






Challenges and lessons learned from digitizing small Brazilian herbaria

André Luís de Gasper^{1*} , Gustavo Heiden² , Leonardo M. Versieux³ , Paula M. Leitman⁴ 
and Rafaela C. Forzza⁴ 

Received: May 27, 2020
Accepted: March 12, 2021

ABSTRACT

The digital era provides new opportunities for taxonomists, as well as for everyone that studies biodiversity. Many herbaria have been able to digitize their collections, a process that started with the typing of label data, moving more recently towards the digitization of each sample with the simultaneous acquisition of high-resolution images. Here we discuss some of the challenges we faced in digitizing samples and provide a series of suggestions to avoid common mistakes for herbaria that have yet to start the process. We used a professional camera, database management software, and a barcode scanner to digitize the collections of herbaria CRI, ECT, FURB, LUSC, and UFRN. Pre-revision of samples with prior restoration when needed, barcode fixation, and a good database allowed faster digitization of samples. Good database software and the formation of a network among small herbaria accelerated digitization and increased the number of images available of Brazilian biodiversity. Thus far, our joint efforts made 118,000 specimen images available online with the purpose of accelerating botanical research.

Keywords: biodiversity data, biological collection, e-taxonomy, exsiccate, plant collection

Introduction

The advent of the computer ushered in a new era for the biological sciences, with new biodiversity analyses employing algorithms, especially in ecology and conservation research (Hampton *et al.* 2013; Nualart *et al.* 2017). The scientific

community understood that digitizing and making public biodiversity data, especially from herbaria and zoological museums, would be essential to achieving a global analysis of the distribution of biodiversity (Guralnick *et al.* 2007). The first step was to publish collection data on the web, with the last two decades seeing specimens begin to be photographed and their images made available online

¹ Herbário Dr. Roberto Miguel Klein, Universidade Regional de Blumenau, 89030-903, Blumenau, SC, Brazil

² Embrapa Clima Temperado, 96010-971, Pelotas, RS, Brazil

³ Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande do Norte, 59078-970, Natal, RN, Brazil

⁴ Jardim Botânico do Rio de Janeiro, 22460-030, Rio de Janeiro, RJ, Brazil

* Corresponding author: algasper@furb.br



(e.g., JSTOR 2020). This was done for specific collections (see Saint-Hilaire herbarium, Pignal *et al.* 2013), libraries (i.e., Biodiversity Heritage Library) and animal and plant collections. Expectations then turned out to be true: new species were discovered, and novel research was being performed based on the examination of such digitized specimens, despite the discussion of the issue of quality (Smith *et al.* 2011; Costello *et al.* 2013; Soltis 2017). These new data demanded new technologies, such as specific software to aid species image analysis (i.e., RECOLNAT 2020) and morphological research (Borges *et al.* 2020), and even machine learning has been employed (Pryer *et al.* 2020).

Brazil started to discuss guidelines and strategies for biological collections in the early 2000s (Peixoto *et al.* 2006) and an initial approach aimed to raise funds for digitizing and databasing collections (see below). This strategy, combined with the Global Strategy for Plant Conservation (GSPC), gave visibility to Brazilian herbaria and the research conduct therein, and increased the visibility of their works, especially those that were digital (Forzza *et al.* 2017a; Maia *et al.* 2017).

The Reflora Project (Reflora Programme 2020) was established to repatriate high-resolution images of Brazilian specimens deposited in European and North American herbaria and to digitalize Brazilian collections (Forzza *et al.* 2015; Canteiro *et al.* 2019). This helped to compile information for an online Flora of Brazil, meeting Target 1 of GSPC for 2020 (Brazilian Flora 2020). The Reflora Project established a network of Brazilian herbaria and established conditions to: (1) share knowledge on the processes of biodiversity databasing and digitization; (2) train staff and students on procedures and best practices for photography; and (3) equip collections with professional cameras, lighting equipment, and management software (Forzza *et al.* 2017b). As a result of this network, a digitization manual was published (Taniguchi *et al.* 2018).

Images of type specimens should usually be obtained at 600 DPI resolution using flatbed scanners and saved in TIFF format (JSTOR 2018). However, digital cameras are widely used to acquire 300 DPI images for non-type specimens (Harris & Marsico 2017; Takano *et al.* 2019; Tulig *et al.* 2012). Digital cameras can provide a metadata file (raw), which is usually converted to JPEG or TIFF format, both (original and converted files) of which should be stored. JPEG files are made available online because the format was adopted as a web standard and because of file size, whereas TIFF files are used for internal purposes such as original metadata files (Downton 2005). Storage of these files can be problematic due to their size, since a raw file is typically about 80 MB (Silva *et al.* 2017), whereas a converted JPEG, which can be used to identify most diagnostic morphological features, is about 10 MB (Borsch *et al.* 2020). Biological collections initially began to store files on CD-ROMs, and then on internal servers (Thiers *et al.* 2016) but image data hosting services and system integration for the remote use

of image data is needed (Tegelberg *et al.* 2014), such as cloud or e-infrastructure (Nieva de la Hidalga *et al.* 2019)

