvested, every record has its scientific name (genus and species) checked by one of the following taxonomic references: Flora e Funga do Brasil, Moure's Bee Catalogue, Catalogue of Life, MycoBank, Algaebase, and LPSN (List of Prokaryotic names with Standing in Nomenclature). Based on these lists, each name receives one of the following tags: *accepted*, *synonym*, *ambiguous*, or *not found*. The status *ambiguous* is given when the same scientific name has two different statuses in the same reference list, normally due to different authors and the tool does not check authors.

The new search interface presents a filter for *taxonomic status*, meaning that users, such as collection curators, may search for synonyms or records with *not found* names to identify problems and this way correct their data. On the other hand, users that require good quality data may choose to only search for records with accepted names.

4. *Geographic coordinates*

The system also checks the informed geographic coordinates of all incoming data. For municipalities of Brazil, depending on whether the coordinate falls within the registered municipality or not, the field is tagged as *consistent* or *suspect*. Other parameters are also checked and the field is tagged. This enables users to search for records with suspect geographic coordinates and possibly correct errors or limit their search to consistent data records. The system also has an automatic georeferencing tool for records collected in Brazil with informed municipality, adding three new data fields with the assigned lat, long and maximum error.

5. *Future data quality reports*

In the future, the idea is to enable users to produce data quality reports as an output of the search interface. Some parameters are already available when visualizing *species*Link's data expressed as numbers through the search interface. This output (Figure 1) indicates the viability of expressing data quality through an output, whether it be for a collection curator or researcher wanting to assess the quality of the data retrieved. Some parameters presented, such as scientific names and identification status already indicate data quality.

6. *Duplicates*

Another valuable tool for the curation of botanical specimens is the possibility to locate and compare specimen duplicates. Through *species*Link's search interface, it is possible to compare the identifications of the same specimen (i.e., duplicates) available in the network. The system looks for records with the same collector name, collector number, and date collected and assumes that those records refer to the same collecting event and specimen. Curators can search and retrieve all records from their herbaria and analyze existing duplicates in the network. This tool allows them to verify whether the network has duplicates of their unidentified material or correct potential misidentifications.

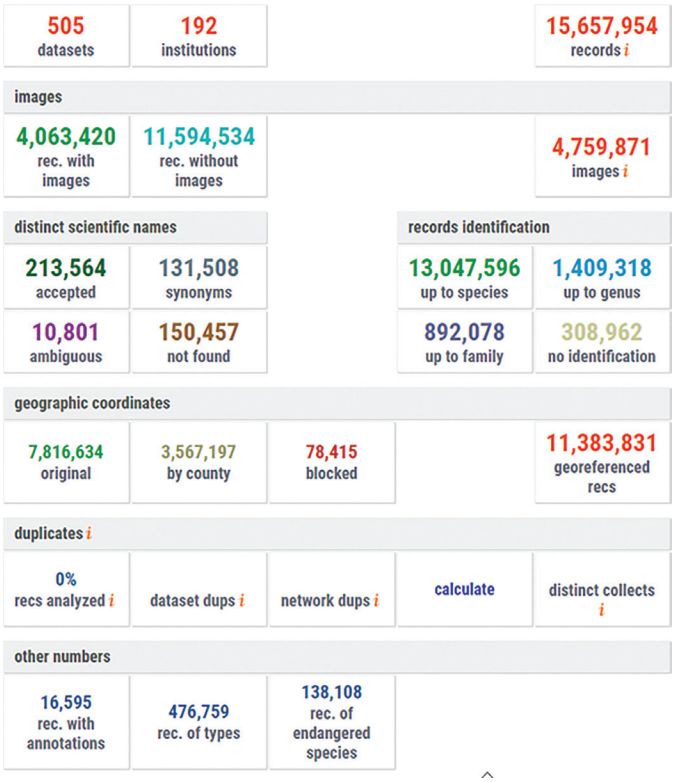

Figure 1. Network output retrieving all *species*Link data expressed in numbers (*species*Link, Jan/2022).

http://dx.doi.org/10.1590/1676-0611-BN-2022-1394

7. *Annotation tool*

*species*Link launched its image service *exsiccatae* in 2011, which, in turn, enabled the development of an *annotation tool* that engages users in improving the quality of online data (Hobern et al. 2013). When finding errors, registered users can edit their comments about a specific data record in a form. The tool sends the form with comments to the collection's curator and automatically adds the comment as an annotation to the record. In ten years, the system received about nineteen thousand annotations, 97% referring to the scientific name. Contributors come from many parts of the world, mostly from institutions that share data with *species*Link.

8. *Lacunas*

This tool has been used for Plants, Algae, and Fungi since 2012 (Canhos et al. 2014) as well as for Neotropical Bees since 2019. The tool *Lacunas* requires a taxonomic list of species that occur in Brazil and their geographic distribution in the country.

Lacunas Flora

This tool helps identify *species*Link's taxonomic and geographic data gaps for native species of plants, algae, and fungi in Brazil. The system displays the status of online data for all valid native species listed in the *Flora e Funga do Brasil*. It also highlights the Brazilian states where specialists indicate that the species occur with the states that do not have occurrence points in *species*Link, representing geographic data gaps within the network. Reports are available for taxonomic groups, families, genera, and species. Selecting angiosperms, for example, the Lacunas' report of July 2022 indicates that the system analyzed 33,172 native species and, using the most inclusive search option, indicates that 1,385 native angiosperm species (4%) do not have any data record in *species*Link. At this level, the system also shows a list of species lacking data for specific Brazilian states. At the family and genus levels, all species for the selected family or genus are listed, classified into four groups: (1) those with no data, (2) those with 1–5 records; (3) those with 6–20 records, and (4) those with >20 records. The report compares the results with previous Lacunas reports so that one can evaluate the degree to which *species*Link is reducing its data gaps. At the species level, the report presents the conservation status, distribution according to Flora e Funga do Brasil and *species*Link, and ecological niche models (when available). Associated with each map is a link to the original information. The report also presents the number of records collected per year, the list of data providers, and the status of the data. This tool helps define strategies to reduce or eliminate gaps, such as new collecting and digitization efforts, inviting new collections to the network, and promoting specialist visits or training.

Lacunas Bees

In July 2019, CRIA adapted and launched the system *Lacunas for Neotropical bees*. The concept is the same as that for plants, algae, and fungi. An important difference is that Moure's Bee Catalog is not limited to bees that occur in Brazil, it refers to bees from the Neotropical region. Therefore, Moure's Bee Catalog supplies information on the list of native bees of the Neotropics and the states where they occur. All species listed are analyzed to also highlight species that are not registered as occurring in Brazil in Moure's Catalog, but for which specimen records collected in Brazil are found in *species*Link. The system produces the same specimen report as for plants, algae, and fungi. Therefore, besides working on *species*Link's data gaps, Lacunas Bees also shows possible data gaps in Moure's Bee Catalog.

9. *OpenModeller and BioGeo*

Relating species occurrence data to the corresponding environmental conditions allow the preparation of ecological niche models that can be used to predict species' geographic potential. Ecological niche modelling is one of the most powerful techniques with which to address current challenges such as the likely impacts of climate change in biodiversity and the potential spread of invasive species, allowing for better conservation decisions and best practices for the selection of the best location for conservation areas, among many others (Peterson et al. 2011). CRIA's prior involvement in the development of Desktop GARP and the early stages of LifeMapper resulted in a new initiative to develop a framework for ecological niche modelling called openModeller (Muñoz et al. 2009). Since its release in 2004, it has evolved into a collaborative effort, effectively resulting in a flexible framework that can run on different platforms, read and write data in different formats, produce models with different algorithms, and be used over different front-end interfaces.

The availability of an in-house ecological niche-modelling tool and millions of species occurrence records through *species*Link led to another development known as BioGeo. BioGeo is a website that allows researchers to navigate across the taxonomy of plants, algae, and fungi that occur in Brazil and generate ecological niche models. Behind the scenes, the website interacts with a series of web services through a complex workflow involving taxonomic data retrieval, occurrence data retrieval, data cleaning, and ecological niche modelling operations. The workflow incorporates different model creation and testing strategies depending on the number of input points available (Giovanni & Bernacci 2015). In response to the scientific names retrieved, researchers can configure the set of names used to search for occurrence data from *species*Link. The system then lists all records retrieved, marking those selected by the system after performing its own set of automatic data cleaning tests (Giovanni et al. 2012). Researchers can review all records, pre-selected or not, and exclude more records than the automated procedures excluded, or can decline automated exclusions. At this point, records confirmed by the researcher, are used to generate a model. As more points are provided to the modelling step, more algorithms can be used to create a richer model ensemble. The resulting model is presented so that the researcher can evaluate the model and decide whether to accept or reject it. Acceptance means making the model publically available (Figure 2). Darker red colors indicate higher environmental suitability for the species and green dots are species occurrence points used to generate the model.

Besides the model, the system presents details as to the number of available and used occurrence records, algorithms used, and a report indicating the specie's real and potential occurrence in Brazilian states and counties. To date, about 5,000 species have distribution models published by BioGeo.

