DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at info@dissco.eu.

Title
MS1.3 Corpus of previous studies on prioritisation of digitisation compiled

Author(s)
Louise Isager Ahl (NHMD)
Henrik Enghoff (NHMD)

Identifier of the author(s)
0000-0002-4727-1882
0000-0002-2764-8750

Affiliation
Natural History Museum of Denmark (NHMD), University of Copenhagen

Contributors
Sarah Rossi De Gasperis (UniFi) - ORCHID 0000-0002-1609-7412
Gianna Innocenti (UniFi) - ORCHID 0000-0002-4504-0765
Lorenzo Cecchi (UniFi) - ORCHID 0000-0001-9706-0947
Luca Bellucci (UniFi) - ORCHID 0000-0002-8464-6153
Anne Koivonen (LUOMUS)
Heli Fitzgerald (LUOMUS) - ORCHID 0000-0002-6754-6409
Sofie de Smedt (Meise) - ORCHID 0000-0001-7690-0468
Patricia Mergen (Meise) - ORCHID 0000-0003-2848-8231
Elspeth Haston (RBGE) - ORCHID 0000-0001-9144-2848
Pierre-Yves Gagnier (MHNH) - ORCHID 0000-0002-7104-898X
Ana Casino (CETAF) - ORCHID 0000-0002-9869-6573
Laura Tilley (CETAF) - ORCHID 0000-0002-2455-7992
Tina Loo (Naturalis) - ORCHID 0000-0002-8250-1786
Laurence Livermore (NHM) - ORCHID 0000-0002-7341-1842
Helen Hardy (NHM) - ORCHID 0000-0002-9206-8357
Lisa French (NHM) - ORCHID 0000-0001-7279-8582

Publisher

Identifier of the publisher

Resource ID

Publication year
2022

Related identifiers

Is it the first time you submit this outcome?
No

Creation date
15/02/2022

Modification date
15/02/2022

Summary of modification
This is the first submission

Version
1

Citation
Ahl, L. I. and Enghoff, H. MS1.3 Corpus of previous studies on prioritisation of digitisation compiled. 2022

Abstract
This report describes progress made by Work Package 1, Task 3 of DiSSCo Prepare Project:
Task 1.3 Establish relevant criteria to identify a prioritisation model for digitisation.

According to the DPP Description of Work, Task 1.3 will, based on the analysis of previous studies, identify relevant criteria and develop them into a basic model for the prioritisation of digitisation of objects held in Natural Sciences Collections (NSCs). Over the past reporting period the focus of Task 1.3 has been to (1) identify new relevant publications, and (2) carry out an investigation of what is currently being done by DiSSCo members in regards to prioritisation of their digitisation efforts.

The result of the work done in Task 1.3 has been the the identification of 12 new publications, and the compilation of replies from 23 DiSSCo partners regarding their prioritisations.

Content keywords
data

Project reference
DiSSCo Prepare (GA-871043)

WP number
WP1

Project output
Milestone report

Deliverable/milestone number
MS1.3

Dissemination level
Public

Rights

License
CC0 1.0 Universal (CC0 1.0)

Resource type
Text

Format
docx

Funding Programme
H2020-INFRADEV-2019-2

Contact email
Louise.ahl@snm.ku.dk

Title: MS1.3 Corpus of previous studies on prioritisation of digitisation compiled

Authors:

Louise Isager Ahl (NHMD)

Henrik Enghoff (NHMD)

Identifier of the author(s) (ORCHID):

0000-0002-4727-1882

0000-0002-2764-8750

Affiliation:

Natural History Museum of Denmark (NHMD), University of Copenhagen

Contributors:

Sarah Rossi De Gasperis (UniFi) - ORCHID 0000-0002-1609-7412
Gianna Innocenti (UniFi) - ORCHID 0000-0002-4504-0765
Lorenzo Cecchi (UniFi) - ORCHID 0000-0001-9706-0947
Luca Bellucci (UniFi) - ORCHID 0000-0002-8464-6153
Anne Koivonen (LUOMUS)
Heli Fitzgerald (LUOMUS) - ORCHID 0000-0002-6754-6409
Sofie de Smedt (Meise) - ORCHID 0000-0001-7690-0468
Patricia Mergen (Meise) - ORCHID 0000-0003-2848-8231
Elspeth Haston (RBGE) - ORCHID 0000-0001-9144-2848
Pierre-Yves Gagnier (MHNH) - ORCHID 0000-0002-7104-898X
Ana Casino (CETAF) - ORCHID 0000-0002-9869-6573
Laura Tilley (CETAF) - ORCHID 0000-0002-2455-7992
Tina Loo (Naturalis) - ORCHID 0000-0002-8250-1786
Laurence Livermore (NHM) - ORCHID 0000-0002-7341-1842
Helen Hardy (NHM) - ORCHID 0000-0002-9206-8357
Lisa French (NHM) - ORCHID 0000-0001-7279-8582

Publication year: 2022

Milestone number	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	MS1.3 Corpus of previous studies on prioritisation of digitisation compiled	3 - UCPH	25	Report submitted to EB

This report describes progress made by Work Package 1, Task 3 of DiSSCo Prepare Project:

Task 1.3 Establish relevant criteria to identify a prioritisation model for digitization

MS3 should be regarded as a stepping stone leading to the final product (deliverable) of Task 1.3, viz.

D1.3 Report on relevant criteria for prioritization of the digitisation

Which is due by the end of the DiSSCo Prepare Project (Month 36).

Task 1.3 is being done by the following DPP partners:

UCPH (task lead), Naturalis, CETAF, Luomus, RBGE, MeiseBG, UniFi, NHM, MNHN,

The task

According to the DPP Description of Work, Task 1.3 will, based on the analysis of previous studies, identify relevant criteria and develop them into a basic model for the prioritisation of digitisation of objects held in Natural Sciences Collections (NSCs).

Modus operandi

Work on T1.3 was initiated at the DPP All Hands online meeting, January 2021. Dedicated T1.3 online meetings were held on 12 March 2021 and 12 January 2022. Apart from these meetings, communication within the task group has been through email.

Analysis of previous studies

Deliverable D2 of the ICEDIG project (Bakker *et al.* 2018) provides an impressive corpus and analysis of digitization criteria and will form a very substantial part of the basis for the final deliverable of DPP Task 1.3.

An earlier report by GBIF (Krishtalka *et al.* 2016: Annex II) includes an analysis of a large-scale survey of digitisation in Natural History Collections and will likewise provide essential input to the final deliverable of DPP Task 1.3.

Search for additional studies on digitization criteria

Based on the previous studies outlined above additional searches were carried out in April 2021 to investigate if additional work had been published on the topic.

The following searches were carried out:

1. Search: ”natural history collections” ”prioritization” since 2017, google scholar: 143 results. Sorted by relevance. By page 2, 3 and 4 nothing was relevant. 4 relevant publications identified.
2. Search: ”natural history collections” ”digitization” since 2017, google scholar: 775 results. Sorted by relevance. 4 relevant publications identified.
3. Search: ”digitization” ”prioritization” since 2017, google scholar: 4.640 results. Sorted by relevance. 2 relevant publications identified.

4. Search: ”natural history collections” ”digitization” ”prioritization” since 2017, google scholar: 46 results. Sorted by relevance. 2 relevant publications identified.