Conceiving an herbarium database is a fundamental step to entering the digital world. It may appear easy at first, thinking that all that is needed is a spreadsheet and some patience (or many workers in large collections) to type-in all the information from voucher labels. Although many Brazilian collections, especially the smaller ones, still do this (Gasper *et al.* 2020; Brazilian Herbaria Network 2020), it is neither the easiest nor the most recommended method. A spreadsheet cannot be considered a secure database since a simple push of the delete button, a copy-and-paste, or an autofill in the wrong direction can erase or corrupt information contained in a cell. Many free and paid softwares are now available to avoid such downfalls, such as Brahms, Specify, SysTax, HERBAR, and JABOT (Silva *et al.* 2017). Another issue is when data structure does not have a pattern, which makes integrating databases difficult, although it is fundamental in this new era (Wieczorek *et al.* 2012). Thus, automatic and semi-automatic quality control of data become useful (Borsch *et al.* 2020). Global standards were created to solve this interoperability issue, such as the Darwin Core (Wieczorek *et al.* 2012), which defines common terms in datasets, thus improving data integration and sharing.

Although datasets hold a great deal of information, quantity is not often related to quality (Chapman 2005). Unfortunately, collection datasets usually have a lot of erroneous information and many problems, especially if the software used for digitizing does identify inconsistencies. Such issues include, for example, misspelled scientific names, coordinates indicating a different locality, and even the lack of reliable identification (Soberón & Peterson 2004). Smaller herbaria experience more difficulties, since access to technical support for databases, collection management and/or, worst of all, good staff of taxonomists, may not be available. One way to partially solve these problems is to publish specimen images online. Many large and small collections now make images of their specimens available, which has changed the scenario. Distinct repositories are also increasing the tools available to consult these images, such as taking measurements and zooming-in with high capacity to observe details in high-quality images.

Small and medium-sized herbaria may house special datasets, such as new species, new occurrence records, and new information about biodiversity. Making these data available to the scientific community is paramount (Colombo *et al.* 2016). According to the Brazilian Herbaria Network, Brazil has 222 (on June 10, 2019, and 175 on December 2021) active herbaria (180 in Index Herbariorum; Thiers 2020). Of these, 49% have less than 10,000 specimens; 32% have between 10,000 and 50,000 specimens; 10% have between 50,000 and 100,000 specimens, and only 9% have more than 100,000 specimens. Together, herbaria with less than 100,000 specimens hold almost 1,000,000 records (Dillen *et al.* 2019) (Figs. 1, 2).



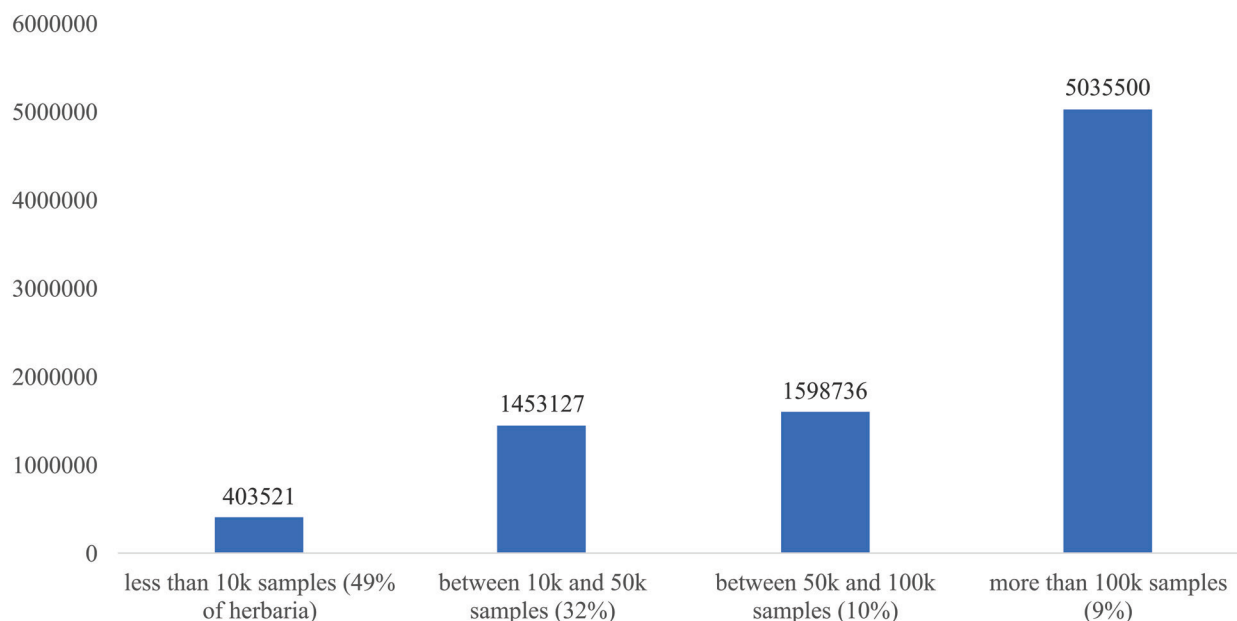


Figure 1. Total number of samples per herbarium size category in Brazil.

