Digitisation of the LUX herbarium collection of the National Museum of Natural History Luxembourg

Thierry Helminger[1], Odile Weber[1] & Paul Braun[1]

[1]Musée national d'histoire naturelle, 25 rue Munster, L–2160 Luxembourg (thierry.helminger@mnhn.lu; odile.weber@mnhn.lu, corresponding author; paul.braun@mnhn.lu)

Helminger, T., O. Weber & P. Braun, 2020. Digitisation of the LUX herbarium collection of the National Museum of Natural History Luxembourg. *Bulletin de la Société des naturalistes luxembourgeois* 122 : 147-152.

Published online 15 July 2020 (ISSN 2716-750X), printed volume 7 December 2020 (ISSN 0304-9620).

Abstract. The LUX herbarium of the National Museum of Natural History is estimated to hold 100.000 specimens documenting the plant and fungal diversity mainly of Luxembourg. In 2019 two batches of herbarium specimens were sent to Picturae for imaging, including several historic herbaria in bound volumes. The resulting 67.116 specimen images give unprecedented access to the herbarium collections to researchers and the general public via the MnhnL Species Observation Database (https://map.mnhn.lu/) and the Global Biodiversity Information Facility (gbif.org). Additional work is needed to maximise the online use of the herbarium collection, including the processing of specimen backlog, data cleaning and geo-referencing.

1. Introduction

The digitisation of natural history collections has taken momentum in the past two decades due to advances in information technology, availability of dedicated funding and the perception that collections are an invaluable source of data for biodiversity research (Meineke et al. 2018). An overview of digitisation projects and techniques is provided by Smith & Blagoderov (2012) in a special issue of the digital journal Zookeys. Access to herbarium specimens is a primary aim of many digitisation initiatives (Nic Lughadha & Miller 2009). A digital herbarium freely accessible online is an unparalleled tool for the study of plants and fungi by specialists and amateurs alike. Many diagnostic characters are visible on good quality images (high resolution, sharp, large depth of field). A reference collection of specimens determined by experts is a helpful resource to help identification. Some taxonomic groups can only be named by comparing specimens. Consulting the digital image of a specimen reduces the number of times the physical specimen needs to be handled and consequently the risks of degradation during these manipulations. The LUX herbarium is the national reference collection for the plant and fungal species of Luxembourg (Colling 2005). In June 2018 the Luxembourg government presented its digital strategy for the national cultural heritage with the focus on making digital collections available online (Ministry of Culture 2018).

2. Materials and methods

2.1. The LUX herbarium

The herbarium of the National museum of natural history in Luxembourg (Musée national d'histoire naturelle Luxembourg, MnhnL), *Index Herbariorum* code LUX, holds an estimated 100.000 plant and fungi specimens to date. The herbarium was started in the middle of the 19th century. The collection is made up of about 50.000 specimens (collection events) of vascular plants glued to sheets of paper. The oldest specimens date back to the early 1800s. Many specimens are from the 1850s and 1860s, the period between 1880 and 1905 and the middle of the 20th century. Heuertz

(1951) gives a brief history of the herbarium collection, which he estimated at that time to hold some 15.000 specimens. There was a peak in collecting activity in Luxembourg from the late 1940s to the early 1960s by local botanists Léopold Reichling, Marcel Etringer, François Léon Lefort and Joseph Witry. The geographic remit of the collection covers predominantly the Grand Duchy of Luxembourg, as well as neighbouring countries. Notable collections from further afield include those of Jean-Marie Mangen (± 900 specimens) documenting the flora of Mount Trikora in New Guinea, Michel Couteaux (± 1.200 specimens) from Tunisia and plants from the Carpathian Mountains collected by Hugo Ilse (± 1.000 specimens).

The main herbarium collection is curated in alphabetical order by family, genus and species within the major systematic groups starting with non-monocot angiosperm and followed by monocots, gymnosperms and ferns & fern allies. A number of herbarium sheet collections are curated separately from the main herbarium, notably those of J.-M. Mangen, M. Couteaux, the Herbier Frère Macedonne and the "Flora Europaea" herbarium of Jules Garnier deposited at LUX by the monastery of Clervaux in December 2016, the "Herbier du Grand-Duché de Luxembourg" assembled by Jean-Pierre Joseph Koltz in 1883 (14 volumes + index) and the Herbier François Auguste Tinant (27 volumes + index).

The non-sheet collections comprise specimens of fungi, lichens, bryophytes and algae. They are prepared in capsules fixed to half-sized herbarium sheets or envelopes and stacked in special boxes.

