adansonia

2024 • 46 • 13

Récolnat Annotate-On: a tool to improve your experience with virtual collections

Marc PIGNAL, Gilles BERTIN, Lisa CHUPIN, Eva PEREZ PIMPARÉ, Srdjan KLASNJA, Régine VIGNES LEBBE & François DUSOULIER

art. 46 (13) — Published on 26 August 2024
www.adansonia.com

PUBLICATIONS SCIENTIFIQUES

MUSÉUM NAL HIST NATURELLE

Couverture / *Cover*:
Réalisée à partir des Figures de l'article/*Made from the Figures of the article*.

Adansonia est indexé dans / Adansonia *is indexed in:*
– Science Citation Index Expanded (SciSearch®)
– ISI Alerting Services®
– Current Contents® / Agriculture, Biology, and Environmental Sciences®
– Scopus®

Adansonia est distribué en version électronique par / Adansonia *is distributed electronically by:*
– BioOne® (http://www.bioone.org)

Adansonia est une revue en flux continu publiée par les Publications scientifiques du Muséum, Paris
Adansonia *is a fast track journal published by the Museum Science Press, Paris*

Les Publications scientifiques du Muséum publient aussi / *The Museum Science Press also publish:* ***Geodiversitas***, ***Zoosystema***, ***Anthropozoologica***, ***European Journal of Taxonomy***, ***Naturae***, ***Cryptogamie*** **sous-sections** ***Algologie***, ***Bryologie***, ***Mycologie, Comptes Rendus Palevol***

Diffusion – Publications scientifiques Muséum national d'Histoire naturelle
CP 41 – 57 rue Cuvier F-75231 Paris cedex 05 (France)
Tél.: 33 (0)1 40 79 48 05 / Fax: 33 (0)1 40 79 38 40
diff.pub@mnhn.fr / http://sciencepress.mnhn.fr

ISSN (imprimé / *print*): 1280-8571/ ISSN (électronique / *electronic*): 1639-4798

Récolnat Annotate-On: a tool to improve your experience with virtual collections

Marc PIGNAL
Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, case postale 39, 57 rue Cuvier, 75231 Paris cedex 05 (France)
marc.pignal@mnhn.fr (corresponding author)

Gilles BERTIN
Lisa CHUPIN
Laboratoire Dicen-IDF, Dispositifs d'Information et de Communication à l'Ère numérique – Paris, Île-de-France, unité de recherche multi-tutelles (EA 7339 – CNAM, UPEM, université Paris-Nanterre)
gilles.bertin@lecnam.net
lisa.chupin@parisdescartes.fr

Eva PEREZ PIMPARÉ
Infrastructure Récolnat, Direction générale déléguée aux collections, Muséum national d'histoire naturelle, case postale 23, 57 rue Cuvier, 75231 Paris cedex 05 (France)
eva.perez@mnhn.fr

Srdjan KLASNJA
Presek Informatique, 2. srpske armije 31, Belgrade (Serbia)
srdjan.klasnja@presek-i.com

Régine VIGNES LEBBE
Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, case postale 48, 57 rue Cuvier, 75231 Paris cedex 05 (France)
regine.vignes_lebbe@sorbonne-universite.fr

François DUSOULIER
Infrastructure Récolnat, Direction générale déléguée aux collections, Muséum national d'histoire naturelle, case postale 23, 57 rue Cuvier, 75231 Paris cedex 05 (France)
francois.dusoulier@mnhn.fr

Submitted on 2 September 2022 | accepted on 26 February 2024 | published on 26 August 2024

Pignal M., Bertin G., Chupin L., Perez Pimparé E., Klasnja S., Vignes Lebbe R. & Dusoulier F. 2024. — Récolnat Annotate-On: a tool to improve your experience with virtual collections. *Adansonia*, sér. 3, 46 (13): 133-148. https://doi.org/10.5252/adansonia2024v46a13. http://adansonia.com/46/13

ABSTRACT
Annotate-On software (https://www.recolnat.org/en/annotate) is freely available to botanists for research on virtual herbarium collections. It contains tools to easily and quickly perform morphological and morphometric analyses on images of specimens including physical measurements (lengths, perimeters, areas, angles, counts) and descriptive annotations. The images derive from various sources such

KEY WORDS
Digitised collections, digital twins, herbarium, e-services, descriptive measurements, teleworking.

as the online Récolnat infrastructure database (https://explore.recolnat.org/), online image URLs or other images. Image libraries can be sorted using the keyword management tool within the software. Annotate-On is built on models structured as lists of editable descriptive characters (observations or measurements) dedicated to a study, a species, an hypothesis, etc. Observations and measurements based on these models can be exported to spreadsheets, data processing software, R applications or deep learning modules. Annotate-On is associated with Xper3 biodiversity data management software (https://www.xper3.fr/).

MOTS CLÉS
Collections numérisées, doubles virtuels, herbiers, services numériques, mesures descriptives, travail à distance.

RÉSUMÉ
Récolnat Annotate-On *: un outil pour améliorer votre expérience avec les collections virtuelles.*
Le logiciel Annotate-On (https://www.recolnat.org/en/annotate) est mis librement à la disposition des botanistes pour mener des recherches sur des collections d'herbiers virtuels. Il contient des outils pour réaliser facilement et rapidement des analyses morphologiques et morphométriques sur des images de spécimens : mesures physiques (longueurs, périmètres, surfaces, angles), marquage et comptage de points d'intérêts, et annotations descriptives. Les images proviennent de sources variées : outre la base de données de l'infrastructure Récolnat (https://explore.recolnat.org/) en ligne, toute URL d'images en ligne, ou encore les images d'un disque dur, d'un appareil photographique, d'un microscope, etc. Ces bibliothèques d'images peuvent être classées à l'aide de l'outil de gestion de mots-clés du logiciel. Les annotations effectuées avec Annotate-On sont structurées par une liste de caractères descriptifs éditables – observations ou mesures – dédiés à une étude, une espèce, une hypothèse, etc. Ce modèle peut être importé depuis le logiciel Xper 3, logiciel de gestion de données sur la biodiversité (https://www.xper3.fr/), ou exporté vers lui. Les observations et mesures fondées sur ces modèles sont exportables vers des tableurs, des logiciels de traitement de données, l'application R ou encore des modules d'apprentissage profond (deep learning).

INTRODUCTION

Collection virtualisation dates from before the birth of computers. Between 1929 and 1939, J. Francis MacBride created a collection of negatives of European types for South American vascular plants (Grimé & Plowman 1986). These negatives are held at the Field Museum in Chicago, Illinois (United States) and can be viewed at https://collections-botany.fieldmuseum.org/. The destruction of the Berlin herbarium by bombing on 1st and 2nd of March 1943 (Merrill 1943) rendered this digital twin collection particularly valuable. More recently, the loss of several million natural history specimens from the Natural History Museum of Rio de Janeiro due to a catastrophic fire on 2nd September 2018, shows that such disasters are not a thing of the past and that digitising collections can provide a safeguard (Coughenour 2018).

Virtual collections, in particular herbaria, are being developed in many countries (Pignal *et al.* 2013; Barkworth & Murrell 2012; Brazil Flora Group 2018). The first large-scale digitisation initiative was carried out in France between 2010 and 2013 at the Paris herbarium (Le Bras *et al.* 2017). This initiative has been extended to all French natural history collections within the framework of the national research infrastructure Récolnat (Pignal & Pérez 2013; Chupin 2017). The optimisation of collection management and access is accompanied by the description and digitisation of natural history collections at the national and international level. At the core of its mission is the production and provision of a database for the study of current and past geodiversity and biodiversity. Récolnat, as the French nodal point, is actively participating in the construction of DiSSCo (Distributed System of Scientific Collections), its mirror infrastructure at a European level (Dusoulier 2022).

The availability of images on the internet gives a broad range of opportunity to make basic observations, but does not allow for the acquisition of more methodical data, in particular, measurements. It is also challenging to make comparisons between specimens, as no appropriate tools are specifically designed for this purpose. Hence, one of the goals of Récolnat was to develop software that would complement digitisation with tools to facilitate the study of virtual collections in the most favourable conditions.

The software Annotate-On was designed to develop the full potential of collection digitisation by reconstituting the tools used by scientists working on herbarium plates. The goal is to replicate working practices and documentation specific to each discipline. Furthermore, it is a question of harnessing the potential inherent in the "virtualisation" of herbaria by implementing, within these new digital workbenches, collaborative practices of discussion and validation between researchers. These previously difficult practices are made possible by digital technology. Deep learning can also be used to recognise characters, morphological traits and, notably, forms. These techniques require clean and validated datasets to feed the models (Zhu *et al.* 2017; Hussein *et al.* 2022).

Digitisation of collections provides a digital twin of a specimen, increasing the number of potential uses and users through online access. Furthermore, with digitised objects, it becomes possible to compare and enrich data and images, cross-validate information and deduce additional data. Virtual herbaria do not replace physical herbaria; in particular, they do not allow for dissection or molecular analysis (Pignal & Pérez 2013).

Characters	#	M	SD
• leaf width	3	2.48mm	0.65mm
Global value	147	2.30 mm	0.54mm
• leaf length	3	90.20mm	8.86mm
Global value	138	70.46mm	22.91mm

Fig. 1. — Examples of mean (**M**) and standard deviation (**SD**). Two characters: blade length (***Fll L***) and blade width (***fll- larg***), number of measures (**#**). First line: image measures; second line: batch measures (global value).