The relevant works that had not been included in the ICEDIG and GBIF reports were identified and scored (1-3) based on relevance for the investigation with 1 being most relevant. A total of 12 new publications were identified from the 4 searches.

Digitisation plans and criteria used by DiSSCo Partners

Considering the completeness of ICEDIG's report (Bakker *et al.* 2018) it was decided to supplement the original task programme with an analysis of digitization plans and criteria used by DiSSCo Partners. På metodefronten står vi derfor noget længere tilbage end eksempelvis DNA-forskningen.
Based on the works identified in the ICEDIG, GBIF and our own additional search, a letter with some guiding questions was designed and distributed among the T1.3 partners to investigate if the same or similar procedures were carried out between the different organisations represented here. As that dataset was fairly limited it was decided to go even further and distribute it among both the DPP and DiSSCo partners, the aim being to get replies from as diverse a group of organisations as possible. As replies were hard to obtain from both DPP and DiSSCo partners we presented our questionnaire at the DiSSCO National Nodes meeting on the 25th November 2021. This sparked more responses and by 22nd of December 2021 we had received a total of 23 replies. These replies were compiled in a master document structured in accordance with the initial letter and organised after country and institution.

The compiled document containing all the information obtained is presented in Appendix 1.

Towards a prioritization model for digitization

With the information collected so far, the milestone target: “Corpus of previous studies on prioritisation of digitisation compiled” can be regarded as reached. Nevertheless, the search for additional information will continue during the first months of 2022. This may take place in the form of interviews and/or smaller targeted questionnaires. The information thus obtained will be analyzed together with that already at hand and developed into “a prioritisation model for digitization”. In line with the way the new information was collected: mainly “free-text under guidance” rather than a strict questionnaire, the focus will not be on statistics; rather, it will be attempted to provide a guide booklet on prioritization of natural science collections for decision makers. Bakker *et al.* (2018) suggested several possible models of decision-making for prioritisation of digitization: 1) Decision tree, 2) Scoring method, 3) Panel review. These will be considered in the light of the new findings.

The current working model for the planned prioritization booklet operates with four main groups of criteria to be considered:

- **Relevance**
- **Data quality**
- **Cost**
- **Feasibility**

Of these, data quality deserves particular attention because although of great importance, this has not been very much considered in previous studies. See Chapman (2005) for a thorough treatment of the data quality concept.

Fig. 1 visualizes the interplay of the four criteria. **Project A** and **Project B** will both deliver data of high quality and high relevance. Although Project B data will be of slighter lower quality, this project

may be chosen because of higher feasibility. **Project C** has little to recommend it, whereas **Project D**, with low data quality, medium relevance and feasibility, and low cost, might be prioritized depending on what the data will primarily be used for.

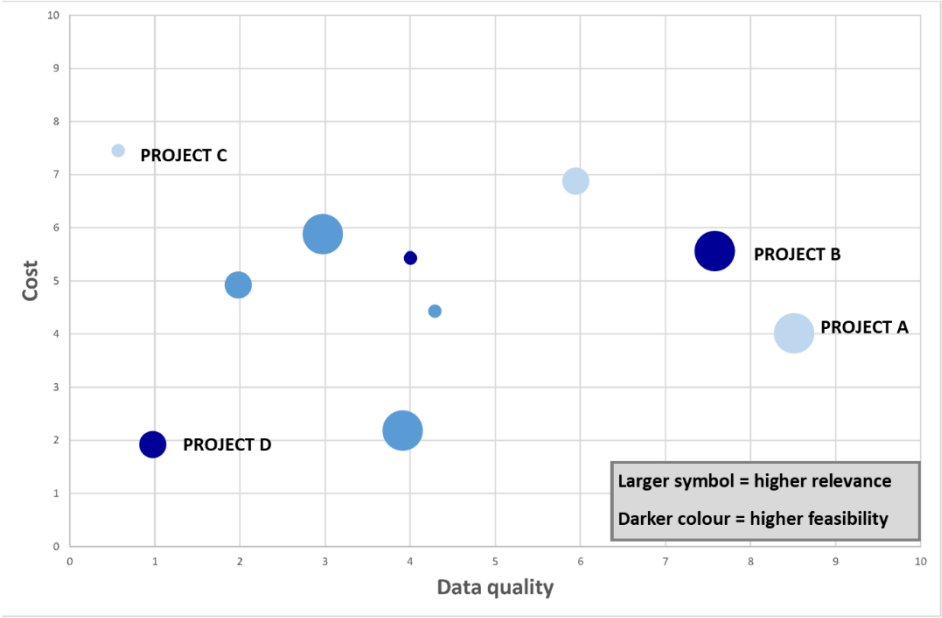


Fig. 1. Four main groups of criteria to be considered when prioritizing digitization projects. In the diagram, **data quality** and **cost** are represented on the horizontal and vertical axes (axis values are arbitrary), whereas **relevance** is represented by the size of the symbols, and **feasibility** by the intensity of symbol colour.

References

Bakker, Hannco P.A.J.; Willemse, Luc; van Egmond, Emily; Casino, Ana; Gödderz, Karsten; Vermeersch, Xavier 2018. Inventory of criteria for prioritization of digitisation of collections focussed on scientific and societal needs. A deliverable (D2.1) of the ICEDIG project. https://doi.org/10.5281/zenodo.2579156

Chapman, A. D. 2005. Principles of Data Quality, Global Biodiversity Information Facility. https://doi.org/10.15468/doc.jrgg-a190

Krishtalka L, Dalcin E, Ellis S, Ganglo JC, Hosoya T, Nakae M, Owens I, Paul D, Pignal M & Thiers B (2016) *Accelerating the discovery of biocollections data*. Copenhagen: GBIF Secretariat. Available online at: http://www.gbif.org/resource/83022.

APPENDIX 1

Search no.	First author	Title	Year	DOI
1	Hedrick, BP	Digitization and the future of natural history collections.	2020	https://doi.org/10.1093/biosci/biz163
1	Willemse, L	Future Challenges in Digitisation of Private Natural History Collections	2019	https://doi.org/10.3897/biss.3.37640
1	Cantrill, DJ	The Australasian Virtual Herbarium: Tracking data usage and benefits for biological collections.	2018	https://doi.org/10.1002/aps3.1026
1	Paton, A	Plant and fungal collections: Current status, future perspectives	2020	https://doi.org/10.1002/ppp3.10141
2	Miller, SE	Building Natural History Collections for the Twenty-First Century and Beyond	2020	https://doi.org/10.1093/biosci/biaa069
2	Nelson, G	The history and impact of digitization and digital data mobilization on biodiversity research	2018	https://doi.org/10.1098/rstb.2017.0391

2	Schindel, DE	The next generation of natural history collections	2018	https://doi.org/10.1371/journal.pbio.2006125
2	Hereld, M	LightningBug ONE: An experiment in high-throughput digitization of pinned insects	2019	https://doi.org/10.3897/biss.3.37228
3	Wetzel, FT	Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe	2018	https://doi.org/10.1016/j.biocon.2017.12.024
3	Daru, BH	Widespread sampling biases in herbaria revealed from large-scale digitization	2017	https://doi.org/10.1111/nph.14855
4	Nekola, JC	Caveat consumptor notitia museo: Let the museum data user beware	2019	https://doi.org/10.1111/geb.12995
4	Veiga, AK	A conceptual framework for quality assessment and management of biodiversity data	2017	https://doi.org/10.1371/journal.pone.0178731

Appendix 2 - Compilation of Prioritisation of Digitisation answers

Table of Contents

Introduction

In the framework of the DiSSCo Prepare Project (DPP), several project partners are engaged in Task 1.3: Establish relevant criteria to identify a prioritisation model for digitisation. The final outcome of the task will be "a basic model for the prioritisation of digitisation of objects held in Natural Sciences Collections".