Data Repatriation

In 2003, GBIF commissioned a study to analyze experiences on data sharing with countries of origin (Canhos et al. 2004). The report indicated that most institutions that answered the survey thought that

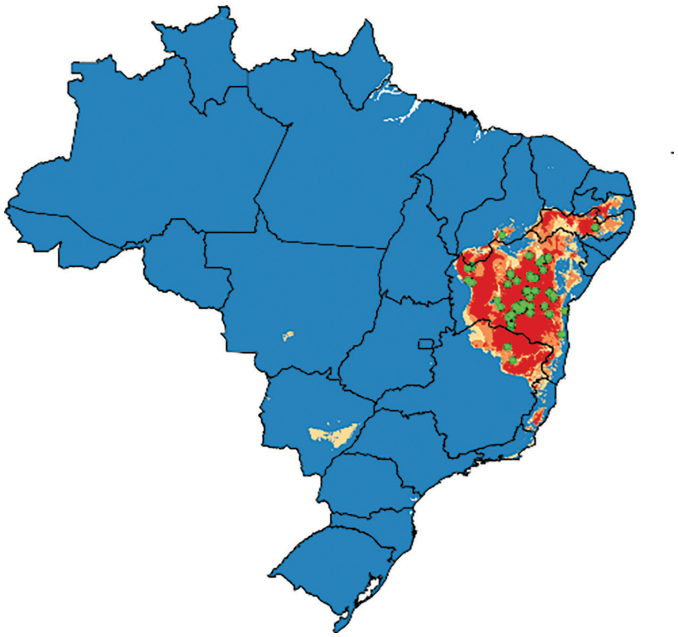

Figure 2. Potential distribution of *Passiflora edmundoi*. (Bernacci & Giovanni, 2013).

data sharing with the country of origin was a valuable spin-off. The rationale was that by making the information freely available, the data becomes available not only to countries of origin but also to anyone else who needs or can benefit from such access. It was clear that making information freely available on the internet was a trend that would continue into the future.

In 2006, New York Botanical Garden was the first collection to share data of samples collected in Brazil with CRIA through *species*Link. As one always assumes that the country that receives the data receives most benefits, it is interesting to learn Barbara Thiers' contribution to this article, expressing her experience as the curator of the largest US herbarium, in sharing data with *species*Link.

> *When, as NY's curator I first met CRIA's team and learned of the plans for species*Link*, the Steere Herbarium of the New York Botanical Garden (NY) was engaged in the digitization of approximately 500,000 Brazilian specimens, a project funded for about 12 years in three phases by the National Science Foundation. NY, like most larger U.S. biodiversity collections, was focused on digitizing subsets of their holdings that could serve the largest number of researchers and students, and that could be completed in the time frame and budget of a standard grant award. Our Brazilian collections were a high priority for digitization because of the strength of our historical holdings as well as more recent ones, e.g. specimens documenting past staff research in the Planalto region, the Flora Amazônica Project, and subsequent work in the Amazon and Atlantic Coastal forests by current NY staff.*
>
> *When approached by CRIA´s team about contributing our Brazilian specimen records to speciesLink, we were happy to do so, though we had never contemplated the export of data and images on such a scale – this was before utilities such as the Integrated Publishing Toolkit (IPT) produced by the Global Biodiversity Information Facility (GBIF) became available. However, the tools already built into CRIA for that import greatly facilitated the process, and once imported, we were delighted to find that NY specimen data became widely available to Brazilian scientists for work on the nascent Checklist and Brazilian flora projects as well as other research projects. Feedback from users indicated that our data needed a lot of cleanups, however – historical specimen records, mostly transcribed by NY staff with no knowledge of Brazilian flora or geography had many errors in collector, plant, and place names. After our data were in speciesLink we could take advantage of the various data cleaning tools provided by CRIA, a rather sobering indication of how far some of our data were from being usable, but a useful guide as to how to improve our data.*
>
> *Although the groundbreaking data cleanup and analytic tools by speciesLink have helped enormously to clean errors in our Brazilian data, and the most egregious errors in our worldwide data, U.S. institutions have not been able to create, either individually or collectively, the sort of tools that speciesLink has long made available for the study of Brazilian biodiversity. speciesLink was consulted during the development of the iDigBio database, and this would be the logical place for the development of the types of data cleaning tools that CRIA has developed; however, the mission of the endeavor and scope of the data provided by iDigBio has not allowed this activity yet. After a decade of digitization of specimens, we are now at a point where we could create a national portal with the type of tools that CRIA provides through speciesLink. Such tools ideally would not only highlight data errors but also provide tools for their correction and tracking of changes made by the database… Should the collections' community in the U.S. find a way to fund the development of tools specifically for the cleanup and standardization of biodiversity collections from within our borders, we will surely depend heavily on the work of CRIA. Not only have they been a continual source of inspiration and new ideas, but they are always willing to discuss and share ideas and thus are invaluable colleagues for U.S. collections and collections worldwide* (Thiers, B. 2022).

CRIA and all *species*Link users greatly benefited from, not only the data and images shared by NY but through the collaboration established. Besides being the first collection from abroad to share data with *species*Link, NY was also the first to share images and expand its geographic scope to all of South America. This is very important for studies of Brazilian biomes, most of which are not limited to its political boundary. NY's participation also set an important and large-scale example, which helped greatly to attract additional collaborators and participating institutions worldwide.

Thematic Networks

Special projects of different taxonomic groups led to the organization of thematic networks within *species*Link. These networks are not just about integrating data, but working as a community to identify specific needs and develop tools and outputs to attend these needs. These networks, of which CRIA's team is part, also represent *species*Link's innovation center. Working together with these different groups became an important strategy not only to meet their requirements, but also to offer many of these developments to all biological collections of the network. Other biological collections have great potential in forming such thematic networks within *species*Link, such as marine biological collections of invertebrates and algae of São Paulo state that already share their data through *species*Link (Borges et al. (in press)).

 http://dx.doi.org/10.1590/1676-0611-BN-2022-1394

1. *Microorganisms*

Microbial collections are an important source of genetic resources and reference material for research and technological development on which biotechnology is founded. The global biotechnology market for products derived from genetic resources at the turn of the century was in the range of US$500–800 B per year (Kate & Laird 1999). This market is continuously expanding, with the provision of new products derived from the prospection of microbiological materials and metagenomes. The prospection of material from unusual and/or extreme environments is impacting the bioeconomy of diverse sectors (Canhos & Manfio 2004, Jorquera et al. 2019, Giudice and Gugliandolo 2019). Technological advances in instrumentation, automation, genomics, and data mining are bringing new perspectives for the sustainable exploitation of biodiversity. New strategies for the prospection of biomolecules are increasingly dependent on the comparative analysis of digital data sets. Open access to quality data on biogeography and ecological context of collection events, taxonomic information, traits, and genomics derived data, including digital sequence information, are fundamental for the development of new strategies (GNAS 2014, GNAS 2021).

To set a basis for the establishment of a global biotechnology infrastructure, the Organization for Economic Cooperation and Development (OECD) Biotechnology Forum established a panel to discuss the challenges and opportunities involved in creating a Global Biological Resource Centers' Network. Publication of the document *Biological Resource Centers: Underpinning the Future of Life Sciences and Biotechnology* (OECD 2001) was followed by the establishment of the OECD Biological Resource Centers Working Group, composed of government representatives and experts. In Brazil, in 2002 a network of biological resource centers *SICol* was launched (MCT 2002). The development of *SICol* by CRIA offered an information system for collections of biotechnological interest as a key component of the MCT/Finep program.

Developed as a thematic network of *species*Link, *SICol* provides a set of applications that allow spatial visualization of data, tools for image analysis, an annotation system, and indicators on samples deposited in collections in the network. Parallel to the network, a management system *micro*SICol specifically designed for microbial collections was also developed. Key features include structured recording of data provenance (i.e., locality, collector, and depositor), quality control, taxonomic information, trait data, and technological applications. In addition, the software allows the insertion of photos, gene sequences, and stock control of preserved strains. The source code and documentation of *micro*SICol are openly available at GitHub.

2. *Pollinators*

Pollinating insects, particularly bees, have attracted great interest over the years thanks to increasing awareness of their critical role in pollination in both natural and agricultural areas (Potts et al. 2010, Klein et al. 2017) and to their bioactive compounds (Costa-Lotufo et al. 2022 (in press)). In addition to being interesting biologically, the value of symbiosis between plants and pollinators for humanity cannot be overstated (Potts et al. 2016). Wild and managed bees contribute to one-third of the total production of food for humans (Klein et al. 2007, 2017). Globally, insects are estimated to contribute more than US$ 235–575 B yearly to the global economy through their role as crop pollinators (Breeze et al. 2016), but actual contributions may be considerably higher. Given their importance to food security and wildland preservation, improving the understanding of pollinators is critical. However, changes in land use, increasing fragmentation and loss of natural habitats, as well as pesticides, pollutants, parasites, diseases, and malnutrition are some of the drivers responsible for reducing local biodiversity throughout the world, raising concerns about declines of native pollinators and potential vulnerabilities of crop and wild plants (Ghazoul 2005, Klein et al. 2007, 2017, Porto et al. 2021).

For the past two decades, *species*Link has become a primary information database for data on pollinating insects in Brazil. Among Brazilian institutions, 28 collections currently contribute data on pollinating insects. *species*Link applies taxonomic concepts from reliable sources to the record data from specimens deposited in biological collections of museums, universities, and other institutions. In the case of bees, the taxonomy uses as reference *Moure's Bee Catalogue* followed by the *Catalogue of Life*.

Moure's Bee Catalogue was originally published as a printed work under the title *Catalogue of the Bees (Hymenoptera, Apoidea) in the Neotropical Region* (Moure et al. 2007). Later it was published in collaboration with CRIA as an open-access online version with three subsequent editions (Moure et al. 2008, 2012, 2022), all coordinated by Prof. Gabriel A. R. Melo, of the Zoology Department at the Federal University of Paraná. This catalog became the main reference on bee diversity of the Neotropics. Its publication facilitated access to comprehensive documentation of 264 years of taxonomic research and joint examination of more than 5,000 species names of bees. It represents a major development in making current information on the number of bee species universally available, contributing to the advancement of research on pollinating insects.