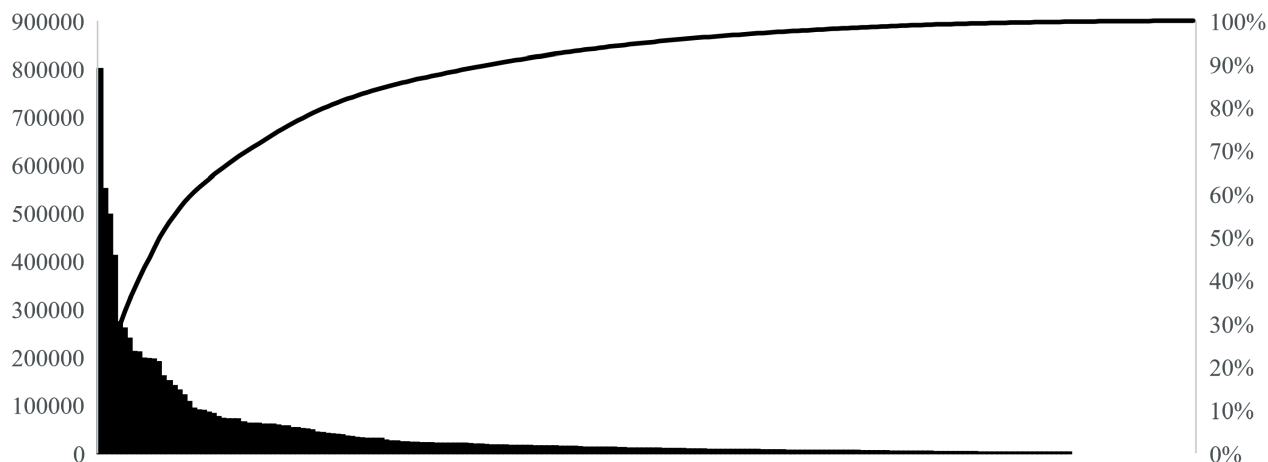


Figure 2. Number of specimens housed in all 221 Brazilian herbaria.

It is not hard to imagine that small collections may experience more difficulties than those that are well-established. Small Brazilian herbaria are usually located far from big cities or state capitals and most are administered by a very limited staff (Gasper *et al.* 2020). On the other hand, many smaller collections are closer to under-collected areas that have not been the focus of historical and intensive botanical activities.

A program of federal public investment in the growth of new university, created many new collections over the last two decades (105), of which 88 (about 49% of Brazilian herbaria) currently have less than 10,000 specimens. It is important to highlight that, in many cases, these “new” herbaria are actually new indexings, and not a new collection (Thiers 2019, continuous adapted). On one hand, such collections may be closer to well-preserved forest fragments

or other natural areas, resulting in collections that are very locally representative of a given area and possessing rare species/specimens that are essential to developing a well-documented Flora (see Thomas *et al.* 2012). On the other hand, such small herbaria may depend only on one or two botanists (and in many herbaria the curator is not a taxonomist), without any collection technicians, and as staff members retire such collections may become neglected (Gasper *et al.* 2020).

Other relevant issues include the lack of institutional support (or respect for small herbaria) and awareness about research conducted at small herbaria. Such collections are usually hosted within a university department and lack internal and bureaucratic visibility within them (Gasper *et al.* 2020).



Challenges for the beginning

First, it is important to understand that the needs and resources of large herbaria differ from those of small herbaria, and there are multiple effective approaches for digitizing these collections (Barkworth & Murrell 2012). For all herbaria, but especially for the smaller ones, the biggest issue is how to capture specimen data fast enough (Smith & Blagoderov 2012), but without great cost, to digitize entire collections (Vollmar *et al.* 2010). This phase is not only limited by equipment (access to cameras, computers, etc.), but also by trained staff with time to perform the work, which needs to be considered. So, what are the possibilities for small collections to make high-quality images with a small budget, perhaps even without access to modern digitalization equipment (see how Muséum National d'Histoire Naturelle proceed; Le Bras *et al.* 2017), to reveal the biodiversity under their care? Almost every Brazilian herbarium provides access to its data via two large national projects: INCT Herbário Virtual da Flora e dos Fungos and ReFlora Virtual Herbarium. All data have recently been made available at the SiBBr website (SiBBr 2020) node to GBIF 2020 (<https://www.gbif.org/>). This was a huge step towards providing access to millions of occurrence records. On the other hand, few Brazilian herbaria have

all their collections imaged, due to reasons cited above: few trained people and many technical difficulties, such as access to a high-quality digital camera and software for image improvement, transfer, and storage.

In summary, photographing an herbarium collection is not an easy task, but it is not impossible either. Here, we describe some case studies of small herbaria with similar approaches.

Case studies

FURB, CRI, and LUSC

FURB (Herbário Dr. Roberto Miguel Klein, Universidade Regional de Blumenau, Blumenau, Santa Catarina) is new relative to many other Brazilian herbaria (Gasper *et al.* 2014). The collection was created in the 1900s and began to be better structured in 2003. A large and systematic project (Floristic and Forest Inventory of Santa Catarina state; Vibrans *et al.* 2010) resulted in its exponential growth from 2007 to 2012. The digitization process began when FURB had 40,000 specimens (now the collection has 69,976). A professional camera and support designed by the herbarium staff were first used (Fig. 3A), and one sample of each species (~8,000) was photographed. Then, with ReFlora and



Figure 3. **A)** First images from the FURB herbarium collection. Initial image of low quality. **B)** The second image with better light and camera equipment.