2.2. Digitisation of the LUX herbarium

Digital processing of the LUX specimen data started in the early 1990s when specimen data from the herbarium index cards was entered into the database SPECIES developed at the museum together with Jim Meisch, curator of the botany department. Data captured from a specimen includes the herbarium specimen number, species name, determiner and determination date, collector(s), collector's number if present, collecting date and locality information (descriptive text and geographical coor-

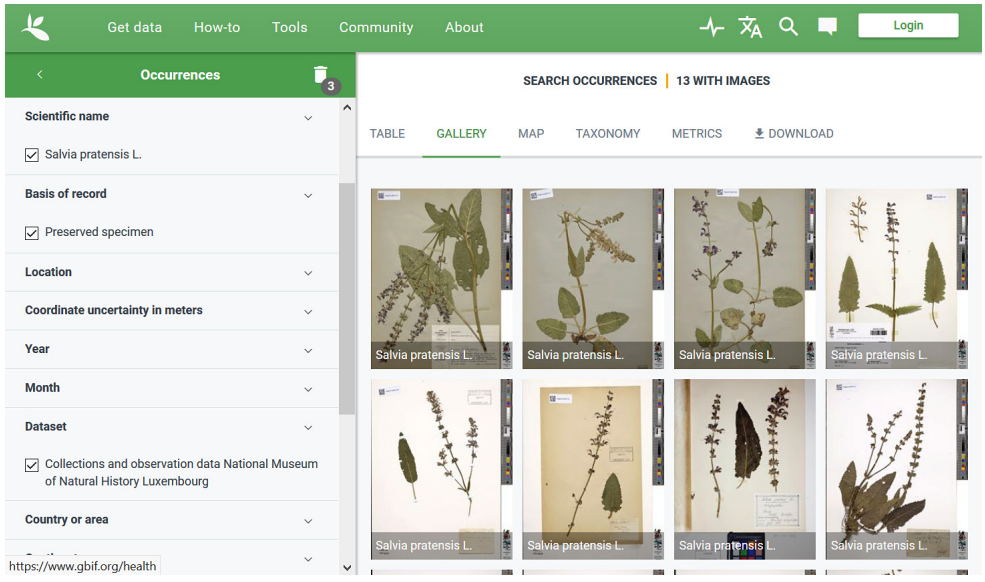


Fig 1. Screenshot of the result of the search for preserved specimens of *Salvia pratensis* L. in the dataset "Collections and observation data National Museum of Natural History Luxembourg", displayed on the Global Biodiversity Information Facility portal (gbif.org) in the gallery menu.

dinates or grid reference). The herbarium specimen number is written out on herbarium sheets.

In 1999, this dataset was imported into BG-BASE, a database application designed to manage information on biological (primarily botanical) collections. From there it was transferred into the current institutional database Recorder-Lux by John van Breda on the 9th of Mai 2010. More than 40.000 specimen records from Recorder-Lux are currently accessible and searchable online via the MnhnL Species Observation Database (mdata.mnhn.lu) and the Global Biodiversity Information Facility (www.gbif.org) (Fig. 1). Approximately one third of these were collected in Luxembourg. They include over 4.000 fungi specimens of the collection of Jean Feltgen with many type specimens of species names published by him.

The metadata of several non-sheet collections has been digitised in separate databases and is awaiting import into Recorder-Lux. These include the fungi collections (8.600 specimens) of Céline Besch, Jean Turk, Adolphe Molitor and Ben Schultheis, the lichen collections by collectors such as F. A. Tinant, J. Feltgen, J. B. Reinhard and H. Olivier (± 3.000 specimens, Paul Diederich, personal communication) currently on loan to P. Diederich and the bryophyte collection of Jean Werner donated to LUX in 2017 (5.400 specimens from Luxembourg). Approximately 550 specimens of bryophytes collected by J. Feltgen have recently been databased and redetermined by bryologist Florian Hans.

A unique opportunity, to considerably advance the digitisation of the herbarium collection arose in 2018 as the Ministry of Culture provided funding for an imaging project. This work was carried out in two parts in 2019 and early 2020.