The system *Knowledge Gaps of Native Bees of Brazil* (or simply *Lacunas Bees*) previously described in this article, represents an important tool for setting data entry priorities, integrating new collections to the network, and determining potentially relevant areas for future biodiversity research.

Research on taxonomy or areas that rely on taxonomic data, improves the amount and quality of data on the diversity of biological taxa. Although publications of scientific articles and books are essential to communicate novel findings, knowledge generated by a taxonomic revision becomes more readily accessible when integrated into a catalog and in searchable online databases (Hedrick et al. 2020). Similarly, the availability of databases that integrate multiple sources of data from various institutions can positively affect the development of taxonomic research with increased breadth (Yeates et al. 2011). Indeed, the plea for an increasingly *integrative* taxonomy has gained traction, as justified by the notion that integrating data can generate more robust scientific hypotheses. In this context, the impact of digitization is indisputably beneficial for biodiversity research because millions of data points from multiple institutions can be integrated. The importance of natural history collections is increasingly recognized for scientific disciplines as diverse as genomics, conservation, morphometrics, phenology, pollination biology, and adaptation, among others (Holmes et al. 2016, Hedrick et al. 2020).

In 2015, CRIA established a partnership with the nonprofit association *Associação Brasileira de Estudos das Abelhas* (A.B.E.L.H.A.). Its mission is to gather, produce, and disseminate scientifically based

information aimed at conserving Brazilian biodiversity and the harmonious and sustainable coexistence of agriculture with bees and other pollinators. In addition to working with the scientific community, the association interacts with beekeepers and the general public. Thanks to this partnership, several advances were possible, namely: (a) live images of bees added to *species*Link, (b) the development of an online database with bee and plant interactions, and, (c) the development of, *info*Abelha. Through *info*Abelha users can retrieve information from a variety of databases, including Moure's Bee Catalogue and *species*Link, searching for species either using its scientific or common name. Using these online systems, A.B.E.L.H.A. created new products for the general public, such as calendars, posters, folders, and specific materials for beekeepers and schools.

For this group of pollinators, it is important to highlight *species*Link's contribution to understanding climate change impacts on its biodiversity. Besides land-use change, climate change has become a powerful driver of global biodiversity loss, perhaps most notably in tropical regions. Using ecological niche modelling (ENM), it is possible to determine the environmental conditions necessary for the persistence of species and use models of those conditions to assess implications of climate scenarios provided by the Intergovernmental Panel on Climate Change (IPCC). These projections can highlight species or populations that are particularly vulnerable, and point to climate-robust zones that can be preserved, or corridors to facilitate dispersal (Campbell et al. 2019, Miranda et al. 2021, Sabatino et al. 2021).

Several studies have analyzed and mapped potential species distribution areas in diverse Brazilian biomes (e.g. Gomes et al. 2019, Giannini et al. 2021, Zwiener et al. 2018). When ecosystem service providers are considered, several examples from the eastern Amazon region have anticipated precipitous declines in populations of bats (Costa et al. 2018), birds (Miranda et al. 2019), and bees (Giannini et al. 2020) under future climate conditions. Consequently, climate change may also impact food production in the region (Giannini et al. 2012, Giannini et al. 2013, Giannini et al. 2017a, Bezerra et al. 2019a, b).

Specifically focusing on pollinating bees, using IPCC scenarios, the potential distribution area for the Caatinga stingless bee *Melipona subnitida*, was assessed in the context of distribution and distributional shifts of the plants on which it feeds. The result was a proposal of several ecological corridors in the landscape (Giannini et al. 2017b). Also in the Caatinga, another small stingless bee, *Plebeia flavocincta,* was studied, focusing on historic climate changes (Maia et al. 2020) to explain its expanded distribution projected over the next 50 years. Lima & Marchioro (2021) assessed climate change implications on several stingless bee species important in meliponiculture, in terms of expansion or reduction of habitable areas. Under different climate change scenarios, Gonzalez et al. (2021) evaluated which Colombian stingless bee species would become important pollinators while Krechemer and Marchioro (2020) assessed bumblebees (*Bombus*) from across South America. All of these studies were based on data derived at least in part from *species*Link, pointing to a key role for data-sharing initiatives in increasing knowledge about likely climate change impacts on Neotropical biotas.

3. Botany

Botany is *species*Link's most structured group due to its organization within the Sociedade Botânica do Brasil and to the development of the National Institutes of Science and Technology (INCT) program, coordinated by the *Brazilian National Council for Scientific and Technological Development* (CNPq). INCTs play an important role in Brazil's national strategy as operators of science, technology, and innovation. One of the projects approved in this program is the Virtual Herbarium of Flora and Fungi (INCT-HVFF). Another important support to herbaria in São Paulo State comes from Fapesp, including research projects and postgraduate scholarships that have increased their holdings and improved their quality (Mamede and Simão-Bianchini (in press)).

INCT-HVFF is a network of herbaria that openly share data and images online through *species*Link. This network began in December 2008 with 25 Brazilian herbaria, two from the United States, as well as CRIA and RNP as associate members. CRIA is responsible for *species*Link´s development and maintenance, and RNP for Brazil's advanced national network for higher education, research, and innovation. Without RNP's academic network (rede Ipê), it would have been impossible to connect herbaria located at remote universities in the country. Today, 138 herbaria in Brazil and 24 from abroad are associated to INCT-HVFF, the latter sharing data of samples collected in Brazil and other South American countries (Maia et al. 2015).

The network also integrates data and images from Georg Marcgrave's *Herbarium Vivum Brasiliense* held at the Natural History Museum of Denmark, with the first plants collected in the Americas between 1638 and 1644, as well as specialized collections, such as the *Solanaceae Source*, a database of the Natural History Museum, London. *species*Link shares more than 11M data records and 4.6M images of algae, fungi, and plants. As all herbaria have full participation, this partnership resulted in the development of new tools and services that respond to the demands of this large and growing network of data providers and users. The fact that *species*Link does not apply a quality filter upon data entry, enabled several developments that address the demands of curators in improving the quality of their data. Important examples include *data cleaning tools* and the *annotation system* that enables specialists throughout the world to point out errors and identify specimens online. Another important aspect is the full attribution of credit to all participants (Maia et al. 2017, Canhos et al. 2019).

INCT-HVFF integrates collections from institutions from all Brazilian states, constituting one of the largest structured networks of algae, fungi, and plant collections of the world. Working as a network, the virtual herbarium (~11 M records online) surpasses the three largest herbaria in the world: Royal Botanic Gardens (K), Museum National D'Histoire Naturelle (P & PC), and The New York Botanical Garden (NY), each with ~8 M specimens (Thiers, 2021). This massive resource demonstrates the importance and impact of a collaborative network.

Brazil's academic network (RNP), offers internet access throughout the country, enabling full participation of small collections. These herbaria today constitute the largest number of data providers, contributing significant amounts of data on Brazilian plant diversity (Monfils et al. 2020) substantially reducing spatial biases across a large country such as Brazil (Sousa-Baena et al. 2013, Oliveira et al. 2016, Daru et al. 2018). Each herbarium in *species*Link receives the same status within the network, regardless of its location, size, or number of specimens (Canhos et al. 2019). The work developed by large and small

herbaria and the visibility provided by the INCT-HVFF network, help prove the value of collections within their institution as an essential infrastructure for teaching, research, and public outreach in botany and mycology.

During INCT-HVFF's first phase (2008–2016), besides associating images to specimen records and serving them online, two analytical tools were developed, *Lacunas* and *BioGeo*. These tools are currently used to determine digitization strategies and in planning new collecting efforts across Brazil (Canhos et al. 2014, 2015). During the second phase (2016–current), a new search interface was developed, incorporating features such as geographic filters (e.g. biomes, conservation units, watersheds), and tools to evaluate and improve data quality. The possibility to analyze specimen duplicates allows careful examination of the specimens deposited in different collections and, in turn, allows determinations to be verified and updated through cross-referencing. This functionality encourages the exchange of duplicates and sending samples from small collections to large herbaria to be studied by specialists. Most graduate programs in biodiversity are linked to at least one herbarium of the network. These herbaria are crucial for advanced training, especially in the fields of systematics, evolution, and biogeography.

Considerable increase in the usage of data shared through INCT-Virtual Herbarium was noted in 2021, due to changes in *species*Link's technology and in the number of outputs offered through the new search interface, enabling faster and more comprehensive searches as well as online data analysis. The average usage in the first semester of 2022 was of about 76 million data records and 44 thousand images *a day* (statistics available at specieslink.net/usage).

The availability of data and associated images, and a variety of tools for specimen curation were particularly important for taxonomists who worked on the *Catalog of the Flora and Fungi of Brazil* and *The Flora of Brazil 2020*. These initiatives allowed Brazil to reach the goals proposed by the Global Strategy for Plant Conservation (GSPC), a program of the Convention of Biological Diversity (CBD), between 2001–2010 (Forzza et al. 2010a) and between 2011–2020 (BFG 2021). To accomplish the goals of both the Catalog of Brazil and the Flora of Brazil 2020, a large number of researchers worked remotely for decades, gathering information on the occurrence of species in the Brazilian territory, feeding the database that today constitutes the fabulous repository of *Flora of Brazil 2020*, maintained by JBRJ.