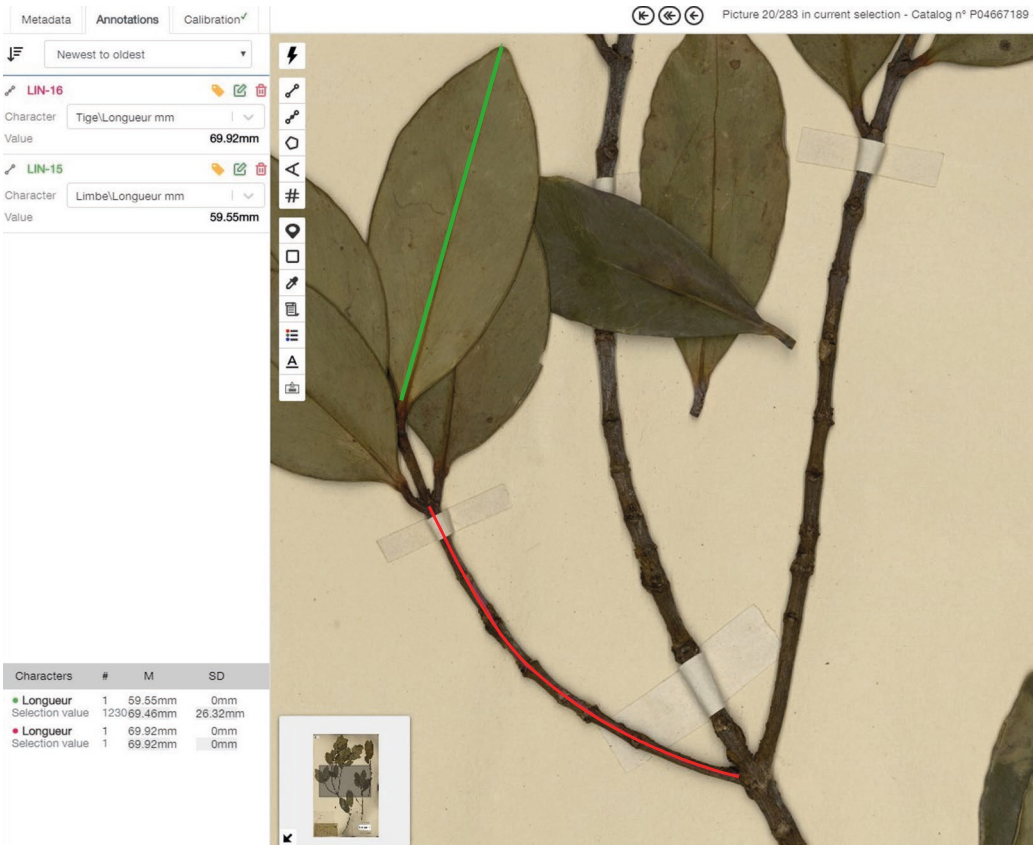

Fig. 2. — Example of measurements: linear (**green line**) or non-linear (**red line**).

TOOL DESCRIPTION

Preliminary study: practices of natural history collections

The management and use of herbarium collections have been impacted by a hybridisation of media since the beginnings of scientific computing. Institutions that have catalogued part of their collections have existing databases linking specimens (identified by a catalogue number or a barcode) with collection data and images, thus forming the basis of a virtual herbarium. Our observations of the uses of these first digital collections (Chupin 2015) reflect that, for many users, they remain a catalogue. This catalogue is useful for preparing a visit to a physical collection, even when those databases include functionalities for enriching, correcting and improving data quality associated with images (e.g. Jstor Global Plants, https://plants.jstor.org/), for understanding the distribution of species by viewing specimen records (e.g. GBIF international, https://www.gbif.org/) or exporting datasets from collections. This last functionality already occurs on existing virtual collection websites and is aligned with the FAIR (Findable, Accessible, Interoperable and Reusable) principles of open data.

Users of physical collections were interviewed during the design study for Annotate-On. One of the main needs that emerged was to equip the virtual herbarium with relevant digital functions so that it would acquire the same scientific material status as the existing physical herbarium (Chupin 2017). The survey and our objectives led to the implementa-

Fig. 3. — Example of measurements: surface and perimeters.

tion of measurement and calibration tools with annotation tools (free or associated with a data model created by the user) within Annotate-On. These coupled functionalities facilitate an operation that is common to multiple works. In systematics, this is specifically the case when describing specimens within a sequence of observations (or measurements) of characters. It was also necessary to identify which computer technologies for document management could be beneficial to user workflow, i.e. the sequence of practices linked to the study of specimens, conservation, research or scientific diffusion. It was also a question of anticipating novel uses.

Powerful software tools already exist, such as ImageJ (https://imagej.net), however they are not designed specifically for natural history objects (Ferreira & Rasband 2012). The dedicated Annotate-On tool is based on our usage analysis and on the Web Annotation Data Model (https://www.w3.org/TR/annotation-model/), making it possible to associate annotations with delimited inner parts of web resources (image, text, etc.) and to share them. A comparison of ImageJ software with Annotate-On is shown in Table 1.

Modules description

Annotations

Annotations can be morphometric (lengths, areas, perimeters, angles or counts) or qualitative (points of interest, etc.) and each can be modified, deleted or associated with previously defined tags. They can also be assigned to a character of a previously defined description model created by the user. All or part of the data produced can be exported in CSV format or XML format for processing and treatment by other software.

Table 1. — Comparison of some specifications for Annotate-On and ImageJ software.

	Annotate-On	ImageJ
Measures		
Calibration assistant	with a measuring tool	declarative
Measures proposed	simple and segmented length, angle, polygon, perimeter, area	idem + segment measurements, freehand lines, multipoint tool
Character state	linked to the data model (via Xper3)	non
Recording the data	in a CSV exportable table with other annotations on the specimen	in a CSV recordable table
Visualization of measurements	by persistent plotting of all measurements made	current measure only
Descriptive data model	association of the measurement with a model character	no
Statistical parameters during measurement	mean and standard deviation per character of the model	no
Mathematical image processing	no	yes
Kind of images		
Formats	jpg, exif	tiff, gif, jpg, png, dicom, bmp, pmg, fis
Dimensions	2D	2D-5D
Images management		
Multi-image management	projects and tabs	stacks
Images library	library with folders	OS directories
Image set selection	filtrage by tags into Annotate-On library	add to stacks
Tags on images	yes	no
Image metadata	editable model, exif	exif, fixed model
Image display	one window for all the images (carousel)	multi-window, one window per image
Zoom	yes	yes
Image treatments		
	no	yes
Areas of interest		
ROI	core areas of interest (points, rectangles and polygons) to mark or cut out areas for external software	wide variety of possible ROIs for image processing and mathematical analysis
Management by tags		
Tags on annotations	yes	no
Tags on images	yes	no
Managing the researcher's work		
Projects management	yes (an Annotate-On project includes a library, several models and a keyword set)	not relevant
Automation	no	yes
Metadata standard		
Darwin Core	yes	no
Dublin Core	yes	no
Exif	yes	yes
Specimen description		
Characters	yes	no
Model (as a set of characters)	yes	no
Data export		
	CSV	CSV
Links with Xper3	yes	no

Morphometric annotations

The mean and standard deviation of each of the characters are shown at the bottom left of the working screen. Two sets of values are calculated: 1) mean and standard deviation of the measurements of a character on an image; and 2) global value of mean and standard deviation of the measurements of a character on the batch of images (Fig. 1).

Length and multiline length tools

There are two distinct modules. The first one is used as a ruler for linear measurements and the second delimits curves or trajectories, which is fully adapted for non-linear measurements such as certain stems or inflorescences. All length measurements require previously calibrated images (Fig. 2).

Surfaces and areas

The tracing of an area to be calculated is based on the definition of a polygon. The coordinates of each point are exportable and can be used to identify different shapes such as leaf blade standards. Surface and perimeter measurements require calibrated images (Fig. 3). Organ surfaces and perimeters are calculated with the same module.