In response to letters sent in May and November 2021 we have by December 22., 2021 received **25** responses. The responses have been compiled in this document and should be read in preparation for the Task 1.3 meeting in January 2022.

Excerpt of the letter with the questions sent to the DiSSCo and Task 1.3 institutions:

In the task group we are discussing what such a “basic model” should look like, and in order to support our discussions and eventual decisions **we want to obtain information from DiSSCo partner institutions on**:

1. Digitisation strategy of the partner institutions (if available, please provide a copy or link).
2. Prioritisation criteria employed for digitisation which has already been done or is in progress.

Rather than presenting you with a lengthy questionnaire, we kindly ask if you could provide information, in free text and preferably no more than **2 A4 pages**, on the above two points as far as your own institution is concerned.

The following few questions might be helpful in highlighting relevant topics:

- Do you have a clear overview of the digitisation status of your institution (how many specimens databased, how many imaged, by which procedural standard etc.)?
- Are you monitoring it? How?
- What is your digitization level: specimen level or higher collection unit level? What are your policies with respect to how much data is acquired (databasing/ transcription of specimen information and/or imaging)?
- Do you have a unique management software or more than one? What kind of protocol are you using for the data digitisation (e.g., ICEDIG guidelines)?
- Do you have a procedure for validating data (e.g., accuracy of identification and georeferenced)?
- What are you planning to digitise next and what projects are planned for further down the line and why?
- If you do not have a defined plan, what are the circumstances driving you to unplanned digitisation actions (e.g., specimens requested for loan, new accessions, specimens involved in an exhibition, etc.)?

It might also be useful to distinguish between

- Mass digitisation or large scale where indeed the questions of prioritization, feasibility etc are very relevant
- Digitisation on Demand
- Opportunistic Digitisation

Replies were asked to be sent to us by **8 December 2021**, a few institutions contacted us and asked for extensions which were granted, and their replies have all be received now.

Compilation of information from DiSSCo partner institutions

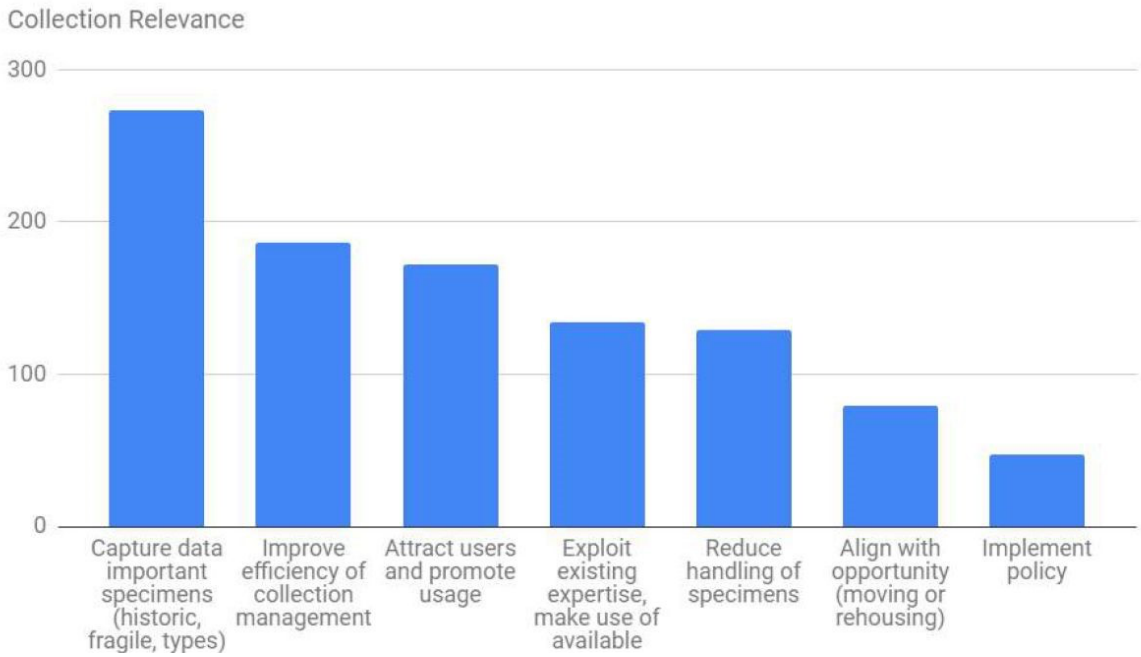
List of all countries in DiSSCo and the institutions from each country that have replied to our questionnaire. Institutions marked with * are partners in task 1.3 lead by NHMD. Only replies and comments obtained by December 22, 2021 are included.

- Austria
- Belgium
  - Royal Belgian Institute of Natural Sciences (BE-RBINS)
  - Meise Botanic Garden*
  - Royal Museum of Central Africa (RMCA)
- Bulgaria
- Czech Republic
- Denmark
  - Natural History Museum of Denmark (NHMD)*
- Estonia
  - Estonian Museum of Natural History (EMNH)
  - Estonian National Node
- Finland
  - LUOMUS*
- France
  - Muséum National d'Histoire Naturelle, Paris (MNHN)*
- Germany
  - MfN Berlin
- Great Britain
  - KEW
  - Natural History Museum, London (NHM)*
  - Royal Botanic Garden Edinburgh (RBGE)*
- Greece
- Hungary
- Italy
  - Florence, UniFi*
- Luxembourg
  - National Museum of Natural History, Luxembourg (MnhnL)
- Netherlands
  - Natural History Museum of Rotterdam (NHMR)
  - Natuur Museum Brabant (NMB)
  - Naturalis*
- Norway
  - Arctic University Museum of Norway (AUMN)
- Poland
- Portugal
- Slovakia
- Spain
  - MAH (Herbarium), Real Jardin Botanico, CSCI
  - MACBH (Herbarium)
  - MAFH (Herbarium)
  - Museo Nacional de Ciencias Naturales (MNCN)
  - Entomology Collection – Complutense University of Madrid (UCME)
- Sweden
  - Herbarium Gothenburg (HBG)

1. Digitisation strategy of the partner institutions (if available, please provide a copy or link).

Belgium

BE-RBINS:
There is no official document, but our digitisation priorities are in line with the ICEDIG survey.



Denmark

NHMD:
The collection of the Natural History Museum of Denmark totals at least 14 million objects, many of which are wet-preserved lots in alcohol. Thus, the total number of specimens is uncertain. In order to get on overview of our collections and to standardize and coordinate future digitization efforts, the museum introduced a new collection management system a few years ago. This replaced more than 60 individual old databases in various parts of the museum ranging from simple spreadsheets to outdated proprietary commercial databases to bespoke systems build by individuals no longer employed by the museum.
The new system is mandatory to use across the museum and transferring legacy data from the old databases is still ongoing. Having a shared database where all registrations are entered is considered crucial for future digitization. To prepare for a future shared database of all digitized Danish natural history collections, all Danish natural history museums have agreed to use the same collection management system as their collection database and a Danish natural history collection database consortium has been established. Based on current data in the database, we know that approximately 5% of the museum collection have been databased and that most of these data represent transcribed label data. No additional effort has been spent to add georeferences to the original label data.