During the Covid-19 pandemic, work involving online consultations with herbaria did not stop, even though visits to the collections closed temporarily. In 2020, INCT-HVFF proposed a collective effort to improve the identification of herbarium samples with associated images online. This activity involved the participation of many taxonomic experts and when compared to previous months, resulted in an increase of more than 25 fold in the numbers of annotated records. During the pandemic, other successful online activities were promoted by herbaria associated with INCT-HVFF using the available annotation tool. These initiatives allowed for the correction of substantial amounts of herbarium label data (i.e. Green September, Botany Day, and Identification Competitions). The plant and fungi identification competition, held during the 71st National Congress of Botany in 2021, motivated 24 botanists from different institutions to remotely annotate about 850 specimens from 56 herbaria.

The Importance of *species*Link for Public Policy, Sustainable Development, and Conservation

Biodiversity databases hold rich and detailed information on the distribution, morphology, and other characteristics of organisms over time. During the past decade, freely available digital biodiversity databases have seen an impressive accumulation of digital biodiversity records (Soberón & Peterson, 2009), greatly expanding their potential uses. Indeed, most recent analyses and studies in taxonomy, biogeography, and ecology have taken advantage of the increasingly important sources of information available online (e.g. Buerki & Baker, 2016). Furthermore, additional uses of these data such as research on environmental impact, agriculture, public health, and disease ecology have also emerged (Ball-Damerow et al. 2019).

*species*Link is used extensively in research since its origin, providing the basis for improved knowledge of the Neotropical biota. CRIA'S 2020 annual report (CRIA, 2020) highlights the citations to *species*Link and its systems in articles, books, thesis, and dissertations. Using Google Scholar and GBIF as references, the report indicates that in 2020 alone, 655 peer-reviewed papers, 22 preprints, 13 books or book chapters, 9 doctoral thesis, 35 master dissertations, and 8 undergraduate reports used *species*Link data as the basis for their studies. Of these, 15 peer-reviewed papers were published in high-impact journals (IF over 10), such as Nature Climate Change, Nature Ecology & Evolution, and Annual Review of Plant Biology.

The compilation of data for floristic and taxonomic studies was greatly facilitated and improved by *species*Link. These studies established the foundation for spatial analyses at different spatial scales, especially those focused on identifying centers of diversity, areas of endemicity, and spatial phylogenetic patterns. Mapping tools available through *species*Link have allowed researchers to visualize data efficiently and detect taxa that are known from only a few collections, often old ones with missing or inaccurate georeferenced data. Identifying temporal, spatial, and taxonomic gaps in biodiversity documentation, sometimes even in areas that have been intensively explored (e.g. Sousa-Baena et al. 2013, Colli-Silva et al. 2019, Colli-Silva & Pirani 2020, Narváez-Gómez et al. 2021a), has allowed us to confidently determine priority localities for improved sampling (Narváez-Gómez et al. 2021a, b). Data of this nature has been especially important in the case of rare and threatened species, allowing one to rapidly spot endangered taxa and establish sound conservation plans (Sousa-Baena et al. 2013). In the next decade, novel research approaches spearheaded by recent developments of machine learning and artificial intelligence approaches (Soltis et al. 2020) will significantly benefit from the vast amounts of data housed at *species*Link and other biodiversity information portals. These data will be especially crucial to expediting taxonomic analyses in the Anthropocene (Grace et al. 2021, Mabry et al. 2022, Gorneau et al. 2022), for high-throughput phenotyping (Gehan & Kellogg 2017), and for the development of new integrative research blending morphology, geology, and ecology, among other data sources.

The *species*Link network also has great potential to contribute meaningfully to the establishment of public policy, sustainable development, and biodiversity conservation, all of which are highly dependent on high-quality biodiversity data. Indeed, it is broadly recognized that many significant changes are required for a more sustainable planet, including ecosystem conservation and reforestation,

more sustainable agricultural practices, changes in the food system, nature-based solutions, reductions in the environmental footprint, and health. Comprehensive data networks such as *species*Link provide crucial data for the global sustainable agenda (e.g., UN Climate Change Conference, COP26, Glasgow, UK), including an improved understanding of the main drivers of global change such as pollution, habitat change, climate change, and species invasions (Lang et al. 2019). Indeed, biological collections are essential for understanding biodiversity in the Anthropocene (Meineke et al. 2018).

The applied uses of biodiversity data repositories and their contributions to society should not be underestimated (e.g., Wen et al. 2015). As outlined in the Shenzhen Declaration on Plant Sciences issued during the International Botanical Congress (2017), scientists must conduct research in the context of a changing world, compile a complete inventory of all plant species, and utilize big data platforms to increase our understanding of nature (Raven 2019). *species*Link provides essential data as we head into a more sustainable future. The potential of herbaria and biodiversity databases for developing new medicines and products is immense, although still underutilized (Souza & Hawkins, 2017). While the current value of large biodiversity databases like *species*Link is immense, their value will greatly increase over time as many species become extinct in nature. As we enter what has been termed the 6th mass extinction, the only known records of several taxa will be those deposited in biological collections and available through biodiversity databases (Raven & Miller, 2020).

As biodiversity documentation and biological studies have entered the "*era of big-data*" (Maldonado et al. 2015), *species*Link is increasingly at the front line of biodiversity research, providing data for accurate documentation of the megadiverse Neotropical biota, as well as for exploration of the relationship between biodiversity data availability and socio-political conditions in time and space (Zizka et al. 2021). Maintaining and funding biodiversity collections, field expeditions, and online repositories such as *species*Link is essential for accurate biodiversity documentation, and crucial to diminish biodiversity shortfalls. It is only through high-quality data and a good understanding of biodiversity patterns that it will be possible to manage, utilize, and preserve biological resources effectively. Indeed, unraveling these distribution patterns and understanding the ecological and historical drivers of species diversity is fundamental for sound public policy, sustainable development, and biodiversity conservation (Narváez-Gómez et al. 2021a, b).

Biodiversity Conservation Policies and e-Infrastructures

Conserving biodiversity and ecosystem services is crucial for our existence. The 2050 vision of the Convention on Biological Diversity (CBD) includes increased efforts to conserve, restore, and safeguard areas that deliver benefits essential to all people. Governments are set to adopt a new set of biodiversity conservation targets to replace the 2020 goals agreed in Aichi, Japan, in 2010. Most of the Aichi targets were missed, despite the agreement of governments to prevent biodiversity decline. The global report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES 2019) warned about the accelerating rates of species extinction and ecosystem degradation.

Achieving nature conservation goals, demands an accurate quantification and mapping of biodiversity and associated ecosystem services at broad spatial scales to help prioritize critical locations for nature conservation and management, especially for evaluating, managing, and establishing protected areas, a central strategy of current global biodiversity conservation initiatives (Mitchell et al. 2021). While the information on ecosystem properties and abiotic environmental conditions are available at many spatial and temporal scales, biodiversity is still often studied locally, many times lacking wide taxonomic breadth, large temporal scale, and spatial coverage (Altermatt et al. 2020, Peterson & Soberón, 2018).

The World Bank report, *The Economic Case for Nature* (Johnson et al. 2021) estimates that the collapse of specific ecosystem services (e.g., pollination, food supply from marine fisheries, and timber from native forests) could result in a decline in the global GDP of US$2.7 T annually by 2030. A drastic reduction in pressure on biodiversity can only occur by systemic changes in the production and consumption of goods and services that impact nature. This depends on the alignment of finance (both private and public) with the combined needs of nature and humans. The institutions that govern global finance should ensure that financial institutions effectively contribute to biodiversity protection. The function of the financial system and its accountability to the ultimate owners of assets and intended beneficiaries, needs good quality, accessible, and standardized information that can be communicated successfully. Recently, during the first part of the Biodiversity Conference of the Parts, COP15 (October 2021), the Chinese government launched the Kunming Biodiversity Fund (US$ 233 Million) for the protection of fauna and flora in developing countries. An effective global information infrastructure for biodiversity – to which *species*Link is a significant contributor – is a crucial element in meeting these challenges.

The Future of e-Infrastructures

The first two decades of the 21st century have seen rapid and massive rises in the quantity, quality, and accessibility of biodiversity data. A worldwide focus on digitization of biological collections and their related environmental, ecological, taxonomic, and supporting data has spurred the development of local, national, regional, and global networks aimed at aggregating, mobilizing, and serving these data to an ever-widening audience (Nelson & Ellis 2018, 2019). The increasing demands and sophistication of this audience are highlighting a plethora of challenges and opportunities that will require novel approaches and innovative solutions for the world's biodiversity data aggregators.

Perhaps most pressing is the design and implementation of the tools and infrastructure that are necessary to effectively integrate and synchronize data across all aggregators and life science domains in ways that bring suites of related data together, allowing for discoveries and unknown relationships (Peterson et al. 2010; Hardisty & Roberts 2013, Peterson, et al. 2015). Collaborative work involving the Global Biodiversity Information Facility (GBIF), Integrated Digitized Biocollections (iDigBio, USA), and the Distributed System of Scientific Collections (DiSSCo, Europe) under the moniker Digital Extended Specimen (DES) has begun. DES represents a convergence of DiSSCo's digital specimen concept (Addink & Hardisty 2020), GBIF's clustering algorithm (GBIF, 2020), and Biological Collections Network's (BCoN, USA) extended specimen network (Lendemer et al. 2020; BCoN, 2019). The unified vision involves increasing collaborations with numerous

additional biodiversity data aggregators, such as *species*Link (CRIA, Brazil), CONABIO (Mexico), and Canadensys (Canada), to name a few.