SiBBr projects, better equipment was acquired, such as an improved lighting system, appropriate lenses, and a barcode scanner. All of this made it possible to photograph the entire collection, including re-photographing specimens from the first phase since it was faster than selecting those that had already been imaged (Fig. 3B). One hundred percent of the database was digitized in less than one year due to good practices of software and database management. The most time-consuming task was adding barcodes to each sample.

Using a Canon 6D camera (20.2 megapixels and 300 dpi resolution), software to help manage digital specimen sheets (Silverimage 2020), a barcode reader, and remote camera managing software (Digital Photo Professional from Canon) made it possible to photograph about 500 specimens per day, a number similar to that cited by Borsch *et al.* (2020) in a German initiative. A color and scale bar were fixed on the support where specimens were placed to be photographed, which saved time during the process (Fig. 3B). Images were also sent to the Reflora Virtual Herbarium server to be published via FTP transferring. File size was ~10 megabytes each, which allowed good analysis of specimens and ensured that storage is possible on a large scale (Borsch *et al.* 2020). Of course, high-performance imaging, as described by Tegelberg *et al.* (2014), is the ideal process, but the same goal was reached with fewer resources and more time.

Once the entire collection was photographed, two different options presented themselves: (1) loan the equipment to other collections; or (2) receive samples to be digitized by the FURB team. The first option was not adopted since an average of 3,500 specimens are added to FURB per year and lending the equipment would prohibit photographing at FURB until the equipment was returned. The second option seemed better, since small pauses in digitizing FURB material were possible. We then proposed to two other nearby herbaria, CRI (Herbário Pe. Dr. Raulino Reitz, Universidade do Extremo Sul Catarinense – UNESC, Criciúma, Santa Catarina) and LUSC (Herbário da Universidade do Estado de Santa Catarina, Lages, Santa Catarina), that they could send us specimens (about 2,000 at a time) to be digitalized by FURB. As the herbarium team was already trained, the process went fast, without the need to stop photographing of the FURB collection. The only requirement was that all specimens needed to be received already barcoded to speed up the process. A total of 17,094 specimens of CRI and LUSC were digitalized.

Four trips were made from Criciúma to Blumenau, from August 22 to October 14, 2016, with an average of five hours to travel 338 km. Each trip transported ~2,500 specimens, for a total of 9,942 photographed specimens and 10,156 images. The herbarium staff found ~7,000 errors, mostly in the database, which were corrected based on sample images. Since the project ended, CRI had photographed 457 specimens with its own resources and ten students, professors, and technicians helping with the process.

The same methodology was used for LUSC, where four trips from Lages to Blumenau, from November 10 to December 15, 2017, with an average of four hours to drive 220 km. A total of 6,938 photographs were taken of 6,938 specimens and 597 errors were detected and corrected in the database.

UFRN

UFRN herbarium (Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte) is also a recently established collection, dating back to the 1990s. After 2010, however, when the number of taxonomists in the University (UFRN) increased, as well as the number of graduate programs, with students conducting their own research and field collections, the herbarium expanded. It is currently the largest collection for the Flora of Rio Grande do Norte State, housing almost 27,000 specimens, of which ~26,000 are digitized.

The procedures used for digitizing were the same as described for FURB, except that the photos were taken during the National Forest Inventory in 2015. A Canon EOS 6D camera (20.2 megapixels and 300 dpi resolution) was used with software to help manage digital specimen sheets (Silverimage 2020) and a professional camera stand with two lights. No *a priori* separation of specimens by size was done since a frame to fit the most common dimensions of specimen sheets (*i.e.*, around 29 x 41 cm) was adjusted, on which the color and metric scales were added. Smaller sheets were eventually cropped manually using the remote camera managing software (Digital Photo Professional, Canon). A maximum of 300 photos was taken per day following the alphabetical order of families, while large loans were suspended during the process. For smaller loan requests, specimens were photographed prior to sending, all of which had the letter F (for Portuguese word “fotografado”, meaning photographed) manually stamped on the left corner of the sheet. This practice proved useful during and after the digitization because it rapidly indicates which specimens have already been photographed. It also made it easier and quicker to sort specimens that were not digitized once the process was finished.

Digitizing at UFRN was less productive than at FURB, and the goal of 300 photos/day was not easily achieved every day. This was due to the process of renaming each file in the manual mode of Silverimage, due to the lack of a barcode label to scan on the first 20,000 specimens, and to computer capacity to manage the files. Also, as three different people were digitizing, individual abilities and basic knowledge of curatorial work affected the results. Nevertheless, UFRN was one of the first Brazilian herbaria to be completely digitalized and is still able to photograph every specimen prior to it entering the collection.

It is important to highlight that before photographing began, specimens on loan to several herbaria were requested to be returned, and database problems were corrected.



These activities ended up occurring concomitantly with digitization and the distribution of nearly 5,000 duplicates accumulated in the collection. Although many curators would wait until they have the best conditions of their specimens to start photographing, this was not the case at UFRN, since the project had a predetermined period and many tasks to accomplish. But the cleaner and better determined, checked, and complete the database is, the better and faster will be the results.