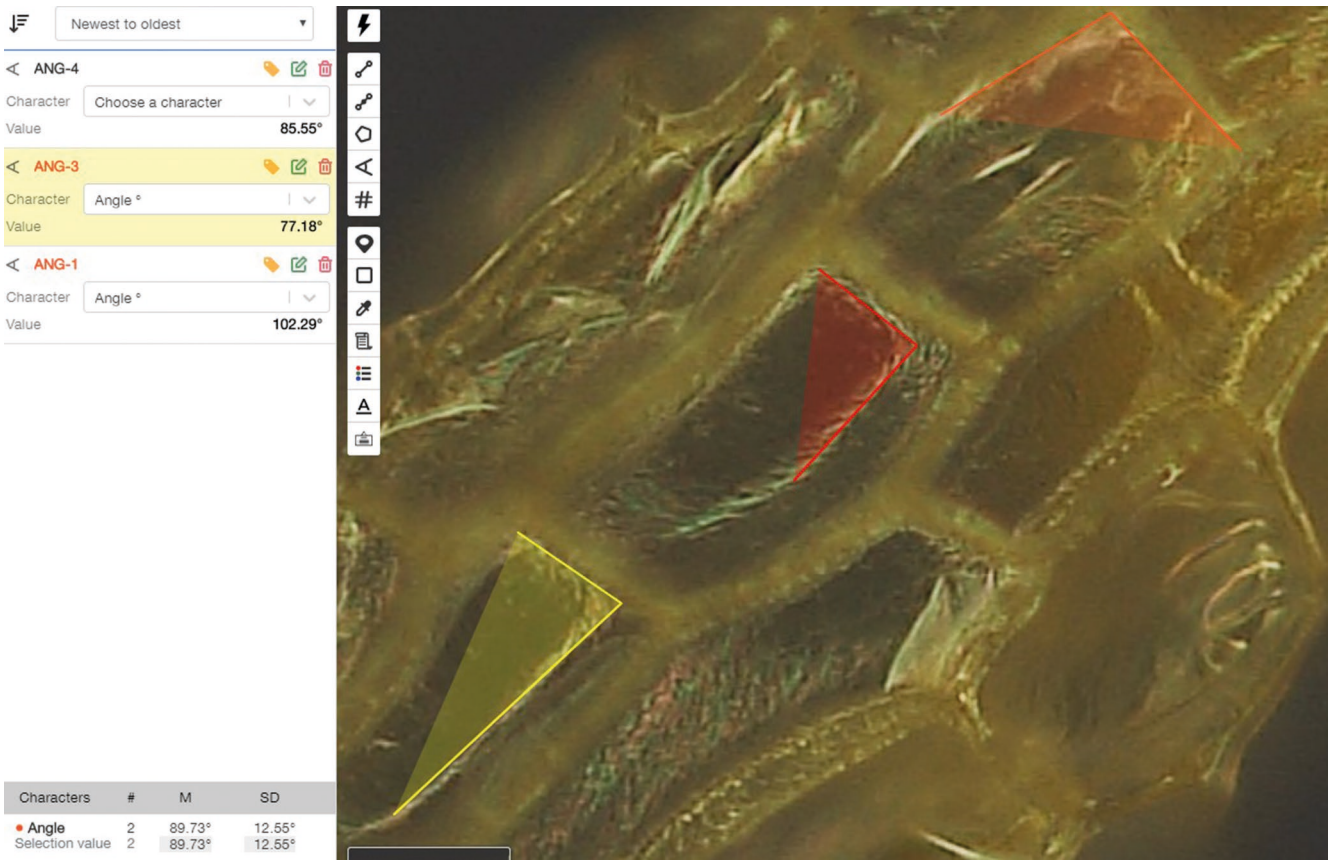

Fig. 4. — Example of angle measurements on an orchid seed from the genus *Phaius*.

Fig. 5. — Example of organ counts: number of rays in an umbel.

Angles

The measurement of angles is useful to accurately describe leaves, seeds or other plant parts. Angle measurement does not require calibration (Fig. 4).

Counting tool

The counting tool allows the counting of objects in an image. Each click adds a cross to avoid counting an object twice. The total number of clicks appears in the left column (Fig. 5).

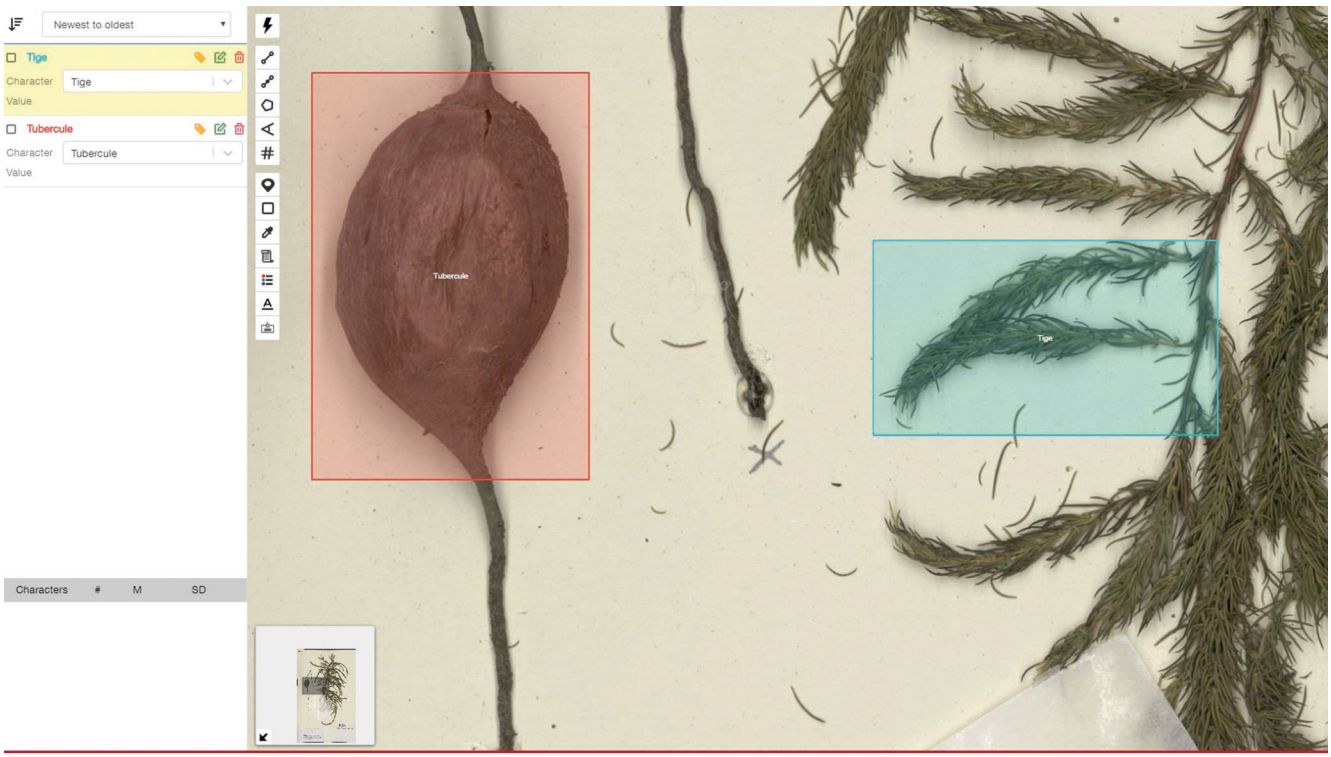

Fig. 6. — Example of the use of rectangle of interest (**ROI**) to show features on a specimen image.

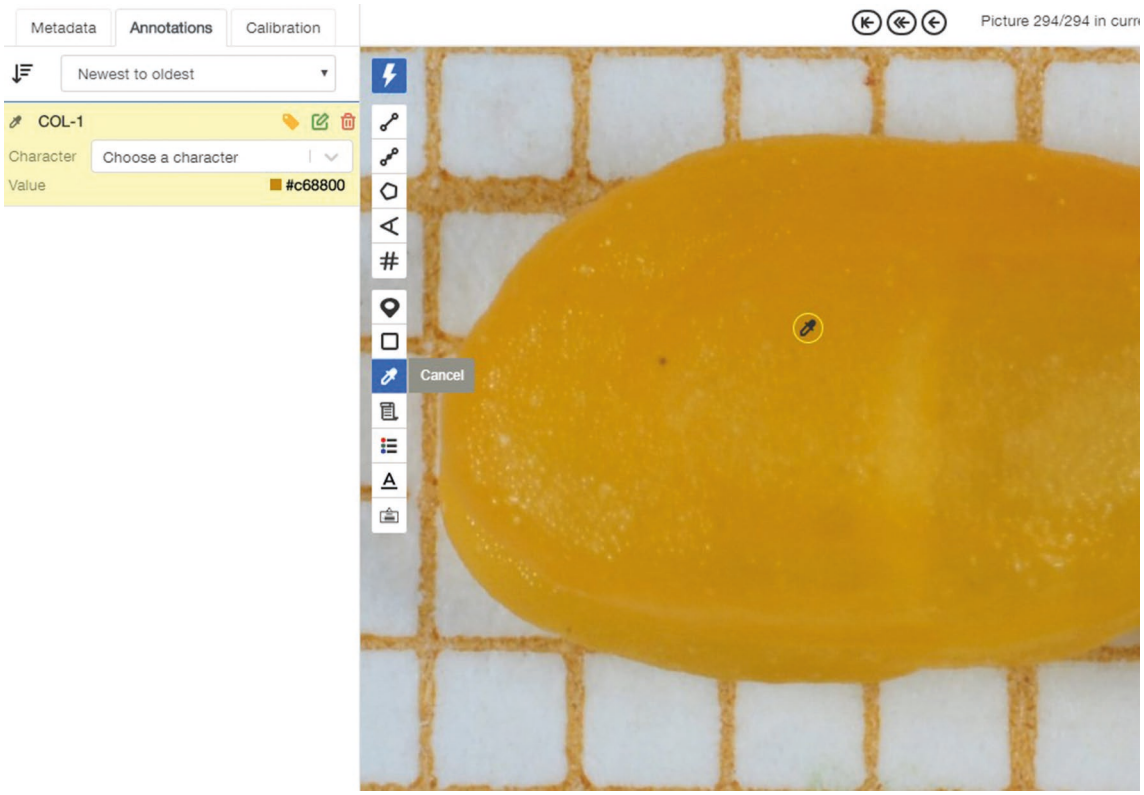

Fig. 7. — Colour tool.