The intended outcome of this work is an easily accessed network of linked, harmonized, and integrated digital resources to increase the quantity, quality, usability, and consistency of all the earth's collections' holdings. Once implemented, a DES-type network of large and small, rich and poor institutions has the potential to allow data providers and users (e.g. researchers, conservation managers, commercial enterprises) to leverage powerful, high-quality data sets at unprecedented size, scale, and versatility (Heberling et al. 2021), ensuring global synchronization of biodiversity data.

Worldwide engagement and collaboration to create a global network of networks across aggregators necessarily require negotiating political and continental boundaries in ways that embrace a common governance structure while simultaneously preserving varying levels of autonomy for network members. Long-term commitments of sustainable funding at national levels to ensure stable support of in-country financial requirements of national aggregators is essential as well as are treaty-like accords that ensure international collaboration. GBIF is currently the only infrastructure with global scope, a well-developed governance structure, and a revenue model that includes membership contributions from its >100 member countries and organizations representing > 1,800 data-sharing institutions or publishers that have been mostly robust and sustainable. The GBIF network also emphasizes inclusion, service to, and data from the world's underrepresented nations, many of which are biodiversity-rich but resource-poor.

The operations of other important infrastructures, including some of those listed above, are often dependent on private or government-funded projects with limited timeframes and the expectation that project staff will pursue strategies for sustainability beyond existing grants of support. For instance, iDigBio has recently entered its second decade of significant funding from the U.S. National Science Foundation (NSF), one of whose roles is to provide limited-time seed funding toward the establishment of potentially important and long-lasting initiatives, yet one of its largest contributors (VertNet) has been unfunded for several years. Although NSF has made it possible for collection-holding institutions in the United States to generate and mobilize a large array of collection data and has funded iDigBio, there is no anticipation that the agency is positioned in scope, resources, or authority to continue this support in the long term. The European Union (EU) and many of its constituent countries are investing heavily in biodiversity data generation, mobilization, and research infrastructure development through DiSSCo and a range of related and significant projects and cost actions (e.g., BICIKL, ICEDIG_eu, SynthesysEU, MobiliseAction).

A driving force for biodiversity aggregation is data used in the scientific literature, both in basic and applied science. Heberling et al. (2021) studied the use of GBIF mediated data in the scientific literature as part of GBIF's strategic planning. They found that ecological niche modelling and related work under the rubric of species distribution modelling, continues to be the most common use of GBIF-served data, but that specific use cases are moving from basic to more applied questions, such as species' distributional responses to climate change. They also found that the use of shared biodiversity data is growing in all major scientific disciplines – these perhaps unintended uses provide opportunities for biodiversity infrastructures to grow and provide value beyond the biological sciences.

To date, most data shared through biodiversity infrastructures has been publicly available, such as data from museums, government-funded programs, and citizen science initiatives, with little data from the private sector. The latest development of the Equator Principles recognizes this shortfall and encourages developers of large infrastructure and industrial projects to share non-sensitive biodiversity data from environmental impact assessments into the GBIF network. This area represents a large potential source of data, currently locked away, that could contribute significantly to filling data gaps in developing areas. The Data4Nature initiative encourages development actors to share the biodiversity data collected during impact assessments of the projects they support for the global commons.

This article emphasizes collaboration and integration, as well as free and open access to data, tools, and systems, in the context of CRIA's efforts to create an e-infrastructure for Brazil and South America more generally. Large and small collections are important, as are global, regional, national, and specialized e-infrastructures. Leaving all responsibility to GBIF as *the* Global Biodiversity Information Facility is not sufficient. Rather, it is important that all networks indeed work as a network, as each entity has a role to play. Local e-infrastructures are also centers of innovation that focus on local needs and feed into international systems. *species*Link works in close collaboration with data providers and users, facilitating and improving data capture, sharing, and use. It integrates data from and shares data with other e-infrastructures such as iDigBio and GBIF while focusing on its users when developing specific tools and outputs for its data. Historically, CRIA/*species*Link's staff works with collaborators from all Brazilian states and from abroad. We believe that this path optimizes data relevancy in research, education, and policymaking, among many other fields. This path is already helping researchers to understand and mitigate the impacts of climate change and to better manage natural resources locally and regionally. At the same time, *species*Link offers data for global integration and analysis. *species*Link's importance is unquestionable, and yet no financial mechanisms exist that guarantee its permanence (Canhos et al. 2015). Indeed, *species*Link is not alone, as large networks such as iDigBio, also do not have their long-term continuity guaranteed, despite the significant budget that they currently hold. Effective steps must be taken to guarantee not only long-term sustainability and continuity, but also to ensure progress in increasing content, usefulness, and usability to an ever-evolving user community.

Acknowledgments

The authors thank all data providers, responsible for *species*Link's 546 data sets, and users who helped CRIA develop all of its tools and systems. We also thank the many funding agencies and institutions who have supported projects, services, and events of importance to *species*Link´s development:

- in Brazil – Fapesp, FUSP, CNPq, Finep, MCTI, CAPES, CGEE, Fiocruz, JBRJ, FACEPE, FAPES, FUNDESPA, Fundação Vitae, Natura, A.B.E.L.H.A., ABC – Academia Brasileira de Ciências, INPA, TaxOnline, Instituto Estadual de Florestas – IEF, UNESP, Universidade de Passo Fundo, Universidade Federal do Paraná, Secretaria de Estado de Meio Ambiente do Acre, and
- abroad: University of Kansas, GBIF, University of California, JRS Biodiversity Foundation, European Commission, FAO,

International Development Research Center (IDRC) – Canada, Gordon and Betty Moore Foundation, Mellon Foundation, University of Colorado, Leibniz Institut für Meereswissenschaften an der Universität Kiel, Joint Nature Conservation Committee, University of Reading, Rutgers, The State University of New Jersey, World Federation for Culture Collections, Cardiff University, European Research Consortium for Informatics and Mathematics, iHub Limited, US Geological Survey, International Council of Scientific Unions ICSU, CODATA.

As to infrastructure services, fundamental partners are RNP – Rede Nacional de Ensino e Pesquisa and Google Cloud.

Conflicts of Interest

The authors declare that they have no conflict of interest related to the publication of this manuscript.

Ethics

This study did not involve human beings and/or clinical trials that require approval by an Institutional Committee.

References

ADDINK W, HARDISTY A. 2020. 'openDS' – Progress on the New Standard for Digital Specimens. Biodiversity Information Science and Standards 4: e59338.

AGENDA 21 (1992). United Nations Conference on Environment & Development. Rio de Janeiro, Brazil, 3 to 14 June 1992. Agenda 21. United Nations Sustainable Development. https://sustainabledevelopment.un.org/content/documents/Agenda21.pdf. (last access in 14/07/2022).

CONVENTION ON BIOLOGICAL DIVERSITY, UNITED NATIONS, 1992. https://www.cbd.int/doc/legal/cbd-en.pdf

CONSTABLE H, GURALNICK R, WIECZOREK J, SPENCER CL, PETERSON AT (2010) VertNet: A new model for biodiversity data sharing. PLoS Biology 8: 1–4.

BALL-DAMEROW, J.E. L. BRENSKELLE, N. BARVE, P.S. SOLTIS, P. SIERWALD, R. BIELER, R. LAFRANCE, A.H. ARIÑO & R.P. GURALNICK. 2019. Research applications of primary biodiversity databases in the digital age. PLoS One 14(9): e0215794.

BCON. 2019. BCoN Report: Extending U.S. Biodiversity Collections to Promote Research and Education – Biodiversity Collections Network. (24 November 2021; https://bcon.aibs.org/2019/04/04/bcon-report-extending-u-s-biodiversity-collections-to-promote-research-and-education/).

BERNACCI, L.C. & GIOVANNI, R. 2013. *Passiflora edmundoi* (v1) in Biogeografia da Flora e dos Fungos do Brasil. INCT Herbário Virtual (http://biogeo.inct.florabrasil.net/proc/46).

BEZERRA A.D. et al. 2019. Data relating to threats to passion fruit production in the Neotropics due to agricultural area loss and pollinator mismatches consequence of climate changes. *Data in brief*, 103802. https://doi.org/10.1006/jdib.2019.103802

BEZERRA, A.D. PACHECO-FILHO, A.J.S. BONFIM, I.G.A. SMAGGHE, G. FREITAS, B.M. 2019. Agricultural area loss and pollinator mismatch due to the climate changes endanger passion fruit production in the Neotropics. *Agric. Syst*. *169*:49–57.

BFG – THE BRAZILIAN FLORA GROUP. 2021. Brazilian Flora 2020: Leveraging the power of a collaborative scientific network. Taxon 00:1-21. https://doi.org/10.1002/tax.12640.

BORGES, M., PETTI, M.A.V., FUKUDA, M.V., CASSANO, V., FUJII, M.T., AMARAL, A.C.Z. 2022. Marine planktonic and benthic organisms: an ocean of diversity in the collections of the State of São Paulo. Biota Neotrop. 22 (Suppl. 1). In press

BREEZE, TD; GALLAI, N; GARIBALDI, LA; LI, XS. Economic measures of pollination services: shortcomings and future directions. Trends in Ecology and Evolution, 31(12):927–939 (2016).

BUERKI, S. & BAKER, W.L. 2016. Collections-based research in the genomic era. Biological Journal of the Linnean Society 117: 5–10.