Presently, the Natural History Museum of Denmark does not have a formalised digitization strategy. However, DaSSCo (Danish System of Scientific Collections) funded by the national roadmap for research infrastructure, and lead by the Museum, will establish such a strategy as its first task. DaSSCo will be the Danish DiSSCo node and funding is provided to set up digitization laboratories at the Museum. This will include both known, established digitization infrastructure and techniques and development of new innovative digitization techniques in close collaboration with the Danish Technical University (DTU) and Department of Computer Science at the University of Copenhagen.

Until now, digitization has mainly been driven by research and funding opportunities. This includes:

- Digital "loan" requests from researchers, where the museum provides pictures of the material requested, and store images together with specimen records in Specify.
- Digitization on request through the Synthesys program.
- Digitization of type specimens. Zoological and Botanical type specimens have been digitized with images and label transcriptions.
- Digitization of material from selected geographical areas.
- Most of the early digitization effort was focussed on the type collections of the museum and a recent donation from a private foundation with a particular interest in Greenland made it possible to digitize the entire Greenland herbarium of higher plants (141.000 herbarium sheets) in the Netherlands (by Picturae). The funding included high-resolution images and transcription of labels.

In summary, the digitization effort at the Natural History Museum of Denmark has been mainly *ad hoc* and/or opportunistic. This will change radically when DaSSCo starts, as one of the main goals is to establish digitisation plans for all major Danish natural history collections in order to kick-off a comprehensive Danish digitization effort. By fall 2021 a steering committee will be established, and further staff will be hired soon thereafter.

<u>Estonia</u>

EMNH:
We have most of the collections digitised and the new material that comes into collections will be digitised during 2 years ideally.

<u>France</u>

MHNH:
Past digitisation efforts at MNHN Paris have so far mostly resulted from the aggregation of a variety of opportunities and guidelines mostly defined at the level of specific collections and curators rather than they are the outcome of a global, institutional policy. Our Museum is, however, willing to establish a written, strategic and comprehensive plan for future computerisation and digitisation plans. For this purpose, we are most interested in sharing best practices, strategy documents and future plans with

other taxonomic facilities. This being said, past computerisation and digitisation campaigns at MNHN Paris over the past 25 years have led to establish a first set of best practices and priorities as detailed below.

Netherlands

NMB:
Practice in the past has been to digitise existing collection where possible and new acquisitions as soon as possible. An official digitisation strategy is in the process of being formulated.

Norway

AUMN:
Botany: The largest part of the collection (98%) is digitized and accessible through GBIF. All specimen information is registered in MUSIT (the current Norwegian collection management system). Some subcollections are still unregistered. Nordic and arctic vascular plants and Nordic macroalgae are mostly photographed, but not all accessible yet. There is currently no plan for photographing new material and handling and storage of images. A large amount of material (mostly fungi and lichens) awaits both curating and digitization.

Zoology: A large part of the collection is digitally registered and accessible through GBIF. We are currently in the process of migrating and publishing the last datasets. Photographs are only available for very few specimens where photography was requested for research or from external stakeholders.

Geology: Approximately 50% of the geology collection is digitally registered in a local database which is currently not accessible externally. Pictures are taken for some objects.

Spain

MAH:
Left blank.

UCME:
There is no strategic plan for the digitization of entomological material from our UCME collection.

2. Prioritisation criteria employed for digitisation which has been done or is in progress.

Belgium

BE-RBINS: Left blank.

Denmark

NHMD:
With an estimated 19 million objects in the Danish Natural History collections and their tremendous difference in size and conservation methods, it will last decades to digitise it all. Accordingly, we consider a couple of decades, roughly corresponding to the estimated minimal lifetime of DaSSCo, a realistic estimate, but this obviously depends on funding. To ensure optimal impact of this effort it is needed to prioritize the digitisation meticulously.

Key criteria for the strategy will be:
*) National collection strength
*) Research and public relevance
*) Digitisation costs and volume
*) Established international policies and archival formats
Prioritizing the Danish collections will be of national importance as it contributions to the new Danish national species portal ("Arter" - https://arter.dk currently being developed by NHMD and the Environmental Protection Agency of Denmark.

Generally the "2-D collections" viz. pinned insects and herbarium sheets, are easy to digitise whereas the whale, the wet and the Geological collections,
are challenging. Presently, automated setups are available for digitisation of herbaria and pinned insects. Thus, such collections are amenable to digitisation both in a relative cost-efficient manner and in large volumes. However, we will during the fully funded, first five years of the DaSSCo's infrastructure's expected 20+ years lifespan start digitising all collection types. We will, in collaboration with other DiSSCo members, establish a range of digitisation procedures aimed at becoming the golden museum standard in the future. Obviously, digitising 19 million objects is a task that stretches far beyond a five-year horizon.

Milestones:
End of year 2021: The DaSSCo Steering group is fully operational and members of the advisory committee are selected. Key staff will be hired for acquiring and implementing key Digitisation setups and establishing the Digitization Laboratory(ies). New staff hired and start-up digitization of prioritized collections.

End of year 2022: 500.000 specimens digitised.
End of year 2023: 1 million specimens digitised.
End of year 2024: 2 million specimens digitised

At the end of the fully funded 5 year period, DaSSCo will have created state-of-the-art Digitization Facilities - An operational, world class collections infrastructure. Additionally, DaSSCo will have established a nationally and internationally important, globally accessible, virtual natural history collection consisting of approx. 5-6 million botanical, geological and zoological objects made digitally available through DiSSCo and GBIF. The digitizing setups and human skills acquired during the funding period will establish the routines needed for DaSSCo to expand the corpus of digitized objects including newly acquired material.

Estonia

EMNH:
Almost everything is digitised, in progress is the last 2-3 years collected material.

France

MNHM:

a. **Computerisation on the fly**

One key institutional recommendation over the past years has been that every new specimen that enters the collection – to be used in study or analysis – and any specimen that leaves the institution on a loan (including loans for exhibition purposes) should be computerised so that it can be identified and tracked through online collection databases. Retrospective computerisation does take place as well, but to a far lesser extent. This recommendation is, of course, implemented at different rates, pending on collection size and staff available in each collection unit.

b. **Collection historical catalogues and documentation**

From 2016 to 2018, MNHN Paris digitised all registries, catalogues and other manuscripts or documents used for the management of its collections. This represents almost 690,871 pages (including catalogue cards). For homeworking staff during the COVID 19 lockdowns, this documentation proved extremely useful but was also enriched thanks to collaborative efforts during this period, which resulted in more than 75,000 new transcriptions of specimen records from digitised catalogue cards. It may also be noted that the MNHN Library, together with MNHN Academic Press have digitised all scientific publications from MNHN Paris since their foundation, all of which are freely available in open access on their digital library as well as in BHL.

c. **Types and figures**

During the 2000s, the majority of collection units focused their computerisation efforts on types and figures specimens.

d. **Ichthyology: the pioneers**

Historically, starting in the 1990s, the curators of Ichthyology engaged the complete computerisation of this collection. It remains to date a very-well-documented collection,

both in terms of databases and digitisation/photography: 445,694 specimens have been computerised, consisting of 136,347 entries, while 16,530 have been digitised (either photographs or radiographs). Those resources are accessible through specific channels such as Fishbase or Worms.

e. **Botany**

In 2008, the project to restore the MNHN herbarium building came together with the objective and resources to digitise herbaria specimens held inside the building and computerised at minima. This program was a strong driver to experiment industrial digitisation at the scale of a very large collection in two dimensions. No less than 5,281,258 herbarium sheets were digitised up to 2012.

f. **The e-recolnat project**

Beginning in 2014, the four-year national program e-recolnat was launched, which benefited from a major grant from the Agence nationale de la recherche (ANR). Although botany was the major focus of the program, e-recolnat also provided resources to computerise and digitise other types and figures at MNHN and across other French collections (other museums and university collections benefited from the program).

