

The Meise Botanic Garden Herbarium Data Management Plan

Mathias Dillen[‡], Laura Abraham[‡], Ann Bogaerts[‡], Sofie De Smedt[‡], Henry Riley Engledow[‡], Frederik Leliaert[‡], Maarten Trekels[‡], Steven Dessein[‡], Quentin Groom[‡]

[‡] Meise Botanic Garden, Meise, Belgium

Corresponding author: Mathias Dillen (mathias.dillen@plantentuinmeise.be), Steven Dessein (steven.dessein@plantentuinmeise.be)

Abstract

This Data Management Plan outlines a comprehensive strategy for handling, storing, and sharing of data generated by digitisation projects of the herbarium at Meise Botanic Garden with Index Herbarium code BR. Its purpose is to establish clear guidelines for both staff and external users, specifying the terms governing data usage and storage. It aims to prioritise the FAIR principles (Findable, Accessible, Interoperable and Reusable), ensure responsible data management, facilitate long-term preservation, uphold legal and ethical obligations, all while aligning with the research excellence mission of Meise Botanic Garden. This plan serves as a guiding document to effectively and efficiently achieve these goals.

Keywords

FAIR, Open Data Policy, digital herbarium, data preservation, specimen digitisation, metadata standards, data sharing, collection management, open access, data backup and recovery, virtual herbarium, specimen imaging, quality control, natural history collections

Aims

- To provide clarity to our staff and external users on the conditions under which herbarium images and data are used and stored.
- To support Meise Botanic Garden's commitment to research excellence.
- To comply with the data requirements of our stakeholders.
- To enhance the value of our collections for both staff and external users.

- To contribute to Meise Botanic Garden's Open Data policy.

Introduction

Data Management Plans (DMP) are now required by many funding agencies (European Commission 2016). A DMP is a valuable exercise that will help researchers and collection managers at Meise Botanic Garden to understand how to handle collection data generated by digitisation projects and what rights and responsibilities we have.

The DMP process was started as a result of the first herbarium mass digitisation project 'Digitale Ontsluiving Erfgoedcollecties' (DOE!), because the decisions documented by the Plan had a direct impact on the development of the project, its legacy and the ongoing costs (Engledow et al. 2018). Without this clarity, it is difficult to plan and progress with confidence. The DMP is helping us ensure that stored materials are in the best format and that their data are well organised and linked to metadata. This will simplify access to data, but also ensure the long-term usefulness of these data. Additionally, the DMP outlines the storage of our data to make it accessible indefinitely. The DMP is not to be considered a static document. It is expected to evolve as the Garden's policies and priorities change. New versions will be published as they are approved by Meise Botanic Garden's Executive Committee.

The DMP is helping us improve data integrity and security, and clarifying what access management is required and who is responsible. The data of the Garden should be used to improve opportunities for research collaboration and funding. It will enhance the research reputation of the Garden, particularly when data are cited properly.

Data generated by the Garden constitute valuable assets, holding significance from scientific, management, historical and cultural perspectives. These data are not only assets of the Garden, but public resources with present and future value. As a public research institution, we need to ensure that we comply with Flemish, Belgian and European regulations on access to and storage of data. We also need to ensure that we are aligned with the policies of other European institutions responsible for biodiversity data, such as natural history museums, universities and other institutes. The Flemish Government recognises an Open Data policy as the standard for public data, as stipulated by the Flemish Open Science Board (FOSB), although the position regarding data from museums and herbaria remains unclear. Long-term digital cold storage is conducted at [Meemoo](#) (formerly VIAA), the Flemish Institute for Archiving, established to provide access to multimedia from the cultural, heritage and media sectors. The Garden's Open Data policy is also aligned with the framework of the European Open Science Cloud (EOSC), as envisaged by the European Commission.

Finally, it is worth acknowledging that the scientific community as a whole faces challenges when it comes to ensuring the reproducibility of research (Baker 2016). By following a DMP, we will support data re-use by making them easier to discover. In order to improve this record, it is essential that data used in research are properly documented and

identifiable. Many of the data we administer require validation and can change as more information becomes available. This presents challenges for research reproducibility from the perspective of tracking and data integrity. However, proper processes will ensure that this can be achieved cheaply and reliably. We strive to align with the recommendations and spirit of the FAIR data principles to ensure that our data are as much as possible findable, accessible, interoperable and reusable (Wilkinson et al. 2016).