CAMPBELL, A.J. GIGANTE CARVALHEIRO, L. GASTAUER, M. et al. 2019. Pollinator restoration in Brazilian ecosystems relies on a small but phylogenetically diverse set of plant families. *Sci Rep 9*, 17383. https://doi.org/10.1038/s41598-019-53829-4

CANHOS DAL, CHAPMAN AD & CANHOS VP (2004). Study on Data-sharing with Countries of Origin. Report number: GBIFS/2003/04. DOI: 10.13140/RG.2.2.22909.10724.

CANHOS VP, MANFIO G (2004). Recursos Microbiológicos para Biotecnologia. iIn Biotecnologia e Recursos Genéticos: Desafios e oportunidades para o Brasil. Orgs: Da Silveira, J.M.F.J, Dal Poz, M.E. e Assad, A.L.D. ppg. 233–250. ISBN -85-87120-05-0

CANHOS, D.A.L, SOUSA-BAENA, M.S, DE SOUZA, S, MAIA, L.C. STEHMANN, J.R, CANHOS, V.P. ET AL. 2015. The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLoS Biology 13(7): e1002204. doi:10.1371/journal.pbio.1002204

CANHOS, D.A.L. SOUSA-BAENA, M.S. SOUZA, S. GARCIA, L.C. DE GIOVANNI, R. MAIA, L.C. & BONACELLI, M.B.M. 2014. Lacunas: a web interface to identify plant knowledge gaps to support informed decision-making. Biodivers Conserv 23, 109–131 (2014). https://doi.org/10.1007/s10531-013-0587-0

CANHOS, D.A.L. SOUZA, S. CANHOS, V. MAIA, L.C. 2019. Brazil's Virtual Herbarium, an Infrastructure for Open Science. In: Contextualizing Openness: Situating Open Science. CHAN, L. et al. (Ed.). Ottawa: University of Ottawa Press, v.1, p. 133–145.

CHAPMAN AD (2005a). *Principles of Data Quality*. Report for the Global Biodiversity Information Facility 2005. 61pp. Copenhagen: GBIF. https://doi.org/10.15468/doc.jrgg-a190.

CHAPMAN AD (2005b). *Principles and Methods of Data Cleaning Primary Species Occurrence Data.* Report for the Global Biodiversity Information Facility 2005. 75pp. Copenhagen: GBIF. http://www.gbif.org/document/80528.

CHAPMAN AD (2017). History of Biodiversity Informatics Part 1. *BITC Global Online Seminar* #20. https://www.youtube.com/watch?v=w5_y6o9GVVs.

COLLI-SILVA M, PIRANI JR (2020) Estimating bioregions and under collected areas in South America by revisiting Byttnerioideae, Helicteroideae, and Sterculioideae (Malvaceae) occurrence data. Flora 271: 151688.

COLLI-SILVA M, REGINATO M, CABRAL A, FORZZA RC, PIRANI JR,VASCONCELOS TNC 2019. Evaluating shortfalls and spatial accuracy of biodiversity documentation in the Atlantic Forest, the most diverse and threatened Brazilian phytogeographic domain. *Taxon* 69: 567–577.

COSTA, W. F. RIBEIRO, M. SARAIVA, A. M. IMPERATRIZ-FONSECA, V. L. and GIANNINI, T.C. 2018. Bat diversity in Carajás National Forest (Eastern Amazon) and potential impacts on ecosystem services under climate change. *Biol. Conserv*. *218*, 200–210. doi: 10.1016/j.biocon.2017.12.034.

COSTA-LOTUFO, L.V.; COLEPICOLO, P.; PUPO, M.T.; PALMA, M.S. Bioprospecting Macroalgae, Marine and Terrestrial Invertebrates & Their Associated Microbiota. *Biota Neotrop*. 22 (Suppl. 1). In press.

CRIA, RELATÓRIO 2020. Centro de Referência em Informação Ambiental. https://cria.org.br/docs/relatorio-anual-2020.pdf (accessed in Feb. 25, 2022).

DARU, B.H. PARK, D.S. PRIMACK, R.B. WILLIS, C.G. BARRINGTON, D.S. WHITFELD, T.J.S. SEIDLER, T.G. SWEENEY, P.W. FOSTER, D.R. ELLISON, A.M. & DAVIS, C.C. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytologist 217: 939–955. DOI: 10.1111/nph.14855.

EDWARDS, J.L., LANE, M.A., NIELSEN, E.S. 2000. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. Science, v. 289 n. 5488, p. 2312–2314. DOI: 10.1126/science.289.5488.231.

FORZZA, R.C. BAUMGRATZ, J.F.A. BICUDO, C.E.M. CARVALHO JR. A.A. COSTA, A. COSTA, D.P. HOPKINS, M. LEITMAN, P.M. LOHMANN, L.G. MAIA, L.C. MARTINELLI, G. MENEZES. M. MORIM, M.P. COELHO, M.A.N. PEIXOTO, A.L. PIRANI, J.R. PRADO. J. QUEIROZ, L.P. SOUZA, V.C. STEHMANN, J.R. SYLVESTRE, L.S. WALTER, B.M.T. & ZAPPI, D.C. (eds.) 2010a. Catálogo de plantas e fungos do Brasil. vol. 1; vol. 2. Rio de Janeiro: Andrea Jakobsson Estúdio Editorial, Jardim Botânico do Rio de Janeiro. 870p. 830p.

FORZZA, R. C. BAUMGRATZ, J. F. A. BICUDO, C. E. CANHOS, D. A. L. CARVALHO JR, A. A. COSTA, A. ... & ZAPPI, D. (2010b). Síntese da diversidade brasileira. Catálogo de plantas e fungos do Brasil, 1, 21–42. https://books.scielo.org/id/z3529/pdf/forzza-9788560035083-04.pdf

GBIF. 2020. New data-clustering feature aims to improve data quality and reveal cross-dataset connections. News, 2020-07-28. (16 November 2021; https://www.gbif.org/news/4U1dz8LygQvqIywiRIRpAU/new-data-clustering-feature-aims-to-improve-data-quality-and-reveal-cross-dataset-connections).

GEHAN, M.A. & E.A. KELLOGG. 2017. High-throughput phenotyping. American Journal of Botany 9: 553.

GNAS GERMAN NATIONAL ACADEMY OF SCIENCES LEOPOLDINA, 2014: Challenges and Opportunities of Integrative Taxonomy for Research and Society – Taxonomic Research in the Era of OMICS Technologies. Deutsche Akademie der Naturforscher Leopoldina e.V. Nationale Akademie der Wissenschaften (German National Academy of Sciences). Halle/Saale, 54 pages. ISBN: 978-3-8047-3291-9. https://www.leopoldina.org/uploads/tx_leopublication/2014_Stellungnahme_Taxonomie_EN_final_01.pdf

GNAS GERMAN NATIONAL ACADEMY OF SCIENCES LEOPOLDINA, 2021: Maintaining open access to Digital Sequence Information – multilateral benefit sharing and Open Science. Halle (Saale). ISBN: 978-3-8047-4220-8. https://www.leopoldina.org/uploads/tx_leopublication/2021_Digitale_Sequenzinformationen_EN_02.pdf

GHAZOUL, J. 2005. Buzziness as usual? Questioning the global pollination crisis. Trends Ecol Evol 20(7):367–373. https://doi.org/10.1016/j.tree.2005.04.026

GIANNINI T.C. ACOSTA A.L. COSTA W.F. MIRANDA L. PINTO C.E. WATANABE M.T.C. ZAPPI D.C. GIULIETTI A.M. AND IMPERATRIZ-FONSECA V.L. 2021. Flora of Ferruginous Outcrops Under Climate Change: A Study in the Cangas of Carajás (Eastern Amazon). *Front. Plant Sci. 12*:699034. doi: 10.3389/fpls.2021.699034

GIANNINI T.C. COSTA W.F. CORDEIRO G.D. IMPERATRIZ-FONSECA V.L. SARAIVA A.M. BIESMEIJER J. *et al*. 2017a. Projected climate change threatens pollinators and crop production in Brazil. *PLoS ONE* 12(8): e0182274. https://doi.org/10.1371/journal.pone.018227

GIANNINI, T.C. et al. 2012. Pollination services at risk: bee habitats will decrease owing to climate change in Brazil. *Ecological Modelling 244*, 127–131.

GIANNINI, T.C. *et al*. 2013. Identify the areas to preserve passion fruit pollination service in Brazilian Tropical Savannas under climate change. *Agriculture, Ecosystems and Environment 171*: 39–46.

GIANNINI, T.C. COSTA, W.F. BORGES, R.C. MIRANDA, L. COSTA, C.P.W. SARAIVA, A.M. *et al*. 2020. Climate change in the Eastern Amazon: crop-pollinator and occurrence-restricted bees are potentially more affected. *Reg. Environ. Change 20*: 9.

GIANNINI, T.C. MAIA-SILVA, C. ACOSTA, A.L. et al. 2017b. Protecting a managed bee pollinator against climate change: strategies for an area with extreme climatic conditions and socioeconomic vulnerability. *Apidologie 48*: 784–794. https://doi.org/10.1007/s13592-017-0523-5

GIOVANNI, R. BERNACCI, L.C. 2015. Progressively approaching the distribution of *Passiflora ischnoclada* (Passifloraceae) from a single occurrence record. *Check List* v.11, 1717–1722. doi: 10.15560/11.4.1717

GIOVANNI, R. BERNACCI, L.C. SIQUEIRA, M.F. ROCHA, F.S. 2012. The Real Task of Selecting Records for Ecological Niche Modelling. *Natureza & Conservação* v.10, 139–144. doi: 10.4322/natcon.2012.018

GOMES, V.H.F. VIEIRA, I.C.G. SALOMÃO, R.P. et al. 2019. Amazonian tree species threatened by deforestation and climate change. *Nat. Clim. Change. 9*, 547–553 https://doi.org/10.1038/s41558-019-0500-2

GONZALEZ, V.H.; COBOS, M.E.; JAREMILLO, J. OSPINA, R. 2021. Climate change will reduce the potential distribution ranges of Colombia's most valuable pollinators. *Perspectives in Ecology and Conservation 19* (2):195–206.