<u>Netherlands</u>

NMB: Prioritisation criteria for further digitisation are in the process of being defined, as part of the new collection digitisation strategy.

<u>Norway</u>

AUMN:

Digitization priorities have been previously established in “Revita-plan for Tromsø Museum 2007 – 2016” (attached [Revita is an abbreviation for Revitalisering av samlingene]). There it was stated that all collections should be digitally registered.
Apart from the Revita document, there is currently no written digitization strategy available for the natural history collections of the Arctic University Museum of Norway. Prioritizing is up to the respective curator; most curators will digitize as time and resources allow after the following priorities:

1. Digitally registering all objects and making them accessible
2. Digitizing type material with high resolution images
3. Digitizing the remaining material as seems sensible (there might be no need to take pictures of each specimen in cases where the collection holds many specimens of the same species [duplicates of common species])

<u>Spain</u>

MAH: Left blank.

UCME:
The criteria prioritized: Digitize the material left by recently retired researchers. Digitization of all the material that is lent for a study, if it was not previously.

Highlighting relevant topics

1. Do you have a clear overview of the digitisation status of your institution (how many specimens databased, how many imaged, by which procedural standard etc.)?

Belgium

BE-RBINS:
Yes. The digitisation is one of the main tasks of the scientific service of heritage. The digitisation is funded by the DIGIT-4 federal program (about 300k€/year for RBINS) and by specific projects at the Belgian and European levels. RBINS and Africa Museum collaborate in this task sharing staff, equipments, workflows, CMS, …).

Estonia

EMNH:
I have overview about how many specimens are databased. I do not have overview how many of those are imaged and how many of them are applying to the standards

France

MNHM:
We have a clear overview of the computerisation and digitisation
of MNHN collections thanks to query systems which provide access to a unified dashboard supporting global monitoring of collections.

Key figures are as follows:

- Computerisation: 7,877,413 records in MNHN collection databases, documenting 10,655,228 specimens among which 432,827 types
- Digitisation (2D): 6,395,161 specimens

A less detailed view is available on the European dashboard expressed in MIDs[1].

Netherlands

NMB:
Yes, the percentage of digitisation is known, the existing data-entry is in the process of being checked and, where necessary corrected, augmented and standardised.

Norway

AUMN:
So far, digitization does not seem to have followed any institutional prioritization criteria. There is a good overview over the digitization status which is estimated every

year for Kulturrådets report (museumsstatisktikk). However, outstanding tasks are not quantified in that report. There is no specific monitoring of the digitization progress in place.

Spain

MAH:
Yes, we have. MA Vascular Plants Herbarium keeps c. 1 million specimens. Right now the number of digitalized specimens is 775,000, which represents 77% of the collection. Two large groups have yet to be digitized (Compositae and Graminae) because they were recently in use for the *Flora iberica* Project.

Digitization occurs in two ways. 1 / Through the databases of the electronic field notebooks provided by the collectors who donate the plants; 2 / The direct digitalization of the herbarium staff, with the specifications in view, and through Specify 6.8.

At present there are around 310,000 images associated with specimens in the database (31%), although we have another 85,000 images of other specimens that have not yet been processed. The imaging process then reaches around 38%.

All new specimens' donations, as well as all loan returns, are digitized and imaged before being incorporated to the herbarium.

UCME:
We do not have this clear vision of the state of digitization due to the enormous work that remains to be done, we have digitized 53,000 copies of the 4 million that we estimate there are in UCME.
All species are documented with at least one photograph.
On the other hand, all the digitized material is labelled so that it can be found immediately in the collection cabinets.

United Kingdom

KEW:
For the Herbarium with an estimated 7 million specimens, nearly one million specimens have been databased and 800,000 imaged. For the Fungarium with an estimated 1.25 million specimens, around 500,000 (around 40%) have been databased, but few imaged (around 2%).

We have a reasonably good understanding of the digitisation status of the collections but there are knowledge gaps. Most specimens are databased within Kew's Collection Management System's (CMS). However, a significant minority of

Herbarium specimens are being databased in individual project or research databases which will be imported into the central CMS at a later which will increase the percentage of specimens digitised that are reported above. We have a large backlog of datasets to import which are currently being audited. All digitisation projects have the procedural standards documented. Day to day digitisation activities is less documented.
All collections images are uploaded to our Digital

Asset Management System (DAMS) Digifolia. Within Digifolia it is easy to see how many images have been created. Although as some specimens may have more than one image it is less easy to see how many specimens have been imaged although this information can be retrieved querying the backend tables of the database. All imaging is currently performed or supervised by a single team, so all images meet the same
standards. Herbarium and Fungarium imaging standards follow those set by the Global Plants initiative projects but have evolved slightly over time. For non-type specimens we no longer open and image capsule contents. For the Fungarium for non-type specimens we are only image packets/labels and not the specimen themselves.

We are assessing quality of our images against FADGI and Metamorfoze standards.

2. Are you monitoring it? How?

Belgium

BE-RBINS:
Annual reporting of the Scientific service of heritage

- Statistics of the collection management systems
- Staff digitization reporting
- Number of files available online

Semestrial and annual reporting of the DIGIT-4 and DiSSCo Fed programs.
This is not a “one button” process and requires several days of work.

Estonia

EMNH: No.

Netherlands

NMB:
Yes, we have a registrar and collections coordinator that monitor the digitisation process.

Spain

MAH:
Monitoring occurs throughout the entire process, including digitization. Quality controls are established and checked in a general way, and specifically for some records and images.

UCME:
All data inputs are automatically monitored.

United Kingdom

KEW:
Yes. Numbers gathered from CMS’s and DAMS. The number of Herbarium images and digital records made externally available Via the Herbarium Catalogue web portal is also monitored.

3. What is your digitization level: specimen level or higher collection unit level? What are your policies with respect to how much data is acquired (databasing/ transcription of specimen information and/or imaging)?

Belgium

BE-RBINS

- Available data and metada of specimens are encoded / imported in CMS
- imaging of drawers, boxes of specimens (50Mpx / view) (e.g. insects boxes) are imported in collections websites with available inventories
- individual imaging of Type specimens (and rare specimens) with several techniques (micro CT, surface scanning, photostacking, multispectral, photogrammetry) depending from the collection requirements
- Individual digitization on demand for scientific studies or popular science valorisation.
- CETAF collections registry. Development and user.