Scope

This DMP pertains to the outputs of herbarium digitisation. The term 'data' is used in the broad sense including both structured, tabular data, such as in databases, and unstructured data, such as notes on herbarium specimens and the images of herbarium specimens. It also includes the metadata, which makes the data understandable, traceable and reusable.

It is important to note that the Garden's data are not "owned" by any specific individual or department, though the work of individual scientists should be protected for a sufficient time so that they are able to validate the data and publish their results.

While this DMP was initially conceived in response to the DOE! projects, it covers all herbarium specimen images and data, including related collections such as alcohol material, wood samples, DNA samples, etc.). Consequently, it serves as an institution-wide policy paper.

The data outputs of the digitisation of herbarium specimens are listed below:

- High resolution images of herbarium specimens;
- Metadata associated with the imaging of the specimen;
- Unique specimen barcodes;
- Transcribed specimen label data which may include either some or all of the fields in Table 1.

In addition, there are data that are interpretations of the label data, such as georeferencing, the accepted scientific name, the collector's identity and country of origin.

Other forms of data that are within the scope are links between our specimens and other data sources, both internal to the Garden and in other institutions. For example, within the Garden, our specimens are linked to literature, particularly the *Flore d'Afrique Centrale* and the Garden's journals, such as *Plant Ecology and Evolution* and original publications, such as protologues. There are links to the living collections where some specimens may have living counterparts and links from the named locations to historical maps, itineraries and gazetteers. There are also links between specimens and illustrations, macro-photographs, micro-photographs, liquid material, silica gel material and DNA sequence data.

Data, specimens and literature held in other institutions can also be connected to our specimens. For example, the taxonomic names we use are linked to the International Plant Names Index (IPNI), Index Fungorum, Mycobank and World Register of Marine Species (WoRMS). The biogeographic data of our specimens are linked to data held at the Global Biodiversity Information Facility (GBIF). Molecular sequence data from specimens are linked to public sequence depositories under the International Nucleotide Sequence Database Collaboration (GenBank, EMBL-EBI/ENA, DDBJ); specimens and names are linked to literature in the Biodiversity Heritage Library (BHL) and, finally, specimens are linked to duplicate specimens within Belgium and abroad.

Roles and responsibilities

Maintaining regular communication between the curatorial and informatics personnel is vital to promptly and transparently address any issues or alterations. Table 2 details the roles and responsibilities for defining, managing, controlling and maintaining herbarium data. Scientists also bear the responsibility of maintaining their own data, particularly making sure those data are made available upon publication of their results.

Data volumes and storage

Meise Botanic Garden is a national repository for botanical specimens and central point for botanical research in Belgium. The Garden already had a large collection of digitised herbarium specimens before the mass digitisation projects started. During the first DOE! project (2015-2018), the whole Belgian and African vascular plant collection of 1.2 million specimens were digitised by a private company. Another 1.2 million vascular plant specimens and macroalgae specimens were digitised for the second mass digitisation project DOE!2 (2018-2021) (De Smedt et al. 2019). Next to these mass digitisation projects specimens are also digitised using our own imaging infrastructure. As of September 2023, we have about 2,714,000 images and the number still grows as all incoming material is digitised after mounting.

As of September 2023, we have about 2,805,506 specimens catalogued in our herbarium database. Additional digitisation and updating of label information is ongoing and will continue in the future.

Data maintenance and citation of specimens

To keep them authoritative, data on physical specimens are best maintained close to the source. The Garden's public [data portal](#) has a feedback system and contact details so that users can communicate issues they find with the data. Users are encouraged to provide corrections and the curator is responsible for updating the Garden's database with these corrections.

It is important to ensure the correct citation of our specimens in scientific publications. We expect our specimens to be correctly cited if they constitute BR material. The data portal provides a permanent Uniform Resource Identifier (URI) that will uniquely identify a specimen and we will guarantee that this URI will always resolve to the data portal page where the image and label data of the specimen is displayed. An example of such a URI is <http://www.botanicalcollections.be/specimen/BR0000024719261>. This URI is the recommended method of citing our specimens. As URIs may not always be supported by publishers or repositories, we also provide recommendations for textual citations. The recommended textual citation of a specimen includes the collector(s) name, the collector's number and its Index Herbariorum code (BR). Citations should always include the specimen's barcode (e.g. BR0000024719261), as that is its unique physical identifier. Examples:

Plant Ecology and Evolution (<https://doi.org/10.5091/plecevo.2020.1670>)

Gabon: Ivindo National Park, 0°15'S, 12°20'E, 10 Apr. 2004, fr., Mougazi 1545 (holotype: BR [BR0000009456501]; isotypes: LBV, WAG [WAG0318084, WAG0122835]).