GORNEAU, J. A., AUSICH, W. I., BERTOLINO, S., BIK, H., DALY, M., DEMISSEW, S., DONOSO, D. A., FOLK, R., FREIRE-FIERRO, A., GHAZANFAR, S. A., GRACE, O. M., AI-QUN HU, KULKARNI, S., LICHTER-MARCK, I. H., LOHMANN, L. G., MALUMBRES-OLARTE, J., MUASYA, A. M., ABEL P, SINGH, Y., SINISCALCHI, C. M., SPECHT, C. D., STIGALL, A. L., TANK, D. C., WALKER, L. A., WRIGHT, D. F., ZAMANI, A., ESPOSITO, L. A. 2022. Framing the future for taxonomic monography: Improving recognition, support, and access. Bulletin of the Society of Systematic Biologists. Vol. 1 No. 1 (2022). https://doi.org/10.18061/bssb.v1i1.8328

GRACE, O.M. O.A. PÉRES-ESCOBAR, E.J LUCAS, M. VORONTSOVA, G.P. LEWIS, B.E. WALKER, L.G. LOHMANN, S. KNAPP, P. WILKIE, T. SARKINEN, I. DARBYSHIRE, E. NIC LUGHADHA, A. MONRO, Y. WOUDSTRA, S. DEMISSEW, A.M. MUYASYA, S. DIAZ, W.J. BAKER & A. ANTONELLI. 2021. Botanical monography in the Anthropocene. Trends in Plant Sciences 26(5): 433–441.

HARDISTY, A. AND D. ROBERTS. 2013. A decadal view of biodiversity informatics: Challenges and priorities. BMC ecology 13:16.

HEBERLING JM, MILLER JT, NOESGAARD D, WEINGART SB, SCHIGEL D. Data integration enables global biodiversity synthesis. Proc Natl Acad Sci U S A. 2021 Feb 9;118(6):e2018093118. doi: 10.1073/pnas.2018093118. PMID: 33526679; PMCID: PMC8017944.

HEDRICK, B.P. HEBERLING, J.M. MEINEKE, E.K. TURNER, K.G. GRASSA, C.J. PARK, D.S. KENNEDY, J. CLARKE, J.A. COOK, J.A. BLACKBURN, D.C. EDWARDS, S.V. & DAVIS, C.C. 2020. Digitization and the future of natural history collections. BioScience 70(3):243–251. https://doi.org/10.1093/biosci/biz163

HOBERN, D. et al. 2013. Global Biodiversity Informatics Outlook: Delivering biodiversity knowledge in the information age. Global Biodiversity Information Facility (GBIF Secretariat). Book. 41 pages. ISBN: 8792020526. http://www.biodiversityinformatics.org.

HOLMES, M.W. HAMMOND, T.T. WOGAN, G.O.U. WALSH, R.E. LABARBERA, K. WOMMACK, E.A. MARTINS, F.M. CRAWFORD, J.C. MACK, K.L. BLOCH, L.M. & NACHMAN, M.W. 2016. Natural history collections as windows on evolutionary processes. Mol Ecol 25(4):864–881. https://doi.org/10.1111/mec.13529

IPBES (2019): Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages. https://doi.org/10.5281/zenodo.3831673

JOHNSON, JUSTIN ANDREW; RUTA, GIOVANNI; BALDOS, URIS; CERVIGNI, RAFFAELLO; CHONABAYASHI, SHUN; CORONG, ERWIN; GAVRYLIUK, OLGA; GERBER, JAMES; HERTEL, THOMAS; NOOTENBOOM, CHRISTOPHER; POLASKY, STEPHEN. 2021. The Economic Case for Nature: A Global Earth-Economy Model to Assess Development Policy Pathways. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/35882 License: CC BY 3.0 IGO.

JOLY, CA, SPEGLICH, E. 2003. Programa BIOTA/FAPESP: Um novo paradigma no estudo da conservação e do uso sustentável da biodiversidade. Cienc. Cult. vol.55 no.3 São Paulo July/Sept. 2003. ISSN 2317-6660.

JORQUERA MA, GRAETHER SP, MARUYAMA F (2019). Editorial: bioprospecting and biotechnology of extremophiles. Frontiers in Bioengineering and Biotechnology, 7: 204.

KATE, K.T. & LAIRD, S.A. (1999). The Commercial use of Biodiversity: Access to Genetic Resources and Benefit-Sharing (1st ed.). Routledge. https://doi.org/10.4324/9780429341540

KLEIN, A.-M. VAISSIÈRE, B.E. CANE, J.H. STEFFAN-DEWENTER, I. CUNNINGHAM, S.A. KREMEN, C. & TSCHARNTKE, T. 2007. Importance of pollinators in changing landscapes for world crops. P Roy Soc B-Biol Sci 274(1608):303–313. https://doi.org/10.1098/rspb.2006.3721

KLEIN, S. CABIROL, A. DEVAUD, J.-M. BARRON, A.B. & LIHOREAU, M. 2017. Why bees are so vulnerable to environmental stressors. Trends Ecol Evol 32(4):268–278. https://doi.org/10.1016/j.tree.2016.12.009

KRECHEMER, F.; MARCHIORO, C.A. 2020. Past, present, and future distributions of bumble bees in South America: identifying priority species and areas for conservation. *J. Appl. Ecol 57*: 1829–1839.

LANG, P.L.M. F.M.WILLEMS, J.F. SCHEEPENS, H.A. BURBANO, O. BOSSDORF. 2019. Using herbaria to study global environmental change. New Phytologist 221: 110–122.

LENDEMER J, THIERS B, MONFILS AK, ZASPEL J, ELLWOOD ER, BENTLEY A, LEVAN K, BATES J, JENNINGS D, CONTRERAS D, LAGOMARSINO L, MABEE P, FORD LS, GURALNICK R, GROPP RE, REVELEZ M, COBB N, SELTMANN K, AIME MC. 2020. The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. BioScience 70: 23–30.

LIMA, V.P. MARCHIORO, C.A. 2021. Brazilian stingless bees are threatened by habitat conversion and climate change. *Reg Environ Change 21*, 14.

LO GIUDICE, A. & GUGLIANDOLO C. 2019. A special issue on microorganisms from extreme environments in Memory of Luigi Michaud (1974–2014). Diversity, 12: 2.

MABRY, M.E. F. ZAPATA, D.L. PAUL, P.M. O'CONNOR, P.S. SOLTIS, D.C. BLACKBURN & N.B. SIMMONS. 2022. Monographs as a nexus for building extended specimens networks using persistent identifiers. Bulletin of the Society of Systematic Biologists http://dx.doi.org/10.18061/bssb.v1i1.8323

MAIA, L.C. VIEIRA, A.O.S. PEIXOTO, A.L. CANHOS, D.A.L. STEHMANN, J.R. BARBOSA, M.R.V. & MENEZES, M. 2017. Construindo redes para promover o conhecimento da biodiversidade: a experiência do INCT-Herbário Virtual da Flora e dos Fungos – Building networks to promote knowledge of Brazil's Biodiversity: The experience of the INCT-Virtual Herbarium of Flora and Fungi. Recife: Livro Editora Universitária da UFPE. ISBN: 978-85-415-0901-5, Páginas: 111.

MAIA, L.C.; VIEIRA, A.O.S. CANHOS, D.A.L. STEHMANN, J.R. BARBOSA, M.R.V. MENEZES, M. PEIXOTO, A.L. (2015). Programa Reflora associado ao Instituto Nacional de Ciência e Tecnologia Herbário Virtual da Flora e dos Fungos: ampliação, integração e disseminação digital de dados repatriados da flora brasileira. Unisanta BioScience 4 (7): 95–111. https://periodicos.unisanta.br/index.php/bio/article/view/613

MAIA, U.M. et al. 2020. Climate induced distributions dynamics of *Plebeia flavocincta,* a stingless bee from Brazil tropical dry forests. *Ecology and Evolution*, *10*: 10130–10138.

MALDONADO, C. MOLINA, C.I. ZIZKA, A. PERSSON, C. TAYLOR, C.M. ALBÁN, J. CHILQUILLO, E. RØNSTED, N. & ANTONELLI, A. 2015. Estimating species diversity and distribution in the era of Big Data: To what extent we can trust public databases? Global Ecology Biogeography 24: 973–984.

MAMEDE, M.C.H. & SIMÃO-BIANCHIN, R. 2022. The Herbaria of the State of São Paulo. Biota Neotrop. 22 (Suppl. 1). In press.