Qualitative annotations

Rectangles (ROI) and points of interest (POI)

The "rectangle of interest" (ROI) and "point of interest" (POI) tools are used to mark a particular area in an image, either to draw attention to or delineate a feature of the specimen, e.g. insect eggs on herbarium sheet, a notable irregularity, etc.,

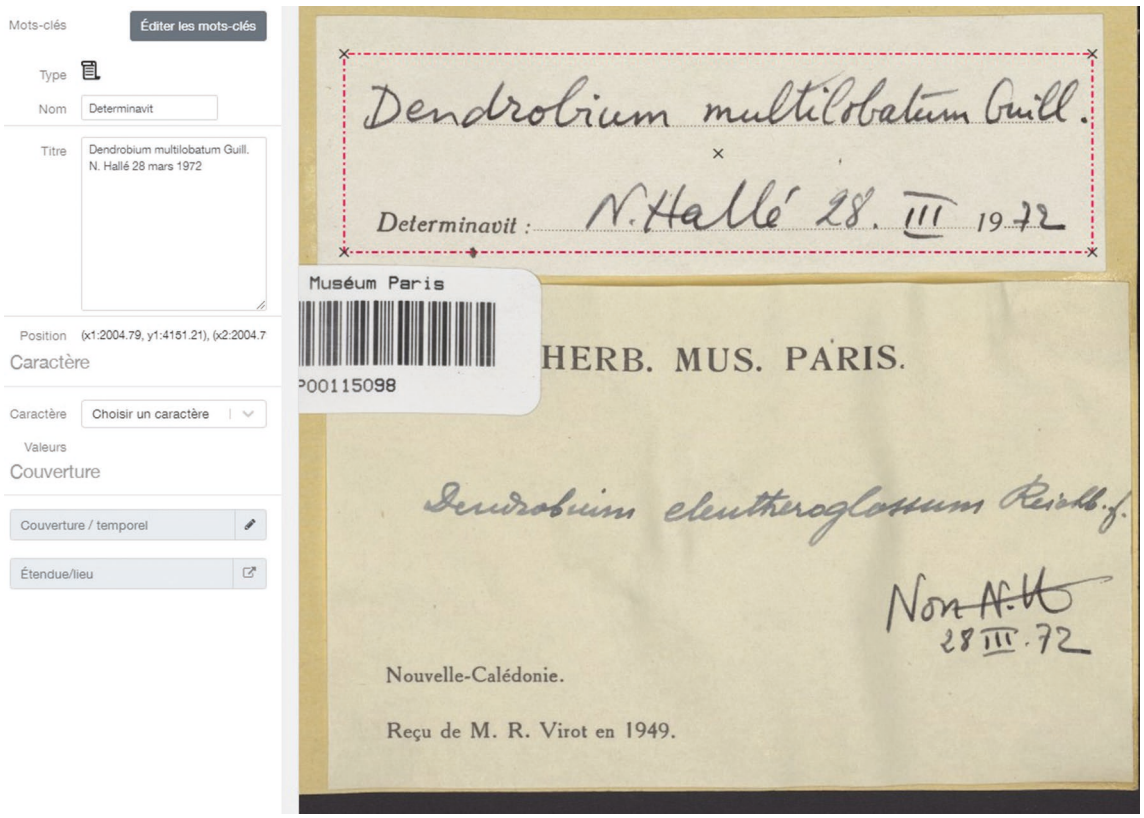

Fig. 8. — Transcription of a written area into a plate.

or to allow it to be used in third-party software (Fig. 6). It can also be used to delineate functional traits. Regarding the ROI, the image area delimited by the rectangle can be automatically exported as a thumbnail image in JPEG format, together with its coordinates in a CSV file. For example, the extracted image parts can be used to initiate the learning phase of a deep learning algorithm.

Colour tool

The eyedropper of this tool allows the RGB colour (hexadecimal Red-Green-Blue code) of an area of the image to be taken from an annotation (Fig. 7). Colour is not frequently used in the study of herbarium specimens, but different shades can eventually serve as a discriminating character.

Transcription tool

This tool makes it possible to transcribe texts exhibited on an image, for example, on labels. It is then possible to use these transcriptions in external software via a CSV spreadsheet export (Fig. 8).

Categorical tool

This tool is adapted for the models produced by the Xper3 tool (https://www.xper3.fr) (Vignes Lebbe *et al.* 2016; Kerner *et al.* 2021). It incorporates characters and associated character states, e.g. "rounded" to describe the apex of a leaf (Fig. 9).

The Xper3 descriptive model offers a controlled terminology. A small number of terminology models are already available, but this list will expand rapidly. A user can annotate with a specific term not included in a previous template. In our context, a character refers to a structure and a property of that structure (e.g. shape of the apex leaf). A value is assigned to the different states of a character.

Observations can then be exported to Xper3 for further processing. As an example, one could refer to the following study: *Architectural Pattern: Study of orchid architecture using tools to take quick measurements of virtual specimens* (Gourraud *et al.* 2021).

Text tool

This tool allows you to add text to the image. The image can then be exported as a JPEG file (Fig. 10).

Label tool

This tool offers the ability to create a label for each image of a set of images. They can then be printed for an exhibition, a book or a website (Fig. 11).

Typical workflow using Annotate-On

An image annotation project starts with the compilation of a dataset of images in the Annotate-On library. The next step is to define a data model by specifying the nature of the organ being measured, e.g. width of a leaf, and the type of annotation to be performed, e.g. measurement of a length, an angle, a point of interest, etc. The last step consists of exporting data to an external processing software to exploit the measurements. Figure 12 summarises the process.

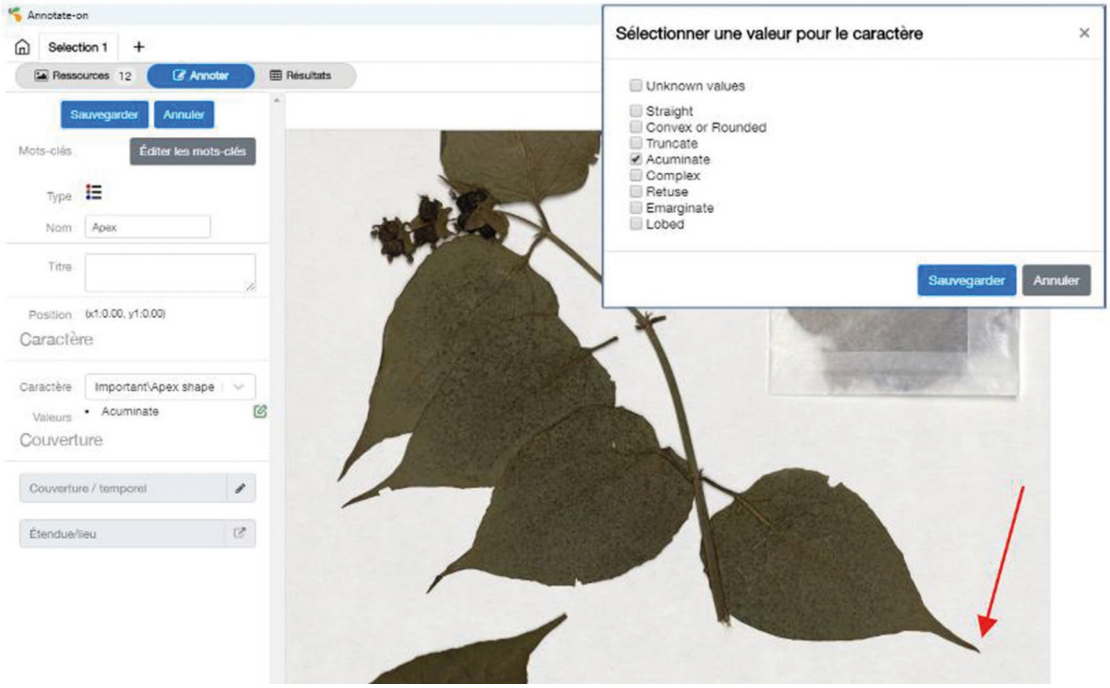

Fig. 9. — Annotation of a categorical character "apex shape" with the character state or nominal value "accuminate" (**red arrow**).

The library can handle hundreds of images. Importation can be done from various sources including the Récolnat database (https://explore.recolnat.org/), URL lists or from files via any hard drive, camera images, microscopes, etc. For very large sets of images, the documentary tagging functionality allows the main batch to be divided into specific subsets, which can be viewed in separate tabs. All of this work is saved as a project via the Annotate-On project management tool for reuse or to make it accessible for other users via a shared cloud directory.

The workflow can be organised as follows:

1) Download the application
Annotate-On is available by download (https://www.recolnat.org/en/annotate) from the French National Network of Naturalist Collection e-services platform (Récolnat 2022).

2) Data model
The data model allows measures to be assigned by character type. It is defined initially and can be completed during the study. A new character must be integrated into a previously created group, e.g. leaf, inflorescence, etc., and the type of measurement must be specified, e.g. physical, enumeration, interest, text, etc. In the case of physical measurements, users are asked to specify whether the measurement is a length, area, angle or a simple enumeration. The mean and standard deviation can be displayed while the measurement is being taken (Fig. 13).