Estonia

EMNH: I am not sure what you mean by „specimen level“ or „higher collection unit level“. All digitised specimens have at least geographical region and is identified mostly to the species level, but there are some specimens that are on family level.

France

MHNM: Answer to the first question depends on collection scale and related resources (staff, in particular). The levels of computerisation and digitisation vary greatly depending on taxonomic groups. For example, entries into the Entomology collection are massive and can be made in tens of thousands at a time, while loans of several thousand specimens at a time are also frequent. Thus, for the Entomology collection, which holds ca. 40 million specimens, the MNHN is considering the implementation of a protocol for batch digitisation of entomological boxes (rather than individual specimens). In other collections such as vertebrates, mammals or birds, for example, computerisation and digitisation are more likely to be processed at specimen level. As a trend, the smaller the specimens, the more likely they are to be computerised in batches rather than individually.

As to the second question, answers vary again depending on collection sets and opportunities, but also depend on conservation practices. Until recently, computerisation was mostly driven by taxonomy and a purely scientific approach (catalogues would inform on: what, where, when…). But there was little information of use for practical conservation management purposes. For instance, there was no specific metadata to help locate collections in the buildings or on the shelves. These practical indications are now being developed.

In contrast, the MNHN herbarium is fully digitised, but with limited general information on the region of provenance and taxonomy. The original geographical organisation of the herbarium sheets on the shelves was replaced by a taxonomic classification, thus the minimal data associated with the barcode for each specimen written in the database was a general geographical origin and higher taxonomic level. Only 18% of the herbarium sheets are fully computerised, of which half had already been computerised. The ongoing "Herbonauts" [http://lesherbonautes.mnhn.fr/]program encourages amateur botanists to use the images of the specimens to read the labels and enter the information they contain. This citizen science program has been a great success, with an average of 30,000 specimens computerised into the museum's database each year.

Netherlands

NMB: Specimen level. Policies are in development.

Norway

AUMN:
There is no policy with respect to how much data is acquired. Overall, the more data and the more detailed the better. Individual assessment by the responsible curator. Species of interest are digitized in higher detail according to request/interest.

Spain

MAH:
The digitization is done at both specimen and its associated elements (liquids, large fruits or seeds, microscopic preparations, DNA). The specimen information is transcribed in the database as literally as possible, respecting the original information. This information can then be standardized in different fields. The specimen imaging is carried out with a scale and a color chart, trying to show all the information that may be hidden in an overhead view, making several images if necessary.

UCME:
As I mentioned previously, 53,000 species of the 4 million that we estimate there are at UCME have been digitized. All of them incorporate into the database their location and identification data that appear on the typical labels of an entomological collection: collector, date of collection, location, georeferencing and biotope; identification, identifier and date of identification. Plus one photograph for each species entered in the database

United Kingdom

KEW:

Digitisation completed at Specimen level. Most specimens have full transcription. However certain digitisation projects have only digitised a subset of core fields which has been documented. We aim to transition to map levels of transcription to Minimum Information about a Digital Specimen - MIDS. Where projects only have funding for some specimens at MIDS 0 we use crowdsourcing to gain the additional fields.

4. Do you have a unique management software or more than one? What kind of protocol are you using for the data digitisation (e.g., ICEDIG guidelines)?

Belgium

BE-RBINS:
We use a main CMS for data and metadata (DaRWIN, WEB CMS, Open source based on Postgress and Symfony, can be embedded in any institutional website using Iframes) and we are importing data from more specific CMS used in some collections (e.g. paleontology). DaRWIN is exporting data to GBIF and use UUID/permanent link according to CETAF requirements. We use a main multimedia server (Open Source Plone in 2021 but migration to Open Source Collective Access with IIIF viewer in 2022).

Estonia

EMNH: PlutoF for biological and SARV for geological collections.

France

MNHM:
20 years ago or so, the idea of buying a commercial product was rejected by both collection curators and the IT department, who shared the vision that no single product (all the more proprietary) could properly answer the very diverse management and scientific needs of all collections. Using Oracle, the Museum's IT services aggregated numerous databases, by collection, and implemented a home-made interface (JACIM) which is clearly obsolete but still in function. Today, the Museum still holds 19 different databases under Oracle. For two years, work has been underway to unify these databases, piece by piece, table by table: geography, people, location, taxonomy, etc. Like many large institutions, reorganisation processes are cumbersome and time-consuming. The new team in the collections department has set itself the task of completely overhauling our collections information system, both for internal organisation and to meet international demand (the DISSCO project, for instance).

Netherlands

NMB: We are in the process of transferring to new software at this moment. This new software works according to the Darwincore standard. A protocol for data digitisation in this new software program is in development.

Spain

MAH:
Yes, we use the software Specify 6.8 for digitization, although for some processes this software does not meet all our needs and expectations and we additionally use Access or Excel. Data digitisation follows our own protocol, adapting other protocols to our needs.

UCME:
Our management software is unique and open source, based on the open source relational database management system MySQL. That you can reached at: ucme.bioucm.es

United Kingdom

KEW:
We have different databases for different collections. However, we are currently part way through a program to migrate all collections into one database. Databasing should follow in-house manual guidelines. Standards in the manual allow us to map to Darwin Core Archive Standards and vocabularies. We follow certain standards for certain fields such as ISOCountry and TDWG for geography.

5. Do you have a procedure for validating data (e.g., accuracy of identification and georeferenced)?

Belgium

BE-RBINS:

accuracy of identification is verified by curator
georeferencing validation is using external services but the procedure concerning the valitaded/extracted data is not yet defined.
The original data is always preserved even after (re)evaluation

Estonia

EMNH: I am not sure what you are asking here.

Netherlands

NMB: Not yet. A procedure is in development.

Norway

AUMN: No special procedure in place. This is the shared responsibility of the respective collection manager and curator.

Spain

MAH:
Identifications are not validated, except for those loans that are returned within the scope of the *Flora Iberica* project, which provides a systematic and taxonomic reference criterion. The validation of the geographic data occurs at the time of georeferencing, by data projection on a cartographic system.

UCME:
All specimens have been identified by specialists in their branch, who certify their correct identification

United Kingdom

KEW:
No the only validation is that employed on data entry through the CMS. Large digitisation projects have Quality Assurance officers to check data entry and provide feedback to digitisers. Curators are consulted by digitisation staff if there are obvious curation

issues. There is not enough digitisation or data manager staff to employ tools such as coordinateCleaner or look at analysis from data aggregator sites. However, we hope that more validation tools could be added to digitisation workflows in the future.

6. What are you planning to digitise next and what projects are planned for further down the line and why?

Belgium

BE-RBINS:
We continue the Type Specimens digitisation and the digitisation on demand with fast valorisation (e.g. Synthesys VA calls).
The next priorities have to be defined but probably linked to the African collections in collaboration with the Africa Museum. The collections from this continent are under represented in the digital data following the One World collection exercise.
The digitisation could also play an important role in the access sharing with African users / countries.

Estonia

EMNH:

We are digitising all new material (herbarium sheets, pinned entomological collections, stuffings etc). We are planning to digitise photography-slides.