2.3. Preparing the herbarium sheets for imaging

During the pre-imaging curation, the tasks outlined by Nelson et al. (2015) have been applied as far as time allowed. About 5.000 unmounted specimens were mounted on archival paper in the usual way (Bridson & Forman 2004). Many of the mounted specimens required some kind of minor intervention such as gluing down the label, putting fragments into a capsule or, less often, writing out the specimen number on additional sheets of multi-sheet specimens and adding a suffix (usually a letter) to multi-specimen sheets. Eight members of the museum staff and tree volunteers worked for an estimated 700 person hours to prepare the first batch of specimens and packaged them in 506 boxes for shipment. Due to the limited time available it was not possible to update the nomenclature and ensure that the name the specimen is filed under is visible on the sheet. We added the previously missing plant family name on the outside of the species covers, to enable a simple workflow when the specimens were returned to the storage units. We used the opportunity this provided to adapt the alphabetical sequence to the APGIV classification (Lambinon et al. 2015, Trias-Blasi et al. 2015 for monocots).

2.4. Digitalisation by Picturae

Imaging took place in the Netherlands on the premises of the firm Picturae, a company that has acquired a high level of expertise in the digitisation of herbaria in recent years (Heerlien et al. 2015, Le Bras et al. 2017, Engledow et al. 2018). They captured high-resolution images (560 dpi) with a depth of field of several centimetres and high colour conformity of each individual herbarium sheet. A printed label with a unique image ID (as human readable text and encoded as machine readable data-matrix) attached to the sheet before imaging provides the file name and identifier for the image. It was linked to the MnhnL herbarium specimen number transcribed from the image following a previously elaborated transcription protocol (Wijkamp 2019), allowing the association of the image to the corresponding specimen entry in the database Recorder-Lux.

The first batch of specimens sent in February and returned on the 18th of March 2019 consisted of the main herbarium with some 35.000 specimens. The specimens were imaged on a special conveyor belt system engineered by Picturae. A second batch



Fig 2. Image of the isotype specimen of *Papuacalia valentini* Veldkamp (Herbarium LUX Number MNHNL85632) taken by Picturae. The availability of a specimen image online enables quick verification of essential data such as the determination history and notes, locality description or collecting date. This reduces the need to handle the actual specimen. The colour chart and scale bar next to the sheet were photographed with the sheet and aligned with the edge of sheet during image processing. The MNHNL logo was also added during image processing. Additional EXIF data embedded into the image file are: the filename which corresponds to Image ID barcode visible on the image; ImageDescription: Herbarium sheet; Artist: Musée national d'histoire naturelle Luxembourg, Copyright; Attribution 4.0 International (CC BY 4.0); and Keywords: Herbarium LUX, Botany, MnhnL, Luxembourg.



Fig 3. Detail from the image of herbarium specimen MNHNL85632, *Papuacalia valentini*, shown in Figure 2 with a 1 cm scale bar. Field depth and resolution of the images are sufficient to allow magnification up to times six on typical computer displays at 96 dpi.

consisting of recently mounted specimens dating from the 1940s to the present day and a number of bound historical herbaria was dispatched on the 19th November 2019 and returned on the 21st of January 2020. Most of these had to be handled with extreme care and were photographed manually page by page.

3. Results

The 67.116 specimen images (ca. 5 Terabyte of data) resulting from the digitisation project with Picturae are currently stored on the MnhnL server and backed up on external hard-drives. The images have a resolution of up to 560 dpi and have been delivered as compressed JPEG (used for online publication) and a master file in JPEG 2000 file format. Each image depicts the specimen itself with the image ID label and has a scale bar, a colour chart (both captured with each image) and a digital Museum logo attached to it (Fig. 2). Figure 3 shows the level of detail visible on an image at maximum zoom. Additional information has been written into the metadata of each image, this includes the name of the MnhnL, copyright information (Attribution 4.0 International CC BY 4.0) and the descriptive keywords: Herbarium sheet, Herbarium LUX, Botany, MnhnL and Luxembourg.

4. Conclusions and outlook

High quality digital images of 34.251 specimens of the LUX herbarium are currently available online. This greatly enhances the visibility of the collection and represents an invaluable resource for researchers working on biodiversity and taxonomy questions. For the herbarium curation team the images will greatly facilitate many curation tasks such as quality control of the label information in the database. Feedback by users of the images already gave helpful hints to correct spelling or other errors in the datasets.

During pre-imaging curation it became apparent that the collection holds a number of hidden, previously not recorded specimens. Specimens from the Carpathian Mountains collected by Hugo Ilse for instance often consist of several samples of the same species from more or less the same locality, but collected on different dates. Ilse tended to regroup samples under a single specimen number. These should be recorded as separate specimens and were assigned a suffix to differentiate them during pre-imaging curation.