Another issue at UFRN was barcoding. The initial photographs were not followed by the addition of a unique barcode and files were renamed manually (manual mode in Silverimage 2020) using the catalog number to “build” the barcode. This procedure delayed the photographing phase, but was not an impediment (but see the section Lessons learned). After trying, and experiencing difficulties, adding a barcode in the same Brahm's generated label for the specimens, the database was moved to the JABOT platform (Silva *et al.* 2017), where this issue was easily fixed. Thus, all specimens from catalog number 20,000 onwards have standard JABOT labels containing automatically generated barcodes (in this case identical to the catalog number, but see limitations in section Lessons learned). Two alternative ways to barcode all of the first 20,000 records were to print new labels already containing the barcode, or simply printing separate barcode labels from number 1 to 20,000 and match these labels with the corresponding specimens. JABOT proved very helpful for the latter option, enabling the generation of PDFs containing lists of barcodes by families, which could be printed on standard Avery adhesive label sheets with a laser printer and easily pasted onto specimens.

Some problems encountered during the photographing process that other herbaria may face include: finding many lost specimens, empty numbers in the database, same collection (same collector and collector number) with different catalog numbers, conflicting catalog numbers (*i.e.*, the same catalog number for different collections), or ending up with photos of specimens that were not databased but were already deposited in the collection (some of them even having a catalog number but lacking data in the database). These totaled around 2,000 problematic records, which were mainly due to the lack of a responsible person for the database. There were several substitutions of students working with the Brahm's database and many problems with the initial herbarium numbers, which were probably related to the loss of specimens and not updating this information in the database. As an example, only in 2016, we found out that the entire fern collection was lost after a loan and many specimens dating back to the 1990s were taken, without consent or any formal loan request, from the UFRN to Museu Câmara Cascudo. The case of Museu Câmara Cascudo is noteworthy because, although belonging to the same institution, the collection was completely abandoned, and the museum staff did not know what to do with it after

the curator retired. After a negotiation, this collection was merged with UFRN in 2016, although most of the specimens demanded time-consuming restoration or were definitively lost. Despite comprising only around 1,000 specimens, most of them were algae and very representative for the state Flora, and the study of these specimens resulted in several new occurrences (Versieux *et al.* 2017). As floristic studies in Rio Grande do Norte are increasing, UFRN has offered support to photograph the collections of Parque das Dunas Herbarium (RN) and the Federal University of Semi-Arid in Mossoró (MOSS), but this has yet to begin.

ECT

The ECT herbarium (Embrapa Clima Temperado, Pelotas, Rio Grande do Sul) started from a collection of vouchers from accession specimens of wild relatives of crops obtained by fieldwork for taxonomical support of gene bank management in 1986. Since 2003, the collection has broadened its focus and increased the number of specimens by the means of floristic surveys, useful plant prospecting, and ethnobotanical studies focusing on temperate and subtropical Brazilian ecosystems, along with the Pampas and Atlantic Forest domains (Heiden & Barbieri 2015).

ECT was registered in Index Herbariorum in 2016, and currently houses 8,826 exsiccates, all of which are photographed and databased, including geographic coordinates for 91 % of the records. The data and images of these specimens are available through online platforms. The collection has only one taxonomist as permanent staff and collecting, drying, and mounting of herborized specimens and sanitary control, as well as duplicate separation, databasing, label typing, identification, and placement in the collection, are responsibilities shared with fellowship holders and undergraduate and graduate students, post-doctoral fellows and other researchers who perform and/or collaborate on tasks demanding taxonomic support, all of them supervised by the curator. To overcome the deficiency of staff and accessibility for herbarium visits, since it is located in a rural area in a medium-sized municipality in a sparsely populated region of Brazil, publishing online was not simply an alternative, but had to be considered a priority and the only way for the collection to be accessible by third parties and keep updated. Complimentary to putting a great effort into digitization, the policy of the herbarium is to immediately distribute duplicates to nine other major Brazilian herbaria, thereby easing the demand for specimen consultation for identification by collaborating specialists.

The herbarium had no acronym prior to 2003, when it adopted HCT, a previously occupied acronym, so it moved to HECT in 2009 and then to ECT in 2014, which was finally indexed in 2016. Prior to 2014, the other acronyms were not consistently adopted and no digitization in public databases existed, except for only some typed label data that were lost due to corruption in the database. ECT joined the REFLOA project in 2015 and discontinued the record number of



specimens with previous unofficial acronyms to avoid mistakes and reproduction of doubtful data. The previous database was completely abandoned, and digitization adopting record barcodes restarted from scratch. It was later decided to adopt only barcodes as reference numbers for specimens to avoid the proliferation of reference numbers. REFLORA funded one fellowship to hire a technician for two years to work on the collection digitization project.

The technician started working before the arrival of the photographing station. Given this constraint, ways were planned to optimize digitization by anticipating problems that could be found during the process. This, in fact, was a good opportunity to develop a pipeline to anticipate any type of problem and seek solutions to optimize photographing. We consider that this could be applied to any small herbaria willing to digitize their collections, while not having immediate funds to acquire, or the ability to borrow, a digitizing station. We began revising the physical content of specimens, standardizing folders and cardboard sizes and colors, checking mounting style and remounting when appropriate, putting small folders with all specimens for detached parts, and subsequently doing a purge for preventive disinfection of any putative contaminated material. This was followed by barcoding and typing, checking all names on IPNI, and updating taxa names to those accepted by Brazilian Flora (2020). During the typing of label data it was decided that, although time-consuming, 100 % of label data should be included, otherwise critical information could never be retrieved and it would be more laborious to return to the same specimens to type additional data in the future or, even worse, the data may never be typed.