3) Images
Select a set of images from either JSON file produced from the Récolnat database (or directly from the Annotate-On application), upload from a URL or any hard drive using a drag and drop functionality (Fig. 14).

4) Image calibration
Calibration is often included in the Exif metadata, even though some images may have Exif errors. Therefore, it is recommended to calibrate the images manually before starting a series of measurements (Fig. 15).

5) Annotations
In this module, each image in the selection is examined. On each one, measurements and annotations are made, associating them with the characters defined in the model (Fig. 16). The real-time standard deviation of the mean wizard allows comparison of a new measurement in relation to others (on an image or a batch of images).

6) Data exportation
Measurements and annotations accumulated on the selected image can be sorted and exported to a spreadsheet in CSV format, or to Xper3 via an XML exchange file using SDD format (https://www.tdwg.org/standards/sdd/) (Fig. 17). Via the SDD export, the descriptions of the specimens (annotations + measurements) can be reused in Xper3 to be compared between them, merged to form the description of a species by generalisation etc. (Hays & Kerner 2020).

Coordinates of the areas of interest can also be exported and can be used to prepare datasets for training of artificial intelligence algorithms.

Technical description

Récolnat Annotate-On is a cross-platform desktop application that can be installed and executed on Microsoft Windows, Mac OS X and Linux Ubuntu operating systems. The initial choice of such an architecture in the era of the full web is

Fig. 10. — Adding text to an image.

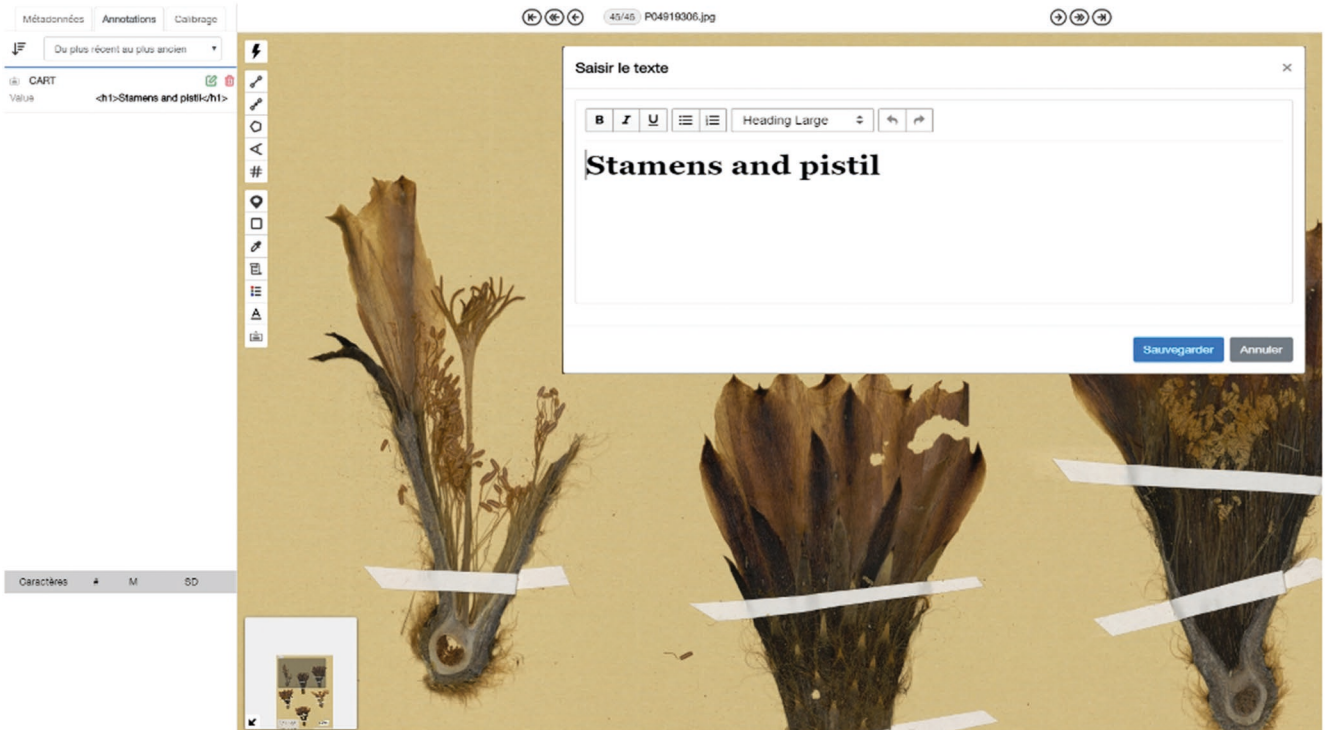

Fig. 11. — Adding a cartel label on an image.

driven by the impossibility of fluidly manipulating several hundred 300 dpi images simultaneously within a browser, and the need to use the tool in the field without an internet connection. However, the development of Annotate-On using web languages (JavaScript), the availability of its code as open source and the growing access to naturalist databases via IIIF mean that it can be extended to a full web solution in the future.

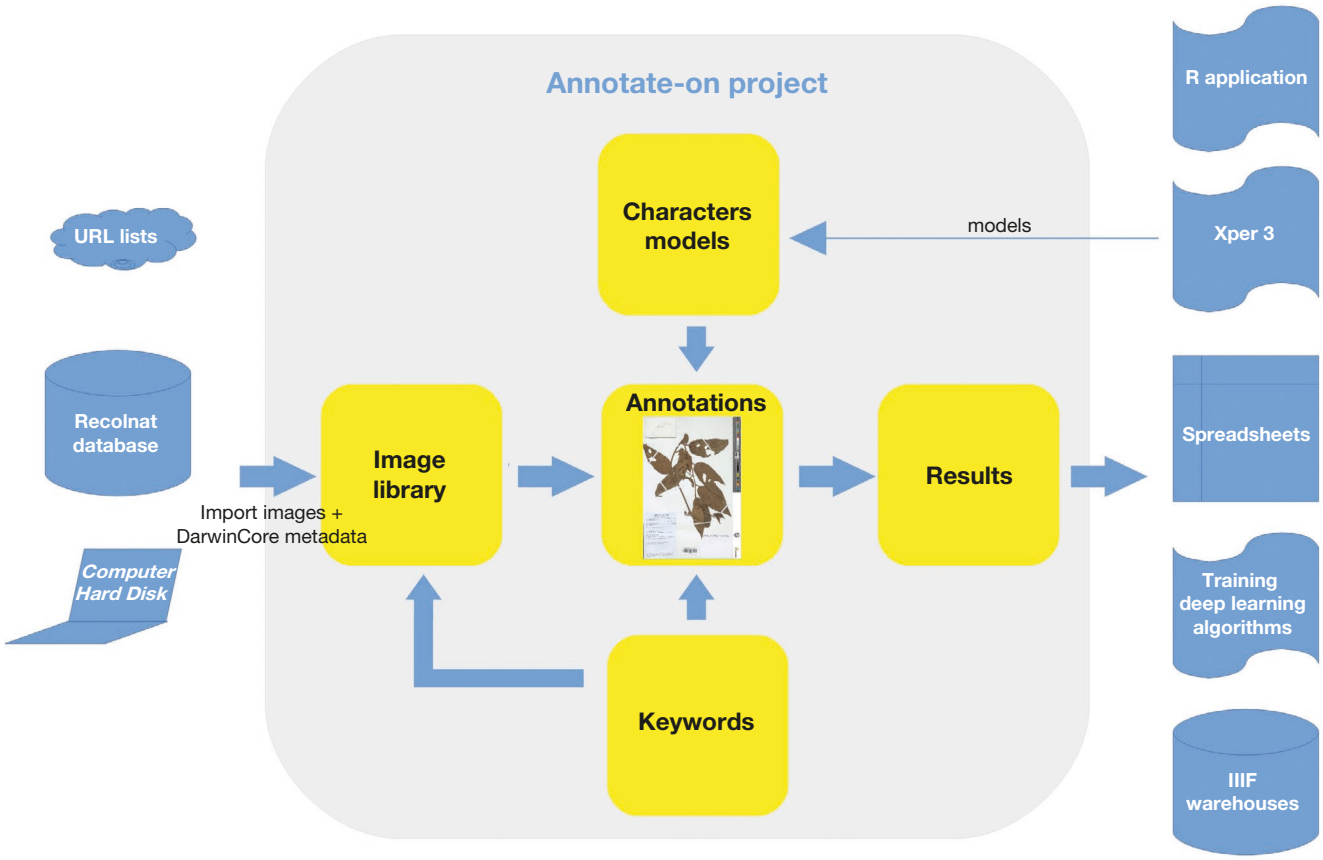

Fig. 12. — Workflow of the Annotate-On tool. A project allows the study, annotation and measurement of objects within images from the Récolnat database, URLs or any hard-drive to constitute a library. It uses one or more of the data models defined previously. Key words are used to select subsets within the library and the results can be exported to various external data processing tools.

Technically, Récolnat Annotate-On is a JavaScript (JS) application based on open-source libraries and frameworks. The main frameworks used are:

Electron (https://www.electronjs.org/), a framework for developing cross-platform desktop applications with web technologies;

React (https://reactjs.org/), a JS library for developing the user interface;

Leaflet (https://leafletjs.com/), a JS library used to manipulate images;

FFmpeg library (https://ffmpeg.org/) for reading and editing video, audio and other multimedia files and streams for non-botanical purposes, especially in digital humanities.