In 2019 two citizen-science transcription projects, using the web-based crowdsourcing tool DoeDat implemented at Meise Botanic Garden (https://www.doedat.be/) were tested by the MnhnL using imaged specimens. A logged-in user is given tasks where he views a specimen image with its label and has to enter, validate and encode information about the specimen into a database. During the science festival in November 2019 people helped to score the phenology of specimens by indicating whether the specimen had flowers or fruits or neither. Such a crowdsourcing tool can be used for a variety of projects, from transcribing entire specimen labels to recording certain characteristics of the specimen depicted. Many specimen records do not have geographical coordinates. They can be georeferenced

post facto based on the more or less precise description of the collection locality (Wieczorek et al. 2004).

In the course of this mass digitisation project the MnhnL was able to image over 90 % of its vascular plants herbarium specimens within less than two years. This marks a major step in making sure the scientific data available within this important natural history collection is made widely accessible to the public and preserved for future generations.

Acknowledgments

The authors would like to express many thanks to their colleagues and to the volunteers who collaborated in preparing the herbarium for shipment, to all the people who contributed to the transcription projects and to Paul Diederich for his valuable comments on the manuscript.

This digitisation project was funded by the Ministry of Culture as part of their 'Stratégie numérique du patrimoine culturel national'.

References

Bridson, D.M. & L. Forman, 2004. The herbarium handbook. 3rd edition, Kew Publishers.

Colling, G., 2005. Red List of the Vascular Plants of Luxembourg. *Ferrantia* 42: 1-77.

Engledow, H., S. De Smedt, Q. Groom, A. Bogaerts, P. Stoffelen, M. Sosef & P. Van Wambeke, 2018. Managing a mass digitization project at Meise Botanic Garden: From start to finish. *Biodiversity Information Science and Standards* 2: e25912.

Heerlien, M., J. van Leusen, S. Schnorr, S. de Jong-Kole, N. Raes & K. van Hulsen, 2015. The natural history production line: An Industrial approach to the digitization of scientific collections. *ACM Journal of Computational Cultural Heritage* 8 (1): Article 3.

Heuertz, M., 1951. L'herbier du musée d'histoire naturelle de Luxembourg. *Bulletin de la Société Royale de Botanique de Belgique / Bulletin van de Koninklijke Belgische Botanische Vereniging* 83: 259-262.

Lambinon, J. & F. Verloove, 2015. Nouvelle Flore de la Belgique, du Grand-Duché de Luxembourg, du Nord de la France et des Régions voisines. 6e éd. avec corrections. *Jardin botanique national de Belgique, Meise* CXXXIX+1195 pp.

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

Meineke, E.K., B.H. Daru, T.J. Davies & C.C. Davis, (Eds.), 2019. Biological collections for understanding biodiversity in the Anthropocene. *Philosophical Transactions of the Royal Society B* 374 (1763).

Ministère de la Culture, 2018. Vers une stratégie numérique du patrimoine culturel national. Conférence de presse. https://gouvernement.lu/dam-assets/documents/actualites/2018/06-juin/26-digital-culture/Presentation-SNPCN.pdf (last accessed on 31/01/2020).

Nelson, G., P. Sweeney, L.E. Wallace, R.K. Rabeler, D. Allard, H. Brown, J.R. Carter, M.W. Denslow, E.R. Ellwood, C.C. Germain-Aubrey, E. Gilbert, E. Gillespie, L.R. Goertzen, B. Legler, D.B. Marchant, T.D. Marsico, A.B. Morris, Z. Murrell, M. Nazaire, C. Neefus, S. Oberreiter, D. Paul, B.R. Ruhfel, T. Sasek, J. Shaw, P.S. Soltis, K. Watson, A. Weeks & A.R. Mast, 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applications in Plant Sciences* 3: 1500065.

Nic Lughadha, E. & C. Miller. 2009. Accelerating global access to plant diversity information. *Trends in Plant Science* 14: 622-628.

Smith, V. & V. Blagoderov, (Eds.), 2012. No specimen left behind: mass digitization of natural history collections. *Zookeys* 209 (Special Issue): 1-268.

Trias-Blasi, A., W.J. Baker, A. Haigh, D.A. Simpson, O. Weber & P. Wilkin, 2015. A genus-level phylogenetic linear sequence of monocots. *Taxon* 64: 552-581.

Wieczorek, J., Q. Guo & R. Hijmans, 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 18: 745-767.

Wijkamp, N., 2019. Protocol for the Transcription of herbarium collection numbers - Herbarium of the Musée National d'Histoire Naturelle de Luxembourg. Internal document, Botany section, MNHNL.