France

MNHN:
The MNHN is continuing its work on types and figures: including maintaining the protocol to computerise specimens entering the collection, used in studies and analyses, and specimens leaving the institution for loans.
Retrospective computerisation has remained on the periphery since the end of the e-recolnat project due to lack of dedicated resources. A new opportunity to promote it is, however, about to emerge as the MNHN was recently allocated a major grant from ANR for a new 3D specimens model program (e-COL+ project). The goal it to digitise 40,000 specimen (vertebrates, marine invertebrates, arthropods) and to develop capacity and national partnerships in 3D digitisation and diffusion, but also to explore AI potential for this material.
MNHN is also willing to develop a long-term institutional program based on a selection of the corpus which would be prioritised and formalised in a multiannual collection development strategy.

Netherlands

NMB: All newly arrived objects (donations) which includes i.a. 400.000 lepidoptera.

Norway

AUMN:
Digitization is done continuously according to the respective curators/ collection managers priorities. There is currently no overall digitization plan.

Spain

MAH:
The plans for the near future are to continue with the digitization of the 23% that is not yet available. On the other hand, imaging continues focussing on new entries, while an effort must be made to work on specimens that have not yet undergone this process.
The complete digitization of the collection allows the integral management of the specimens through the database, improving and optimizing work times and efforts, which currently have to be done in two different ways (digitally and manually).
Any plan in this sense needs funding resources, with extra staff, which is not available right now.

UCME:
Finishing the digitization of the UCME collection is for now our only goal that will still take many years to complete.

United Kingdom

KEW:
Business as usual activities include: digitisation of new types, discovered types already in the collection but not recognised as types before, image requests, digitisation of new Accessions for those Genera completed digitised. When the new CMS is in place for Herbarium specimens it is planned to digitise all new accessions and ensure workflows are in place for all Kew collected material in the field. In addition, we hope to work with Key institutions to transfer electronic data along with the physical duplicate specimens. We will work towards more "born digital" workflows. All Fungarium new accessions are already databased before integration into the collection.
Funded projects currently running include digitisation of selected species of plants and fungi from Colombia. This is part of the larger funded program Useful Plants and Fungi of Colombia. Digitisation of Dianthus through Call one of Virtual Access through Synthesys +, the second call outcome is expected next month.

Proposals have been submitted to digitise the whole Fungarium and Herbarium and we are also seeking philanthropic funding. Priority groups for digitisation are aligned with Kew's Science and Collection Strategies but also are matched to funders/individual donor interests. Priority groups include grasses and legumes and other groups actively worked on by Kew staff. Where possible digitisation is completed in the most efficient way that matches collection storage e.g., taxonomically of a complete Genera or family. Where possible we avoid digitisation by species due to a much larger digitisation cost per specimen

due to time taken to select material however this is not always possible, and exceptions include image requests and project driven needs e.g., useful plants and fungi of Colombia project.

7. If you do not have a defined plan, what are the circumstances driving you to unplanned digitisation actions (e.g., specimens requested for loan, new accessions, specimens involved in an exhibition, etc.)?

Belgium

BE-RBINS:
Digitization on demand: specimens requested for loan, specimens involved in an exhibition, **destructive sampling**
Opportunistic Digitisation: related to scientific projects or new acquisitions

Estonia

EMNH:

There is no need, because we have almost everything digitised and with new material it is ongoing process where every curator is dealing with her/his material.

France

MHNH
We are in the process of reviewing and unifying our information system. The choices that will be made will contribute to the definition of our future computerisation and digitalisation plan.

Netherlands

NMB: Blank

It might also be useful to distinguish between

- Mass digitisation or large scale where indeed the questions of prioritization, feasibility etc are very relevant
- Digitisation on Demand
- Opportunistic Digitisation

Norway

AUMN:
Unplanned digitization actions follow the opportunities that arise, ie. requests from outside, guest researchers that come to visit the collections, specimens used in exhibitions or for media contacts, digitization projects that get funded through external sources.

Spain

MAH:
Already answered above, but: Occasionally, part of the digitalizasion and imaging is done on demand that have increased enormously in the last 3 years also due to Covid-19. These tasks are also done taking advantage when small loans are requested.
UCME:
The examples they put are exactly the exceptions that fall out of the chosen plan. Works in which specimens from our collection are involved, if they were not yet digitized, they immediately become available, as well as those that are chosen for exhibitions.

United Kingdom

KEW:
As mentioned above there are a lot of ongoing BAU activities. Unplanned digitisation might include visitor, intern or student requests for digitisation where we will provide support and training to help them digitise the specimens, they need for their research.

Information put together by Sarah Phillips (Digitisation Manager) and reviewed By Alan Paton (Head of Collections)

Additional notes and comments

Belgium

RMCA, Belgium

Digitisation Strategy and Prioritisation at the Royal Museum for Central Africa

On a daily basis collection specimens are digitised at the Royal Museum for Central Africa (RMCA). An important step in the digitization is a migration of all collection data stored in different mediums (Excel, Access, DataPerfect, FileMakerPro…) to a centralized collection management system DaRWIN. Several techniques are used to digitise the collection specimens and label (metadata) information. The most common are SfM (structure from motion, i.e., photogrammetry), structured light scanning, 2D+ imaging, µCT scanning, multispectral imaging, etc. A handy tool has been created by RBINS and RMCA for Synthesys+ Task7.2. to determine the best digitisation technique to digitise a specimen: https://digit.naturalheritage.be/digit-key/.

As our digitisation specialists report on a regular basis to the head of collections on the amount of digitised specimens or collection drawers, the amount of specimens currently digitised can be easily retrieved. The percentage of the collection that is digitised to date is a bit more difficult to tell as the exact number of specimens for certain collections is estimated, but not exactly known.

At the RMCA the main focus is high quality digitisation at the specimen level. Because of the extensive collection the digitisation policy is to first digitise the type and figured specimens and rare and highly valuable collection items. Besides this digitisation work, a daily pipeline to digitise collection drawers is also operational. This pipeline helps to better understand the number of specimens present, the condition of the specimens, the exact location of the specimens in the collection, etc. This information helps our collection managers and requesters to process/make specimen requests without the need to physically enter the collection.

For each specimen digitised, the metadata accompanying the specimen is entered or checked into our CMS (DaRWIN). Within the CMS a UUID (Universally Unique IDentifier) is created to link the data of the CMS to the digitised imaging/3D data. For each technique also a metadata sheet is filled online. The digital images and digital twins created within the digitisation process are made available through the museum's virtual collections page.

The guidelines used to digitise the collection are those set up by Synthesys3, Synthesys+ and published in Keklikoglou et al. 2019 and Brecko & Mathys, 2020.

The collections currently digitised are the types and figured specimens from the Zoology collections and the precious collection items from Antropology. Aside from the above mentioned daily digitisation, often digitisation on demand is done as well. It includes individual requests, for example request for a loan, as well as requests in the context of research projects demanding the digitization of an entire collection of interest (see Synthesys+ Virtual access projects: https://www.africamuseum.be/en/research/discover/news/virtual_access)

Specimens are digitised prior to a loan or simply to replace the loan by sending the digital data instead. For exhibitions digital copies are sent or 3D prints are made in house to keep

the original specimens in the collection. Figure 1 below visualises the RMCA's digitisation pipeline.