Open code source

Récolnat Annotate-On was initially developed by Dicen-IDF lab engineers and is currently subcontracted to the software company Presek Informatique. An Annotate-On forge in open access under GPL license was created on GitHub in 2021 at https://github.com/Annotate-On (Bertin *et al.* 2021). This open source code opens up the possibility of additional measurement tools, new processing, data import and export and interconnections with other software. The interface can also be modified in this repository to enable any language community to use Annotate-On in its own language (currently English, French, Portuguese and Spanish).

Récolnat Annotate-On is open: a) upstream to images from all sources;

and b) downstream by exporting annotations and user metadata in appropriate formats (JSON, CSV, RDF, XMP) and standards such as IIIF collections. The export of thumbnails extracted from specific image areas greatly facilitates the preparation of the learning phase of deep learning algorithms.

Maintainability of the tool

Towards algorithms and artificial intelligence

The tool is part of the software package developed by the members of the Récolnat infrastructure consortium and the e-COL+ programme (ANR21_ESRE-0053). It is therefore intended to expand its functionalities and is constantly being improved with a view to enabling 3D management in the coming years. In addition, all these sources are freely available on GitHub so that any research team can correct bugs, extend the software's functionality or even generate a separate version for their specific needs.

Towards digital humanities

Researchers in social sciences and humanities are more and more often called upon to analyse heterogenous corpuses of photos and videos due to their abundance and their benefits to science and heritage. As a result, they increasingly require

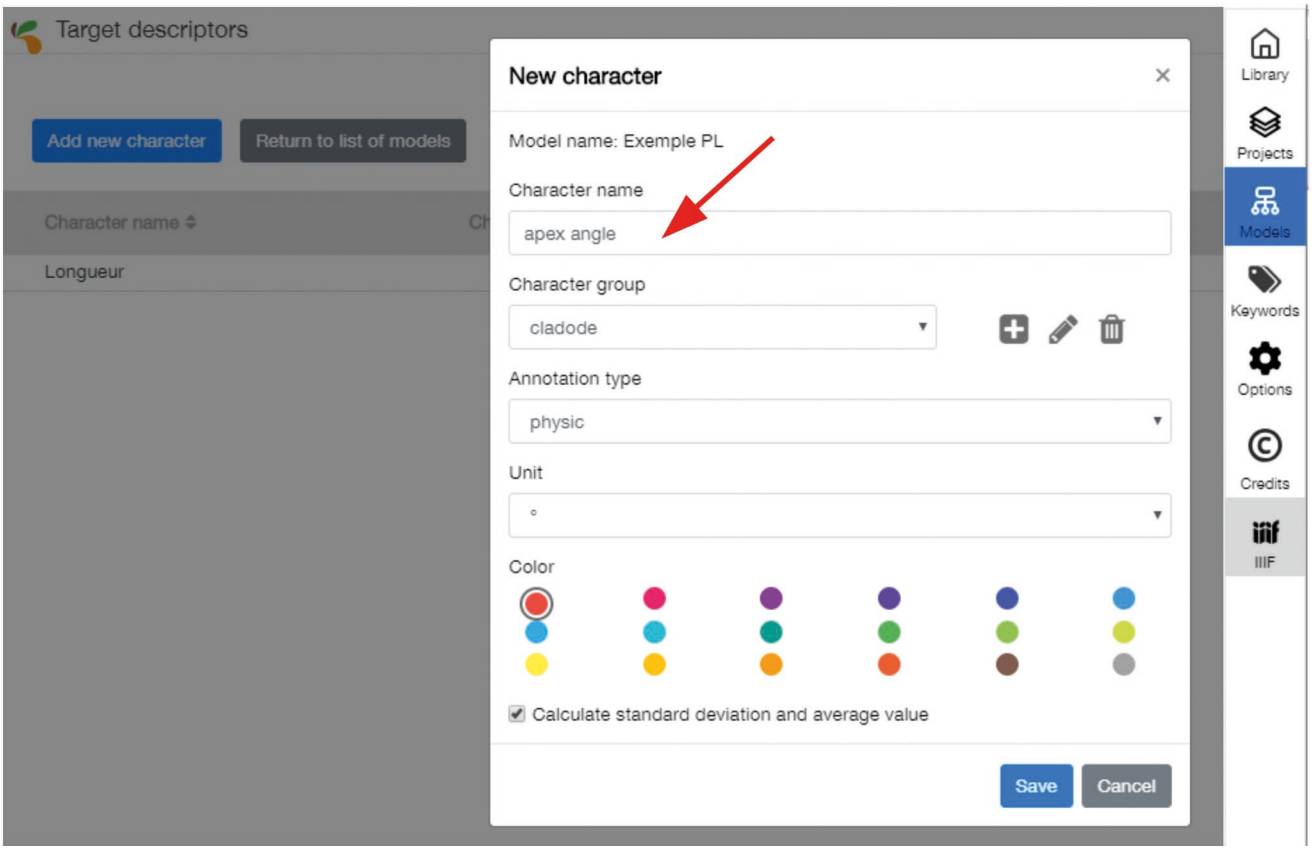

Fig. 13. — Definition of a new character ("apex angle") for the model "Asparagus", here a physical measurement (angle unity: "°") into the group named "Cladode".

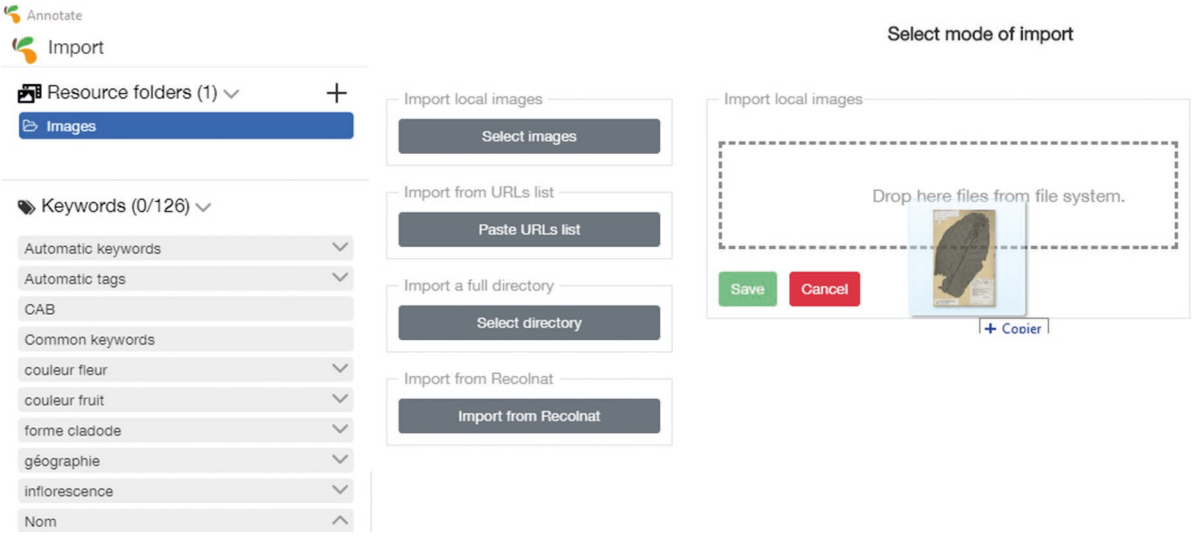

Fig. 14. — One of the three modes of importing images, drag and drop from a computer hard drive.

training and access to analytic data in order to perfect algorithms and data visualisation tools. Today, analysing these types of corpuses remains challenging.

Récolnat Annotate-On has been expanded beyond botany into digital humanities within the framework of OPAHH IIIF (Bertin 2022). It has been extended to include the chrono-thematic annotation of videos and live events, broadening its use to oral and animated archives and their constitution within the framework of OPAHH. This project is specifically dedicated to open annotations within the field of digital humanities. Analyses carried out with Annotate-On can combine all types of annotations including physical measurements, zones of interest, chrono-thematic annotations from heterogeneous datasets from fixed images and animated and live events.