MEINEKE, E.K. T.J. DAVIES, B.H. DARU & C.C. DAVIS. 2018. Biological collections for understanding biodiversity in the Anthropocene. Phil. Trans. R. Soc. B. 374: 20170386.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA – MCT (2002). Sistema de. Avaliação da Conformidade de Material Biológico. Brasília, SENAI/DN, 2002. 102 p. ISBN 85-7519-077-6. https://www.fcav.unesp.br/Home/servicos/CREBIO1852/sistema-de-avaliacao-da-conformidade-de-material-biologico.pdf

MIRANDA, L.S. IMPERATRIZ-FONSECA, V.L. AND GIANNINI, T.C. 2019. Climate change impact on ecosystem functions provided by birds in southeastern Amazonia. *PLoS One 14* :e0215229. doi: 10.1371/journal.pone.021522

MIRANDA. L.S.; AWADE, M. *et al*. 2021. Combining connectivity and species distribution modeling to define conservation and restoration priorities for multiple species: a case study in Eastern Amazon. *Biol. Cons. 257*: 109148.

MONFILS, A.K. KRIMMEL, E.R. BATES, J.M. BAUER, J.E. BELITZ, M.W. CAHILL, B.C. CAYWOOD, A.M. *et al*. 2020. Regional Collections Are an Essential Component of Biodiversity Research Infrastructure. BioScience 70 (12): 1045–47. https://doi.org/10.1093/biosci/biaa102.

MOURE, J.S. URBAN, D. & MELO, G.A.R. 2007. Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region.

MOURE, J.S. URBAN, D. & MELO, G.A.R. 2008. Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region – online version. Available from: http://www.moure.cria.org.br/catalogue

MOURE, J.S. URBAN, D. & MELO, G.A.R. 2012. Catalogue of Bees (Hymenoptera, Apoidea) in the Neotropical Region – online version. Available from: http://www.moure.cria.org.br/catalogue

MUÑOZ, M.E.S., DE GIOVANNI, R., DE SIQUEIRA, M.F. et al. openModeller: a generic approach to species' potential distribution modelling. Geoinformatica 15, 111–135 (2011). https://doi.org/10.1007/s10707-009-0090-7.

NARVÁEZ-GÓMEZ, J.P., GUEDES, T.B. & LOHMANN, L.G. Recovering the drivers of sampling bias in Bignonieae (Bignoniaceae) and identifying priority areas for new survey efforts. Biodivers Conserv 30, 2319–2339 (2021). https://doi.org/10.1007/s10531-021-02195-7.

NARVÁEZ-GÓMEZ JP, SZUMIK CA, GOLOBOFF PA, LOHMANN LG. 2021b. Unravelling distribution patterns of Neotropical lianas: an analysis of endemicity of tribe Bignonieae (Bignoniaceae), Botanical Journal of the Linnean Society, Volume 199, Issue 1, May 2022, Pages 470–495, https://doi.org/10.1093/botlinnean/boab066

NELSON, G. & ELLIS, S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B 374: 20170391. http://dx.doi.org/10.1098/rstb.2017.0391

NELSON, G. & ELLIS, S. 2019. The history and impact of digitization and digital data mobilization on biodiversity research. Phil T Roy Soc B 374(1763):20170391. https://doi.org/10.1098/rstb.2017.0391

OECD (2001), Biological Resource Centres: Underpinning the Future of Life Sciences and Biotechnology, OECD Publishing, Paris, https://doi.org/10.1787/9789264193550-en

OLIVEIRA, U. PAGLIA, A.P. BRESCOVIT, A.D. DE CARVALHO, C.J.B. SILVA, D.P. REZENDE, D.T. LEITE, F.S.F. BATISTA, J.A.N. BARBOSA, J.P.P.P. STEHMANN, J.R. ASCHER, J.S. DE VASCONCELOS, M.F. DE MARCO, P. JR, LÖWENBERG-NETO, P. DIAS, P.G. FERRO, V.G. & SANTOS, A.J. 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Diversity and Distributions 22: 1232–1244. https://doi.org/10.1111/ddi.12489.

PEIXOTO, AL, BARBOSA, MRV, MENEZES, M, MAIA, LC, VAZOLER, RF, MARINONI, L, CANHOS, DAL (2006). Strategies for the modernization of Brazilian biological collections and the consolidation of integrated information systems on biodiversity. Brasília, DF. Centro de Gestão e Estudos Estratégicos. ISBN CDU 574:32:5/6(81).

PETERSON AT et al. 2011. Ecological Niches and Geographical Distributions. Princeton: Princeton University Press.

PETERSON, A.T., SOBERÓN, J. Essential biodiversity variables are not global. Biodivers Conserv 27, 1277–1288 (2018). https://doi.org/10.1007/s10531-017-1479-5

PETERSON, A.T.J. SOBERÓN, AND L. KRISHTALKA. 2015. A global perspective on decadal challenges and priorities in biodiversity informatics. BMC ecology 15:1–9.

PETERSON, A.T.S. KNAPP, R. GURALNICK, JJ. SOBERÓN, AND M.T. HOLDER. 2010. The big questions for biodiversity informatics. Systematics and Biodiversity 8:159–168.

PORTO, RG; CRUZ-NETO, O; TABARELLI, M; VIANA, BF; PERES, CA; LOPES. AV. Pollinator-dependent crops in Brazil yield nearly half of nutrients for humans and livestock feed. Global Food Security, 31, 100587 (2021) doi.org/10.1016/j.gfs.2021.100587

POTTS, S.G. BIESMEIJER, J.C. KREMEN, C. NEUMANN, P. SCHWEIGER, O. & KUNIN, W.E. 2010. Global pollinator declines: trends, impacts and drivers. Trends Ecol Evol 25(6):345–353. https://doi.org/10.1016/j.tree.2010.01.007

POTTS, S.G. IMPERATRIZ-FONSECA, V. NGO, H.T. AIZEN, M.A. BIESMEIJER, J.C. BREEZE, T.D. DICKS, L.V. GARIBALDI, L.A. HILL, R. SETTELE, J. & VANBERGEN, A.J. 2016. Safeguarding pollinators and their values to human well-being. Nature 540(7632):220–229. https://doi.org/10.1038/nature20588

RAVEN, P.H. & S.E. MILLER. 2020. Here today, gone tomorrow. Science 370: 149.

RAVEN, P.H. 2019. Saving plants, saving ourselves. Plants People Planet 1: 8–13.

SABATINO, M.; ROVERE, A. & MELI, P. 2021. Restoring pollination is not only about pollinators: combining ecological and practical information to identify priority plant species for restoration in Pampa grasslands of Argentina. *J. Nat. Cons*. *61*: 126002.

SOBERÓN, J. & A. T. PETERSON. 2009. Monitoring biodiversity loss with primary species-occurrence data: Toward national-level indicators for the 2010 Target of the Convention on Biological Diversity. AMBIO 38:29–34.

SOLTIS, P.S.G. NELSON, A. ZARE & E.M. MEINEKE. 2020. Plants meet machines: Prospects in machine learning for plant biology. Applications in Plant Sciences 8(6): e11371.

SOUSA-BAENA, M. S. L. C. GARCIA, and A. T. PETERSON. 2013. Completeness of Digital Accessible Knowledge of the plants of Brazil and priorities for survey and inventory. Diversity and Distributions 20:369–381.

SOUZA, E.N.F. & J.A. HAWKINS. 2017. Comparison of Herbarium label data and published medicinal use: Herbaria as an underutilized sources of ethnobotanical information. Economic Botany 71: 1–12.

THIERS, B. 2021. The World's Herbaria 2021: A Summary Report Based on Data from Index Herbariorum. Issue 6.0, published February 2022. http://sweetgum.nybg.org/science/wp-content/uploads/2022/02/The_Worlds_Herbaria_Jan_2022.pdf. (last access in 13/07/2022)

THIERS, B. 2022. NYBG's experience in sharing data with speciesLink. Contribution to this paper.

UNITED NATIONS CONFERENCE ON ENVIRONMENT AND DEVELOPMENT, RIO DE JANEIRO, BRAZIL, 3–14 JUNE 1992. United Nations Conferences. Environment and sustainable development. Rio 1992. https://www.un.org/en/conferences/environment/rio1992 (last access in 25/02/2022)

WEN, J. ICKERT-BOND, S.M. APPELHANS, M.S. DORR, L.J. & FUNK, V.A. 2015. Collections-based systematics: Opportunities and outlook for 2050. Journal of Systematics and Evolution 53: 477–488.

WIECZOREK J, BLOOM D, GURALNICK R, BLUM S, DÖRING M, GIOVANNI R, ROBERTSON T & VIEGLAIS D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715

YEATES, D.K. SEAGO, A. NELSON, L. CAMERON, S.L. JOSEPH, L. & TRUEMAN, J.W.H. 2011. Integrative taxonomy, or iterative taxonomy? Syst Entomol 36(2):209–217. https://doi.org/10.1111/j.1365-3113.2010.00558.x

ZIZKA, A., RYDÉN, O., EDLER, D., KLEIN, J., PERRIGO, A., SILVESTRO, D., JAGERS, S. C., LINDBERG, S. I., & ANTONELLI, A. (2021). Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. Journal of Biogeography, 48, 2715– 2726. https://doi.org/10.1111/jbi.14256

ZORZETTO, R. (2002). Rede da vida. Software permite integração de bancos de dados com informações sobre plantas, animais e microrganismos. Revista Fapesp 82: 24-27. https://revistapesquisa.fapesp.br/wp-content/uploads/2002/12/24_megaci%C3%AAncia.pdf

ZWIENER, V.P. LIRA-NORIEGA, A. GRADY, C.J. PADIAL, A.A. and VITULE, J.R.S. 2018. Climate change as a driver of biotic homogenization of woody plants in the Atlantic Forest. *Glob. Ecol. Biogeogr*. *27*, 298–309. doi: 10.1111/geb.12695.

Received: 16/07/2022
Accepted: 20/07/2022
Published online: 15/08/2022