Phytokeys (<https://doi.org/10.3897/phytokeys.133.38694>)

MOZAMBIQUE. Manica Province: Magorogodo hills, Zomba Community, 19°54'28"S, 33°11'4"E, c. 559 m alt., fl. and fr. 28 October 2013, B.T. Wursten BW897 (holotype: BR, BR0000020700003)].

European Journal of Taxonomy (<https://doi.org/10.5852/ejt.2022.801.1685>)

MADAGASCAR - Antananarivo Province • forêt

d'Ambohitantly, jardin botanique; 1464

m a.s.l.; 6 Feb. 1999; fr; De Block & Rakotonasolo 736;

holotype: BR[BR0000022757661]; isotypes: BR[BR0000022757616], BR[BR0000022757623].

Data organisation and documentation

Prior to photographing, all specimens will be labelled with a unique number in the form of a barcode. This barcode is printed on archival quality labels and we are currently using the Code 128 barcoding format (ISO/IEC 2007).

New incoming material will be digitised before they are stored in the collection.

Image file formats

The image formats have been selected after consultation with Meemoo.

Important criteria that we considered were:

- Long term sustainability.
- Storage capacity.
- Image quality.
- Functionality (viewable over Internet *via* a browser-based image viewer).

For reference, other sources of information on image formats include the CEST website, Cultureel Erfgoed Standaarden Toolbox (<http://projectcest.be/wiki/Hoofdpagina>), Arms and Fleischhauer (2005), Gillesse et al. (2008).

Image formats we use in the Garden:

- The baseline uncompressed TIFF 6.0 is used for the cold storage of our images: this format is the most commonly advised format for long-term archiving.
- JPEG-2000 Part 1 Lossless (JP2) is used for displaying images on our data portal: it gives a smaller file size than TIFF, but its compression algorithm means that a JP2 is 40% smaller than the equivalent TIFF, while still providing the technologies such as quality layers and tiling for viewing the images over the internet. Conversion parameters from TIFF to JP2 were fixed after consultation with Meemoo and Picturae NV.
- The older standard JPEG (www.jpeg.org/jpeg/) is used at 50% quality level from the original TIFF to facilitate downloading of a smaller image from the Internet.
- During the first DOE! project, archival TIFF files were sent to Meemoo on LTO6 tapes after the internal quality control process. During the second DOE! project, archival TIFF files were sent to Meemoo via FTP. Currently approximately 270 TiB in total is archived by meemoo. Only the JP2 and JPEG files are stored in the Garden, with backup.

Data access and sharing

Users of the public data portal will have freedom to browse and search the digital data of the Herbarium without the need to login. They will be able to view data and view high quality images anonymously.

Images available for download will be at least 420 dpi in JPEG format. These images will be licensed with a [Creative Commons Attribution 4.0 International licence](#).

Data are available for download directly on the website. A valid e-mail address is required to request bulk data downloads. Data will be downloadable in [simple Darwin Core format](#) to facilitate compatibility with other software systems. They will be distributed under a Creative Common Attribution 4.0 International Licence, except for the specimen's barcode,

country of origin and scientific name, which will instead be published under a [CC0 \(Public Domain\)](#) licence.

Data will also be downloadable in RDF/XML format, following the [Darwin Core RDF guide](#) and the [CETAF Specimen Preview Profile](#) more specifically. In this format, data can be more easily accessed and interpreted by machines, including Linked Open Data initiatives.

Embargoing data

It is recognised that scientists expend considerable effort collecting and identifying their own specimens, as well as digitising the data on those specimens. They should have a grace period of exclusive use of those data, before they are available to the community as a whole. Should a scientist wish to do so, they can embargo use of the specimens they work on and their associated data. This embargo will last for four years from the date of digitisation. This embargo only applies to data collected by a scientist or digitised on request of the scientist. The scientist has to justify why the data should be blocked. When the project has concluded, the scientist should inform the database manager so that the embargo can be lifted. No specimen should be blocked that is specifically referred to in a publication. If a specimen has already been published to the data portal, it cannot be removed. When the embargo period has expired, the embargo will be lifted after consultation with the scientist, who has the possibility to extend the embargo. While data are embargoed, they will be invisible to users of the data portal.