During digitization, the opportunity was taken to standardize collector names and update obsolete geography data, when strongly supported, while keeping original information. Meanwhile, duplicates were separated for distribution and occasionally other unnoticed duplicates from the same collections (but recorded twice or thrice in the herbarium) were found and separated to be shipped to other herbaria avoiding overlapping and redundant specimens within the same collection. This process was done in batches, mostly organized by family, since the same problems tended to be repeated in similar specimens from the same taxonomic group. After all these specimens were ready, a further step was added and specimens with no original coordinates were georeferenced, always taking care to inform that added coordinates were not original and pointing out the level of accuracy of the data added (such as flagging when coordinates were plotted over cities and not countryside localities, for example). The entire pre-digitization process proved to speed up the digitization process itself later, which made this our most important. After some months of working on batches of specimens to be ready for photographing, the equipment arrived, and we acted as a focal point, sharing training and facilities with

other nearby countryside herbaria from the of the state of Rio Grande do Sul.

The digitization process was the same as previously explained for FURB and UFRN (same station design, camera model, and software). Two courses were offered to train students and professionals on herbaria digitization and a talk was given at the annual state meeting of herbaria explaining the pathways described above, which other small herbaria can follow to anticipate problems and accelerate digitization when equipment becomes available.

ECT has now finished its digitization and is sharing the REFLORA photographing station with other herbaria from the state of Rio Grande do Sul. ECT data are now fully accessible at *Herbário Virtual Reflora*, *SiBBR*, *speciesLink*, and *GBIF*. The station is currently at PEL (25,541 specimens) and soon is going to be moved to HUCS (50,000) and HVAT (8,000) allowing a circuit of digitization and 83,541 more samples digitalized. Other smaller herbaria from the state are willing to borrow the equipment with those that have already done all the previous steps necessary to speed up photographing being prioritized. Meanwhile, these herbaria are not provided access to the equipment.

Lessons learned

We can highlight some of the lessons learned from our experiences digitizing distinct herbaria in Brazil with different realities and difficulties:

Sharing equipment optimizes the process of digitizing smaller herbaria since costs (or investments) are minimized, and improved use of the equipment minimizes downtime, since while one collection is digitalizing, other(s) can solve curatorial problems and restore and/or gather new specimens to be photographed. Shipping equipment also creates a collaboration network that aids in solving problems inherent to the process. Furthermore, sharing training lessons, if there is such opportunity, and having the same photographer perform the procedures tends to result in more homogeneous work;

Whether the equipment will be transferred between institutions or materials will be sent to a collection where the equipment will be stationary, is something to be decided between curators, considering the size of each collection, annual input, and distances to be covered. However, we advise not shipping all specimens at once because of the risk of accidents in the route. Also, the process should not be done within the herbarium room itself, so specimens do not need to be purged of pests;

Prior to photographing, mounting and labeling should be revised and specimens restore on demand. It is not recommended to start digitization if a collection needs to be recovered, if plants are not well fixed onto the cardboard, if labels are missing or are covered by specimens, or if specimens need restoration for cleaning pest damage and removing bugs. This takes time and should be done before taking photos, especially with regard to purging — to avoid



cross-contamination and spreading of pests. The same is true for pasting barcodes and typing label data; if an herbarium lacks a digitization facility, these steps should be done in advance of photographing, thus speeding up the process when a camera becomes available;

Choose good management software for the collection by taking into consideration the internet access of the institution and the number of people working in the team. Be sure that almost all, if not all, vouchers are databased before photographing. This facilitates the work, but it is not mandatory since the images themselves can be used to build and fix the database. It is preferable to choose a software that automatically generates a unique identifier number for each specimen. Otherwise, this information should be manually entered into the database, especially if the intention is to publish it in a large biodiversity repository such as GBIF. Brazilian herbaria usually use a collection number, but it is relatively common to have errors, with different specimens under the same number due to curatorial or collecting problems. Thus, having a unique identifier with an independent sequence of numbers makes digitalization quicker and safer.

Images must be named with the barcode number, adding a numerical or alphabetical suffix to collections with multiple sheets (e.g., parts of leaves and inflorescence of *Arecaceae* and *Bromeliaceae*). With most software, this is the information that will relate the image to its metadata and biodiversity repositories. Fixing a barcode label onto the specimen and using a barcode scanner to name the image files speeds up the process and avoids typing errors. Using the same number (collection number = unique identifier/barcode) is time-consuming since there is a need to find the specific material related to the barcode;

When possible, material with cardboard of the same size and of similar color should be arranged first because it saves time during the batch process by minimizing the need to make several manual adjustments, such as to brightness and contrast and rotating and cutting the image.

Conclusion

Small herbaria are generally less accessible and less visited, but many host important data for local Floras and, thus, worldwide plant knowledge. Making herbaria occurrence information available through biodiversity platforms must be a continuous program so the data can be more easily accessed by the scientific community and policymakers.