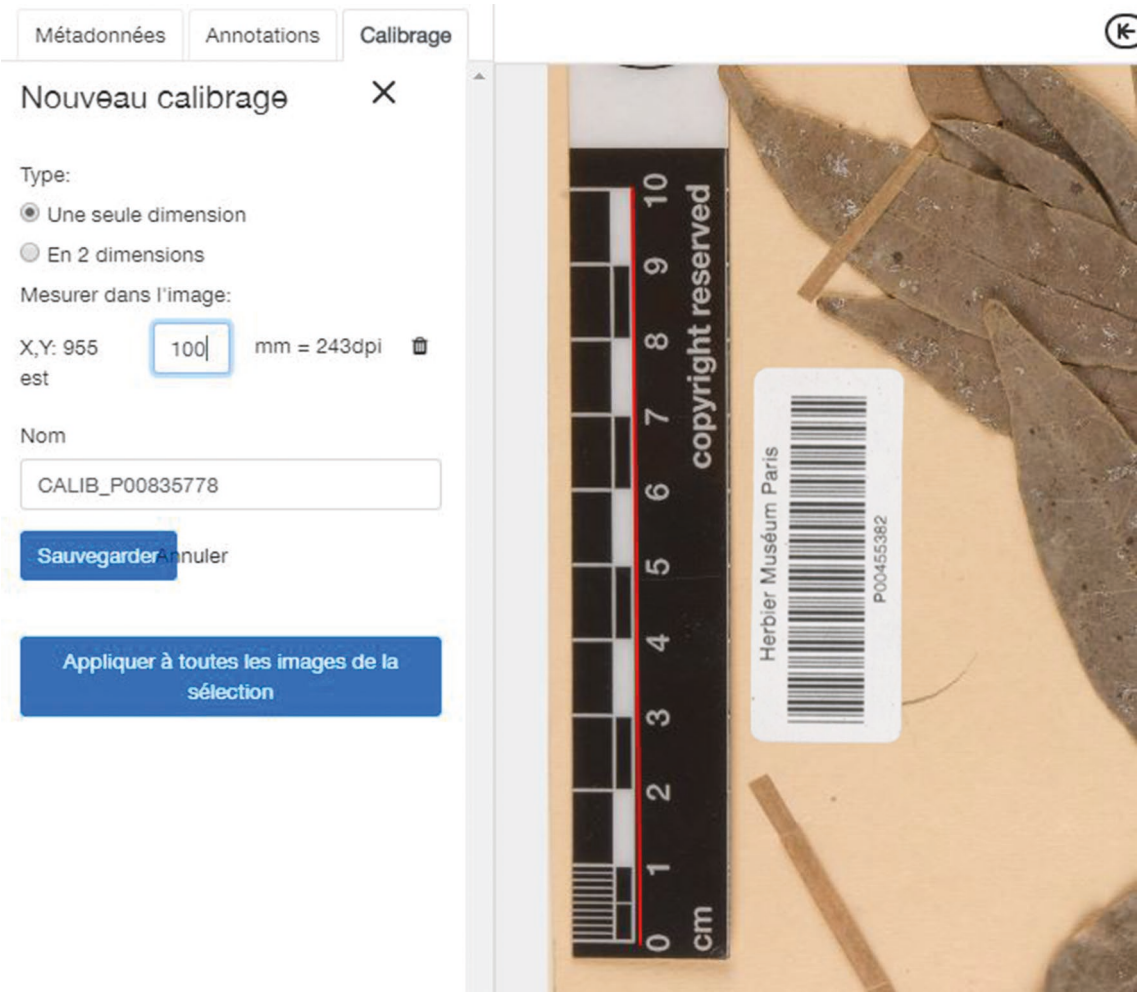

Fig. 15. — Scales on Paris herbarium sheets. The vertical line on the left measures 10 mm, on the right the label of the linear barcodes measures 50 mm in length. In addition, the photos of the nomenclatural types have a scale.

DISCUSSION

Récolnat Annotate-On is an original tool, developed and adapted specifically for remote work on the digital twins of 2D specimens in natural history collections. In particular, it allows an intensive use of measurements and annotation for the study of herbarium specimens. This intensity allows for considerable time-saving in the study of numerically important sets characteristic of naturalist collections. A botanist who spends a week taking 500 measurements on botanical specimens can now execute the same in half a day using this tool. Moreover, these measurements can be made on virtual duplicates of collections geographically distributed in a multitude of institutions. This ability to explore, simulate and validate experiments and hypotheses in a short period of time is perfectly suited to the writing of scientific articles, collaborations between researchers, taxonomic revisions, etc. The Annotate-On tool is suitable for one-off measurements or as part of a descriptive model, and may also be used for other applications where users can devise more personal protocols.

The strengths of Annotate-On are its ergonomics and flexibility of use. Its design means that physical measurements can be performed rapidly and in large numbers. It can thus meet the needs of a researcher who wishes to accumulate and process a large corpus of images, as well as those of rapid measurements and/or annotations. Within each project, image imports are flexible and easily accessed from the database of the Récolnat infrastructure, via a URL or a hard drive. In addition, the data models proposed allow direct and rapid exploitation. The ergonomics extend to the rapid constitution of a data model, which allows it to be easily used or adapted in the case of a taxonomic revision in botany, according to particular morphological characters. These models can be shared between several different projects and are, as such, exchangeable with other researchers and students.

Xper3 is an online collaborative tool. This property allows for the editing of new models in an open and consensual way. It also allows for working collaboratively on descriptions exported from Annotate-on to Xper3.

All data produced with Annotate-On is easily exportable. Annotations with their models and metadata can easily be

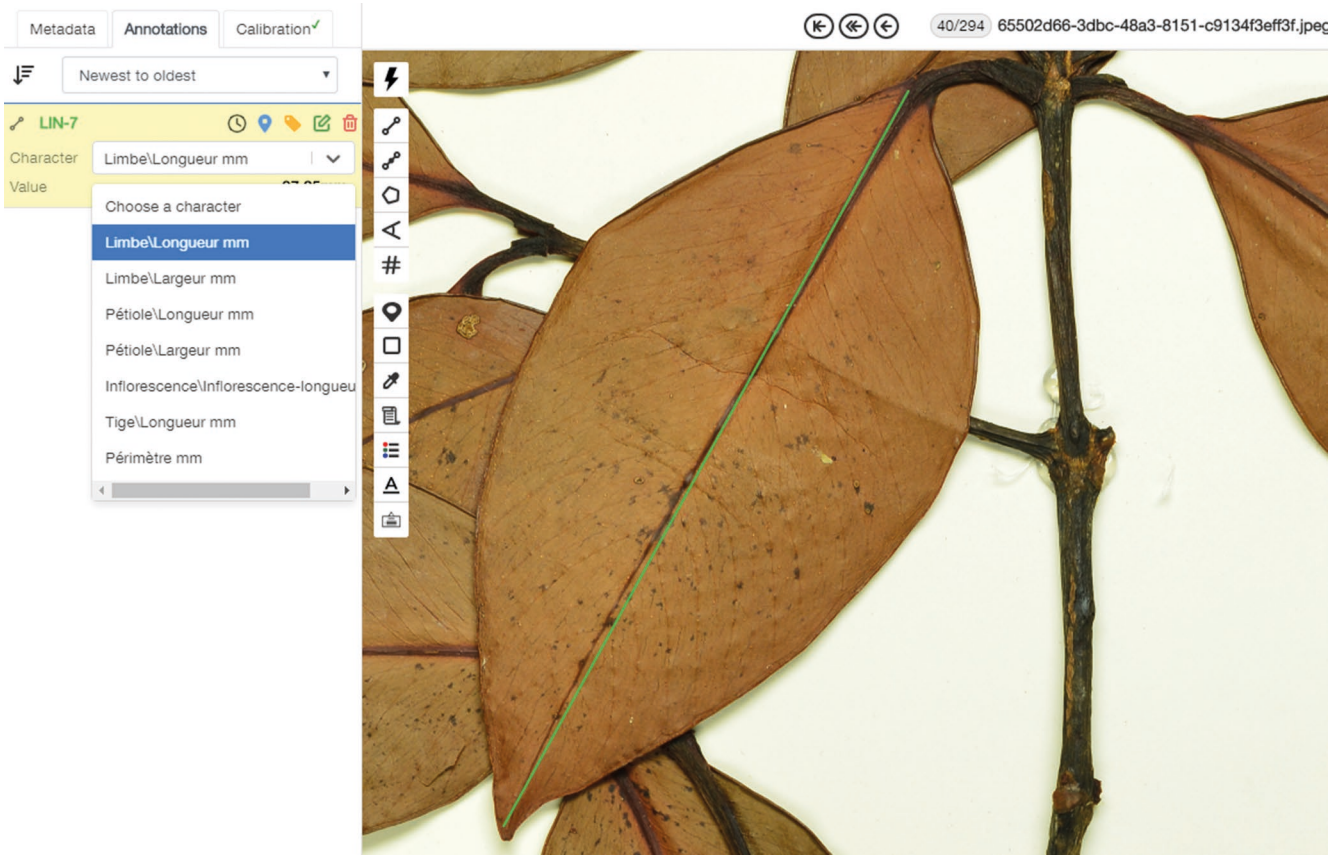

FIG. 16. — Character attribution.

sent in CSV format to third-party tools and software for image analysis, mathematical and statistical data processing or visualisation. Eventually, this tool will contribute to the production of data related to machine learning. The projects carried out in Annotate-On are already exportable to IIIF format to enrich the image databases of institutions open to this new international exchange standard. This opens up an important field of international collaboration between teams of researchers on major societal issues. These collaborations are made easier because Annotate-On is freely available on the Récolnat website (https://www.recolnat.org/en/annotate) and the code is deposited on GitHub (Bertin *et al.* 2021). It has the capacity to evolve to integrate new measurement and processing tools and export formats in order to adapt to the emergence of new requirements and new disciplines.

Researchers have already used the tool to adapt it to their own field of research, such as Hays & Kerner (2020), who have used it to study archaeocyaths. The development of the software continues in the direction of other natural science disciplines. Qualitative annotations open up uses for analysis, reflection and preparation for writing in different disciplinary fields. In addition, an application to images of 2D specimens outside of botany could bring interesting results, notably in the field of morphometry or comparative anatomy. Studies on digital twin specimens in palaeontology or entomology could be performed to test its relevance for these disciplines. Furthermore, an experiment on 3D specimens will take place within the framework of the e-COL+ programme. This transdisciplinary application is already well established outside the natural sciences in the field of digital humanities.