Note that it is important that specimens collected on expeditions are digitised and mounted as soon as possible upon return from the mission. This will protect Meise Botanic Garden by ensuring that documentation procedures are followed and that details related to collecting and export permits are digitised immediately. We will also continue to look for opportunities to digitise some of our historical collections that have not been part of our past digitisation efforts as they require adjustments to the methodology, such as oversized herbarium sheets, slides with diatoms, wood samples and preserved fungi.

Sensitive data

There may be occasions where we are requested to restrict access to data. The reason may be to protect the sites of rare species, such as those listed under CITES, but there may be other reasons, for example protecting the biographical data of living collectors who may not want these shared.

The data on plants and their localities will not be considered sensitive by default. Should we be asked to obscure information on the grounds of sensitivity, we will review it on a case-by-case basis, considering issues of whether the data are available elsewhere and whether the benefits of secrecy outweigh the potential risks. Currently, we do not restrict access to any specimens on grounds of sensitivity and have never been asked to restrict access for this reason.

We will only make full biographical information available for dead collectors. This information will be added to Wikidata. Living collectors will be identified only by their name.

Data re-use

Europeana

Images have been supplied to Europeana with the same image quality and licensing as previously. CC0 data will be provided through the Garden's Integrated Publishing Toolkit ([IPT](#)) server and JPEG images on a dedicated image web server.

Global Biodiversity Information Facility (GBIF)

Datasets that are supplied to GBIF will be hosted on the Garden's IPT server. Data will be supplied to GBIF for the whole collection, except for embargoed specimens and specimens with wrong and/or missing basic data that need to be corrected.

JSTOR Global Plants

More than 74000 images and data of type and historical material are available on [JSTOR Global Plants](#).

IPNI & Tropicos

We will continue to exchange nomenclatural information with IPNI and Tropicos on an ad hoc basis. Where possible we will make links between the names in our database and these other databases. This will help valorise our data and reduce errors in all databases.

Data Preservation

Two backup procedures are used to ensure permanent data preservation. The TIFF images are stored offsite, at Meemoo (formerly VIAA), where they are stored on tape (LTO6) in three different data centres. At one of those data centres, there is also a hard disk repository of lower resolution JPEG images held for rapid viewing of the images in the tape archive.

At the Botanic Garden, JP2 and JPEG images are stored locally on our servers in two ICT rooms located in separate buildings. Each image is therefore stored on a 'production' storage area network (SAN) and a second copy is located on the 'backup' SAN. Synchronisation from production to backup site is scheduled every night. These SANs use redundant disk arrays to ensure continuous and reliable access to the images. Additionally, to improve availability of our servers and storage, these devices have dual power supplies and are connected to an uninterruptible power source (UPS).

Ethical considerations: decolonisation

The collections at Meise Botanic Garden have a global scope, with a focus on Central Africa, originating from botanical expeditions in the 19th and 20th centuries. We acknowledge the historical context of colonisation in the Garden's history. While altering the past is not possible, our commitment and willingness to learn from this history is certain. By embracing fairness and openness in our collection management and practices, we aim to foster a more inclusive and equitable approach to botanical research and conservation (Park et al. 2023). Digitisation plays a crucial role in this endeavour by enhancing accessibility for a wider audience. Many of our works on Central Africa are openly available through our websites, such as our [Central African Flora](#) series and on data aggregators, such as our publication of Central African vernacular names on GBIF (Meise Botanic Garden and Dillen 2018). In managing these collections, we prioritise ethical stewardship, cultural sensitivity and historical awareness, fostering a deeper understanding of our shared botanical heritage. We are open to collaborating with individuals from the countries where these collections originate, aiming to enhance our offerings and increase accessibility beyond Belgium.

Funding programme

The development of the Data Management Plan for Meise Botanic Garden was made possible through the support of various funding sources. The Research Foundation – Flanders (FWO), contributed funds as part of the Flemish contribution to the DiSSCO Research Infrastructure under grant n° I001721N and through the funding provided by the Flemish Government for the DOE! (IWT140930) and DOE!2 project (VR 2018 0806 DOC. 0603/1 en DOC.0603/2). Furthermore, this project received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 871043, as part of the DiSSCO Prepare and under grant agreement No. 777483, as part of the ICEDIG initiative. Lastly, grant agreement No. 101007492, as part of the BiCIKL project.