Acknowledgements

We are thankful to Sistema de Informação sobre a Biodiversidade Brasileira (SiBBR) and Inventário Florestal Nacional (IFN), as well CNPq/Reflora Project for equipment acquisition; the IT staff of JBRJ, and COPPE-UFRJ, who have

been working non-stop to keep the infrastructure operating. We thank FAPERJ for grants to PML (E-26/202.435/2018), and RCF (E-26/202.778/2018). We also thank CNPq for the productivity grants to ALG (311303/2020-0), GH (n°314590/2020-0), RCF (303059/2020-6), LMV (383382/2015-7).

References

- Barkworth ME, Murrell ZE. 2012. The US virtual herbarium: Working with individual herbaria to build a national resource. *Zookeys* 209: 55-73.
- Borges LM, Reis VC, Izbicki R. 2020. Schrödinger's phenotypes: Herbarium specimens show two-dimensional images are both good and (not so) bad sources of morphological data. *Methods in Ecology and Evolution* 11: 1296-1308.
- Borsch T, Stevens A-D, Häffner E, *et al.* 2020. A complete digitization of German herbaria is possible, sensible and should be started now. *Research Ideas and Outcomes* 6: e50675. doi: 10.3897/rio.6e50675
- Brazilian Flora 2020. Jardim Botânico do Rio Janeiro. <http://floradobrasil.jbrj.gov.br/>. 25 May 2019.
- Brazilian Herbaria Network. 2020. <https://www.botanica.org.br/a-rede-brasileira-de-herbarios/>. 26 Jan. 2020.
- Canteiro C, Barcelos L, Filardi F, *et al.* 2019. Enhancement of conservation knowledge through increased access to botanical information. *Conservation Biology* 33: 523-533.
- Chapman AD. 2005. Principles and methods of data cleaning—primary species and species-occurrence data. version 1.0. Copenhagen, Report of the Global Biodiversity Information Facility.
- Colombo B, Kaehler M, Calvente A. 2016. An inventory of the Bignoniaceae from the Brazilian state of Rio Grande do Norte highlights the importance of small herbaria to biodiversity studies. *Phytotaxa* 278: 19-28.
- Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE. 2013. Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution* 28: 454-461.
- Dillen M, Groom Q, Chagnoux S, Güntsch A, Hardisty A, Haston E. 2019. A benchmark dataset of herbarium specimen images with label data. *Biodiversity Data Journal* 7: e31817. doi: 10.3897/BDJ.7.e31817
- Downton AC. 2005. Online acquisition of scientific archive documents – a survey and manual. In: Häuser CL, Steiner A, Holstein J, Scoble MJ. (eds.) *Digital imaging of biological type specimens. A manual of best practice*. Stuttgart, Germany, European Network for Biodiversity Information. p. 56-86. <http://www.gbif.org/resource/80576>. 01 May 2020.
- Forzza RC, Carvalho Junior A, Andrade AC, *et al.* 2017a. As coleções biológicas do Jardim Botânico do Rio de Janeiro à luz das metas da GSPC/CDB. *Revista Museologia & Interdisciplinaridade* 5: 125-141.
- Forzza RC, Dalcin EC, Silva LAE, *et al.* 2017b. Plantas do Brasil: Resgate Histórico e Herbário Virtual para o Conhecimento e Conservação da Flora Brasileira-REFLORA <http://dspace.jbrj.gov.br/jspui/bitstream/doc/104/1/Relat%C3%B3rio%20REFLORA%20FAPERJ-CNPq-SiBBR%202011-2017.pdf>.
- Forzza RC, Filardi FLR, Condack JPS, *et al.* 2015. Herbário Virtual Reflora. *Unisanta BioScience* 4: 88-94.
- Gasper AL, Vibrans AC, Funez LA, Rigon-Jr MJ, Bittencourt F, Vieira C. 2014. Dr. Roberto Miguel Klein Herbarium (FURB), Blumenau, Southern Brazil. *PhytoKeys* 37: 21-37.
- Gasper AL, Stehmann JR, Roque N, Bigio NC, Sartori ALB, Grittz GS. 2020. Brazilian herbaria: an overview. *Acta Botanica Brasílica* 34: 352-359.
- GBIF – Global Biodiversity Information Facility. 2020. <https://www.gbif.org/>. 26 May 2020.
- Guralnick RP, Hill AW, Lane M. 2007. Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10: 663-672.
- Hampton SE, Strasser CA, Tewksbury JJ, *et al.* 2013. Big data and the future of ecology. *Frontiers in Ecology and the Environment* 11: 156-162.