At a time when the concept of "extended specimens" is emerging, Annotate-On will find a place in the ecosystem of shared, accessible and interoperable data, bringing significant added value to physical specimens within naturalist collections.

Acknowledgements

Annotate-On has been funded and developed as part of the e-ReColNat program (ANR-11-INBS-0004) within the framework of the French National Network of Naturalist Collections Research Infrastructure (Récolnat RI). The authors are particularly grateful to Germinal Rouhan, Nicholas Wilding, Jérôme Munzinger and Michel Guiraud and for their valuable comments, advice and feedback at different stages during the process. We thank Christian Laudereau for permission to reproduce the image in Figure 3. We also thank the three anonymous referees for their valuable and insightful comments that refined the quality of this manuscript. Last but not least, we are kindly indebted to Charlotte Gillbanks for improving the English text.

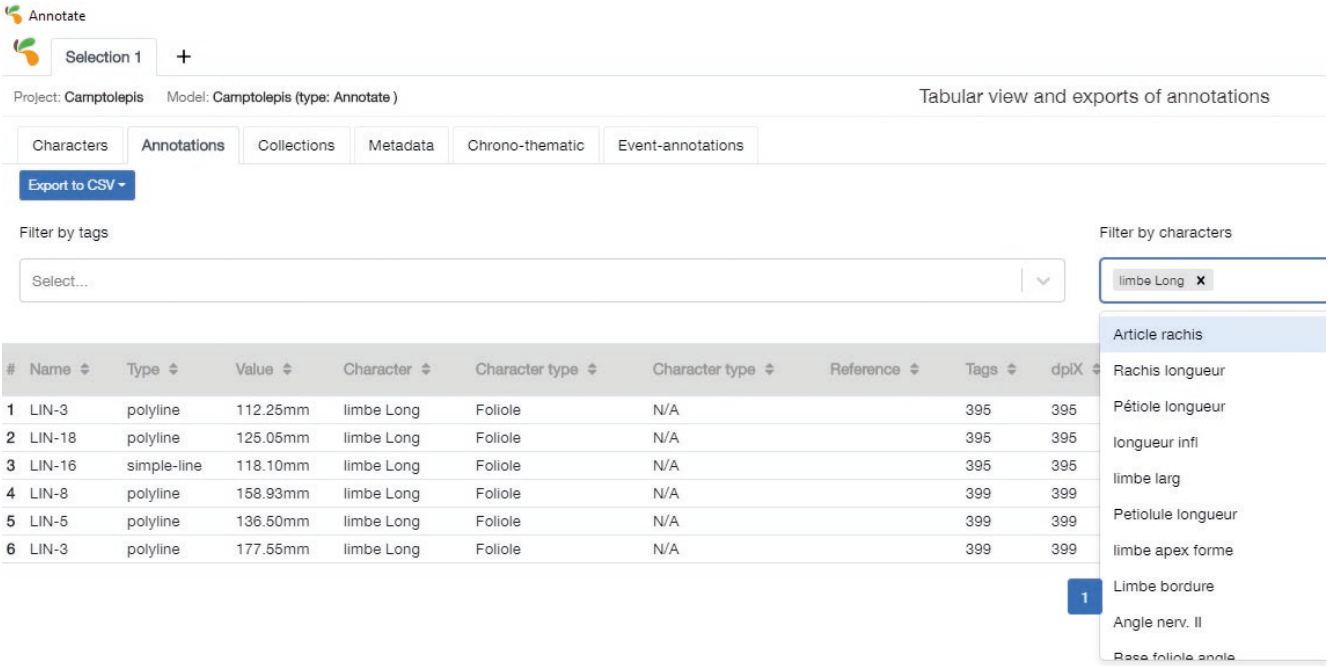

Fig. 17. — Data export.

REFERENCES

Barkworth M. & Murrell Z. 2012. — The US Virtual Herbarium: working with individual herbaria to build a national resource. *ZooKeys* 209: 55-73. https://doi.org/10.3897/zookeys.209.3205.

Bertin G. 2022 — Open Pictures Annotator for Heritage and the Humanities with IIIF OPAAH iiif. [Retrieved from http://passes-present.eu/en/open-pictures-annotator-heritage-and-humanities-iiif-opaah-iiif-44487].

Bertin G., Pignal M., Klasnja S., Bottini T., Scopsi C., Zacklad M., Perez E.., Dusoulier F., Chupin L., Rouhan G. & Le Gall L. 2021. — Annotate-On. *Github.* https://github.com/Annotate-On

Brazil Flora Group 2018. — Brazilian Flora 2020: Innovation and collaboration to meet Target 1 of the Global Strategy for Plant Conservation (GSPC). *Rodriguésia* 69 (4): 1513–1527. https://doi.org/10.1590/2175-7860201869402.

Coughenour C. 2018. — Inside Brazil's National Museum on Google Arts & Culture [Retrieved from https://g.co/arts/NcPQNoqbRyvesDSx5, accessed 2019-09-17].

Chupin L. 2015. — Enjeux de la numérisation des herbiers pour l'information et la communication scientifiques : de la transformation des matières documentaires à l'évolution des pratiques. *Les Enjeux de l'Information et de la Communication* 16 (2): 69-82. https://doi.org/10.3917/enic.019.0069

Chupin L. 2017. — *Enjeux communicationnels de la conception de dispositifs de médiation documentaire augmentée pour les herbiers numérisés*. Thèse de doctorat en Sciences de l'information et de la communication. Conservatoire national des arts et métiers, Paris, 220 p.

Dusoulier F. 2022. — *The DiSSCo consortium in France*. [Retrieved from https://www.dissco.eu/fr/, accessed 6 May 2022].

Ferreira T. & Rasband W. 2012. — ImageJ User Guide IJ 1.46r, 187 p. http://imagej.nih.gov/ij/docs/guide

Gourraud A., Vignes Lebbe R., Kerner A. & Pignal M. 2021. — Architectural Pattern: Study of orchid architecture using tools to take quick measurements of virtual specimens. *Biodiversity Information Science and Standards* 5: e75752. https://doi.org/10.3897/biss.5.75752

Grimé W. E. & Plowman T. 1987. — Type photographs at the Field Museum of Natural History. *Taxon* 36 (2): 425-426. https://doi.org/10.2307/1221436

Hays L.-M. & Kerner A. 2020. — Virtual Systematics with Annotate: The Mexican archaeocyaths example. *Biodiversity Information Science and Standards* 4: e59106. https://doi.org/10.3897/biss.4.59106

Hussein B. R., Malik O. A., Ong W.-H. & Slik J. W. F. 2022. — Applications of computer vision and machine learning techniques for digitized herbarium specimens: A systematic literature review. *Ecological Informatics*: 101641. https://doi.org/10.1016/j.ecoinf.2022.101641

Kerner A., Bouquin S., Portier R. & Vignes Lebbe R. 2021. — The 8 Years of Existence of Xper3: State of the art and future developments of the platform. *Biodiversity Information Science and Standards* 5: e74250. https://doi.org/10.3897/biss.5.74250

Le Bras G., Pignal M., Jeanson M. L., Muller S., Aupic C., Carré B., Flament G., Gaudeul M., Gonçalves C., Invernón V. R., Jabbour F., Lerat E., Lowry II P. P., Offroy B., Pimparé E. P., Poncy O., Rouhan G. & Haevermans T. 2017. — The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. *Scientific Data* 4: 170016. https://doi.org/10.1038/sdata.2017.16

Merrill E. D. 1943. — Destruction of the Berlin Herbarium. *Science* 98 (2553): 490-491. https://doi.org/10.1126/science.98.2553.490

OPAHH-IIIF 2022. — Open Pictures Annotator for Heritage and Humanities with IIIF, Labex "Les passés dans le présent", [accessed 2022-05-02].

Pignal M. & Pérez E. 2013. — Numériser et promouvoir les collections d'histoire naturelle. *Bulletin des bibliothèques de France (BBF)* 5: 27-30.

Pignal M., Romaniuc-Neto S., Souza S. D., Chagnoux S. & Canhos D. A. L. 2013. — Saint-Hilaire virtual herbarium, a new upgradeable tool to study Brazilian botany. *Adansonia,* sér. 3, 35 (1): 7-18. https://doi.org/10.5252/a2013n1a1

Récolnat 2022. — Récolnat-ANR-11-INBS-0004 [Le réseau national des collections naturalistes]. https://www.recolnat.org/ (accessed 01.03.2024).

Vignes Lebbe R., Chesselet P. & Diep Thi M. H. 2016. — *Xper3: New Tools for Collaborating, Training and Transmitting Knowledge on Botanical Phenotypes*. UNESCO, Paris: 228-239. (Botanists of the twenty-first century: roles, challenges and opportunities).

Zhu Y., Durand T., Chenin E., Pignal M., Gallinari P. & Vignes-Lebbe R. 2017. — Using a deep convolutional neural network for extracting morphological traits from herbarium images. *Biodiversity Information Science and Standards* 1: e20400. https://doi.org/10.3897/tdwgproceedings.1.20400

Submitted on 2 September 2022;
accepted on 26 February 2024;
published on 26 August 2024.