Status of the Data Management Policy

This Policy (version 8.0) was officially adopted in late November 2021 by the Executive Committee (Directiecomité) of the Meise Botanic Garden, to bring the Data Management Plan (DMP) in line with current policy after the two Digitale Ontsluiting Erfgoedcollecties (DOE!) projects.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Arms C, Fleischhauer C (2005) Digital Formats: Factors for Sustainability, Functionality, and Quality. Archiving Conference 2 (1): 222-227. <https://doi.org/10.2352/issn.2168-3204.2005.2.1.art00047>
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533 (7604): 452-454. <https://doi.org/10.1038/533452a>
- De Smedt S, Bogaerts A, Engledow H, Groom Q (2019) Different Approaches between First and Second Mass Digitisation Project for the Herbarium (BR) at Meise Botanic Garden. Biodiversity Information Science and Standards 3 <https://doi.org/10.3897/biss.3.37259>
- Engledow H, De Smedt S, Groom Q, Bogaerts A, Stoffelen P, Sosef M, Van Wambeke P (2018) Managing a Mass Digitization Project at Meise Botanic Garden: From Start to Finish. Biodiversity Information Science and Standards 2 <https://doi.org/10.3897/biss.2.25912>
- European Commission (2016) Guidelines on Data Management in Horizon 2020. URL: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Gillesse R, Rog J, Verheusen A (2008) Life Beyond Uncompressed TIFF: Alternative File Formats for the Storage of Master Image Files. Archiving Conference 5 (1): 41-46. <https://doi.org/10.2352/issn.2168-3204.2008.5.1.art00009>
- ISO/IEC (2007) ISO/IEC 15417:2007: Code 128 bar code symbology specification. 2. URL: <https://www.iso.org/standard/43896.html>
- Meise Botanic Garden, Dillen M (2018) Vernacular names of the Flore d'Afrique Centrale. 1.7. GBIF. URL: <https://doi.org/10.15468/hwrwlp>
- Park D, Feng X, Akiyama S, Ardiyani M, Avendaño N, Barina Z, Bärtschi B, Belgrano M, Betancur J, Bijmoer R, Bogaerts A, Cano A, Danihelka J, Garg A, Giblin D, Gogoi R, Guggisberg A, Hyvärinen M, James S, Sebola R, Katagiri T, Kennedy J, Komil T, Lee B, Lee SL, Magri D, Marcucci R, Masinde S, Melnikov D, Mráz P, Mullenko W, Musili P, Mwachala G, Nelson B, Niezgoda C, Novoa Sepúlveda C, Orli S, Paton A, Payette S, Perkins K, Ponce MJ, Rainer H, Rasingam L, Rustiami H, Shiyani N, Bjorå CS, Solomon J, Stauffer F, Sumadijaya A, Thiébaud M, Thiers B, Tsubota H, Vaughan A, Virtanen R, Whitfield TS, Zhang D, Zuloaga F, Davis C (2023) The colonial legacy of herbaria. Nature Human Behaviour 7 (7): 1059-1068. <https://doi.org/10.1038/s41562-023-01616-7>
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 (1). <https://doi.org/10.1038/sdata.2016.18>

Table 1.

List of data elements that may be transcribed from specimen labels and mapping to Darwin Core where there is one.

collector name (dwc:recordedBy)	locality (dwc:locality)	country (dwc:country)	geographical coordinates (dwc:decimalLatitude and decimalLongitude)
ocean or sea	collection date (dwc:eventDate)	collector number (dwc:recordNumber)	habitat (dwc:habitat)
type status (dwc:typeStatus)	determiner (dwc:identifiedBy)	scientific name(s) (dwc:scientificName)	vernacular name (dwc:vernacularName)
determination date (dwc:dateIdentified)	ecological data	ethnobotanical uses	information on collection permits
associated species	associated specimens	macroscopical and or microscopical characteristics	details of former herbaria where the specimen was deposited
herbarium code (Index Herbariorum acronym) (dwc:institutionCode)	unique identifier (barcode) (dwc:occurrenceID)		

Table 2.

Responsibilities and corresponding job titles for herbarium digital data and images.

Responsibility	Role
Post imaging processing	images manager
Integrity of database	database manager
Ensuring backups of images	images manager
Ensuring backups of data	database manager
Updating the portal	portal manager
Portal maintenance	portal manager
Maintaining image metadata	images manager
Loans management	scientific manager Herbarium
Ensuring barcoding uniqueness	database manager
Digitisation prioritisation	scientific manager Herbarium
Imaging	scientific manager Herbarium
Transcription	scientific manager Herbarium
Decisions on data sharing	scientific manager Herbarium