- Harris KM, Marsico TD. 2017. Digitizing Specimens in a Small Herbarium: A Viable Workflow for Collections Working with Limited Resources. *Applications in Plant Sciences* 5: 1600125. doi 10.3732/apps.1600125
- Heiden G, Barbieri RL. 2015. Herbário da Embrapa Clima Temperado, Rio Grande do Sul (ECT). *Unisantia BioScience* 4: 209-212.
- JSTOR – Journal Storage. 2018. JSTOR Global Plants: Guidelines for scanning specimens. https://guides.jstor.org/ld.php?content_id=31764146. 26 May 2020.
- JSTOR – Journal Storage. 2020. Global Plants. <https://plants.jstor.org/> 26 May 2020.
- Le Bras G, Pignal M, Jeanson ML, *et al.* 2017. The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. *Scientific Data* 4: 1-16.
- Maia LC, Vieira AOS, Canhos DAL, Stehmann JR, Barbosa MRV, Menezes M. 2017. Construindo redes para promover o conhecimento da biodiversidade brasileira: a experiência do INCT - Herbário Virtual. Recife, Editora Universitária da UFPE
- Nieva de la Hidalga A, Cazenave N, Agosti D, Wu Z, Dillen M, Nielsen L. 2019. Use of European Open Science Cloud and National e-Infrastructures for the Long-Term Storage of Digitised Assets from Natural History Collections. *Biodiversity Information Science and Standards* 3: e37164. doi 10.3897/biss.3.37164
- Nualart N, Ibáñez N, Soriano I, López-Pujol J. 2017. Assessing the Relevance of Herbarium Collections as Tools for Conservation Biology. *Botanical Review* 83: 303-325.
- Peixoto AL, Barbosa MRV, Menezes M, Maia LC. 2006. Diretrizes e estratégias para a modernização de coleções biológicas brasileiras e a consolidação de sistemas integrados de informação sobre biodiversidade. Brasília, Centro de Gestão e Estudos Estratégicos, Ministério da Ciência e Tecnologia.
- Pignal M, Romaniuc-Neto S, Souza S, Chagnoux S, Canhos DAL. 2013. Saint-Hilaire virtual herbarium, a new upgradeable tool to study Brazilian botany. *Adansonia* 35: 7-18.
- Pryer KM, Tomasi C, Wang X, Meineke EK, Windham MD. 2020. Using computer vision on herbarium specimen images to discriminate among closely related horsetails (*Equisetum*). *Applications in Plant Sciences* 8: e11372. doi 10.1002/aps3.11372
- RECOLNAT. 2020. *Annotate v1.4.1*. <https://www.recolnat.org/en/annotate/>. 26 May 2020.
- REFLORA Programme. 2020. Refflora Programme. <https://www.gov.br/cnpq/pt-br/aceso-a-informacao/acoes-e-programas/programas/reflora/>. 2 May 2020.
- Silva LAE, Fraga CN, Almeida TMH, *et al.* 2017. Jabot - Sistema de Gerenciamento de Coleções Botânicas: a experiência de uma década de desenvolvimento e avanços. *Rodriguésia* 68: 391-410.
- Silverimage. 2020. Silverimage. <http://www.silverbiology.com/products/silverimage/>. 26 May 2020.
- SiBBR – Sistema da Informação sobre a Biodiversidade Brasileira. 2020. <https://www.sibbr.gov.br>. 01 Feb. 2020.
- Smith GF, Roux JP, Raven P, Figueiredo E. 2011. African Herbaria Support Transformation on the Continent. *Annals of the Missouri Botanical Garden* 98: 272-276.
- Smith VS, Blagoderov V. 2012. Bringing collections out of the dark. *Zookeys* 209: 1-6.
- Soberón J, Peterson AT. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* 359: 689-98.
- Soltis OS. 2017. Digitization of herbaria enables novel research. *American Journal of Botany* 104: 1-4.
- Takano A, Horiuchi Y, Fujimoto Y, Aoki K, Mitsuhashi H, Takahashi A. 2019. Simple but long-lasting: A specimen imaging method applicable for small- and medium-sized herbaria. *PhytoKeys* 118: 1-14.
- Taniguchi M, Condack JP, Leitman PM, *et al.* 2018. Manual de digitalização REFLORA. http://dspace.jbrj.gov.br/jspui/bitstream/doc/103/1/Manual_Digitalizacao_exsicatas_2018.pdf.
- Tegelberg R, Mononen T, Saarenmaa H. 2014. High-performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. *Taxon* 63: 1307-1313.
- Thiers B. 2019, continuous adapted. The World's Herbaria 2018: a summary report based on data from *Index Herbariorum* http://sweetgum.nybg.org/science/docs/The_Worlds_Herbaria_2018.pdf.
- Thiers B. 2020, continuous adapted. *Index Herbariorum*: A global directory of public herbaria and associated staff. <http://sweetgum.nybg.org/science/ih/>. 26 May 2020.
- Thiers B, Tulig MC, Watson KA. 2016. Digitization of The New York Botanical Garden Herbarium. *Brittonia* 68: 324-333.
- Thomas WW, Forzza RC, Michelangeli FA, Giulietti AM, Leitman PM. 2012. Large-scale monographs and floras: the sum of local floristic research. *Plant Ecology & Diversity* 5: 217-223.
- Tulig M, Tarnowsky N, Bevans M, Kirchgessner A, Thiers B. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 113: 103-113.
- Versieux LM, Dávila N, Delgado GC, *et al.* 2017. Integrative research identifies 71 new plant species records in the state of Rio Grande do Norte (Brazil) and enhances a small herbarium collection during a funding shortage. *PhytoKeys* 2017: 43-74.
- Vibrans AC, Sevegnani L, Lingner DV, Gasper AL, Sabbagh S. 2010. Inventário florístico florestal de Santa Catarina (IFFSC): aspectos metodológicos e operacionais. *Pesquisa Florestal Brasileira* 30: 291-302.
- Vollmar A, Macklin JA, Ford L. 2010. Natural History Specimen Digitization: Challenges and Concerns. *Biodiversity Informatics* 7: 93-112.
- Wieczorek J, Bloom D, Guralnick R, *et al.* 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLOS ONE* 7: e29715. doi: 10.1371/journal.pone.0029715