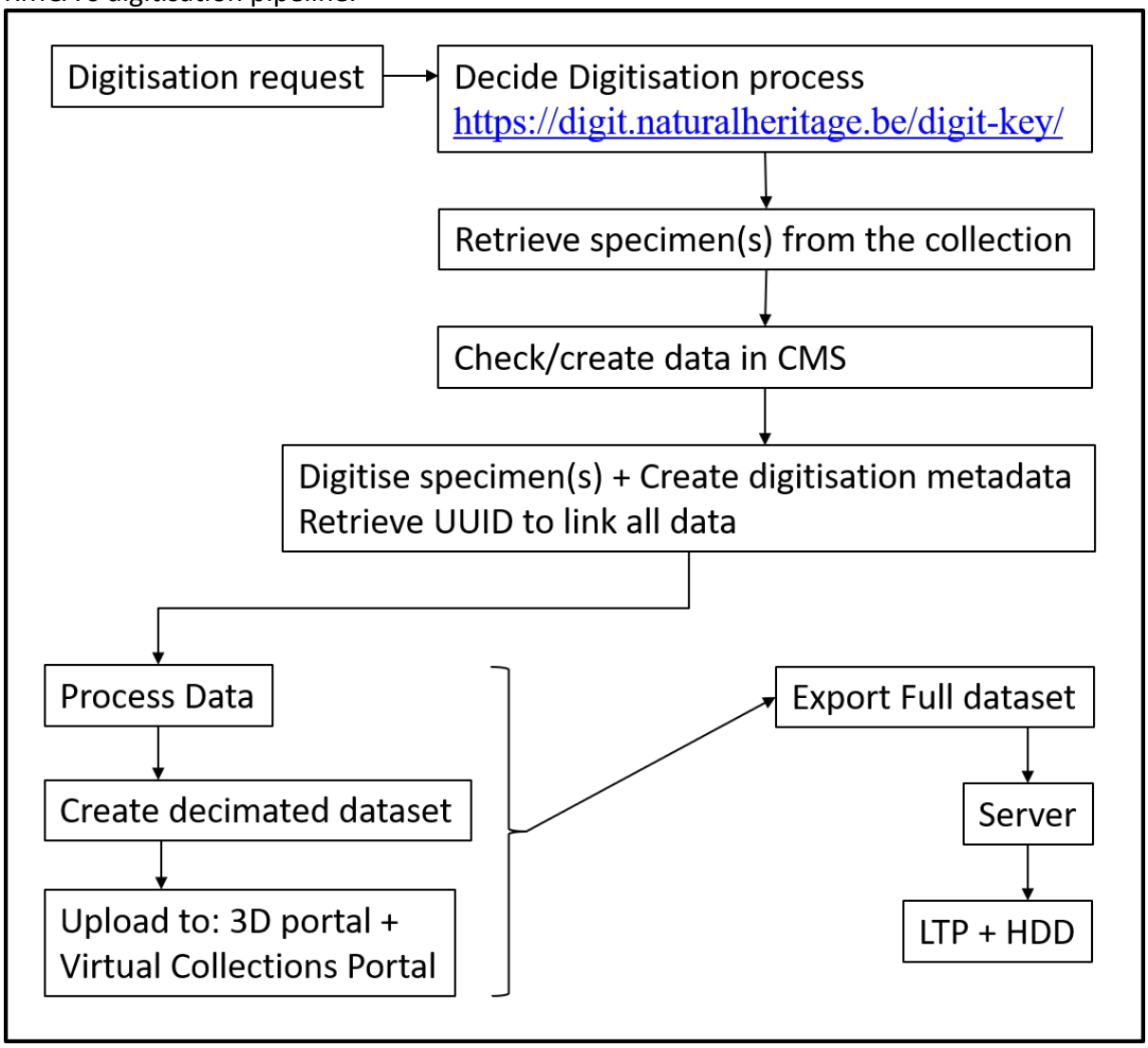


Figure 1: Digitisation Pipeline at the RMCA.

References

Keklikoglou, K., Faulwetter, S., Chatzinikolaou, E., Wils, P., Brecko, J., Kvaček, J., Metscher, B., & Arvanitidis, C. (2019). Micro-computed tomography for natural history specimens: a handbook of best practice protocols. European Journal of Taxonomy, (522). https://doi.org/10.5852/ejt.2019.522

Brecko, J., & Mathys, A. (2020). Handbook of best practice and standards for 2D+ and 3D imaging of natural history collections. European Journal of Taxonomy, (623). https://doi.org/10.5852/ejt.2020.623

Meise, Belgium

Most of our workflows and approach of our mass digitisation projects is published in the following publications:

https://doi.org/10.3897/BDJ.8.e47051 (Designing an Herbarium Digitisation Workflow with Built-In Image Quality Management)

https://doi.org/10.5281/zenodo.3524263 (P 25 case study mass digitisation projects DOE! and DOE!2 at Meise Botanic Garden)

https://doi.org/10.3897/rio.6.e56211 (4.6 transcription information for Meise Botanic Garden's first mass digitisation project DOE!)

https://docs.google.com/spreadsheets/d/13n_3GCyAu8wJZSF6DPRPrvMS01A9hQ7c8H5mODV8IFU/edit#gid=1029684834 (digitisation status of our collection as provided for the Synthesys+ project)

Extra information:

Prioritisation

- Our cryptogamic collection is almost completely databased as the database is used to track the specimens in the herbarium. We are not planning to make images of these specimens as we think it is not useful to have an image of a dried mushroom, a moss or a lichen. An image would only be useful to see the label information and the quantity and quality of the specimen.
- Our liquid collection is almost completely databased for the same reason as the cryptogamic collection.
- Our silica gel collection is almost completely databased and will not be imaged.
- Our botanical curiosities collection is being databased and will be imaged in the future.
- We started in 2004 with digitising (imaging and databasing) all our vascular plant type specimens (for the African Plant Initiative, Latin American Plant Initiative and Global Plants Initiative (GPI))
- Also for the GPI project, we digitised
    - all black and white drawings and their related specimens
    - historical specimens from the Martius herbarium (partially)
    - type specimens of the macro algae collection

-> ca. 100,000 specimens were digitised within 10 years (in-house using herbscans)

Specimens mounted on (A3) sheets were chosen to be digitised (imaging and databasing) first as it is the easiest.

- In 2015 our first mass digitisation project DOE! started. We received a grant from the Flemish Government to digitise (imaging and databasing) our African vascular plant herbarium and our Belgian herbarium of vascular plants (1,2 million specimens).
    - The central African (Congo DR, Rwanda and Burundi) collection was chosen as a priority because 85% of all specimens ever collected in that region are stored in our collection (due to our colonial past).

    - Same for the Belgian herbarium. We have the most important collection for Belgium. For the transcription of these labels we called in the help of our citizens. We therefore developed the DoeDat platform (www.doedat.be). This platform has been built using the Open Source project crowdsourcing platform DigiVol, built by the Australian Museum in collaboration with the Atlas of Living Australia. We have since extended the interface so it can accommodate different languages and made our own code open so that other institutions can set up their own multilingual version of Digivol.
    - All specimens are openly available on our virtual herbarium www.botanicalcollections.be and on GBIF.
    - The digitisation was outsourced to Picturae.
- In 2018 we received a second grant from the Flemish Government to digitise (imaging and databasing) the general collection, the macro algae collection and vascular plants of the Van Heurck collection (another 1,2 million specimens).
    - These collections were chosen because we wanted to complete the digitisation of all our specimens mounted on sheets.
    - The digitisation is outsourced to Picturae.
    - Specimens will be available online by the end of 2021 on www.botanicalcollections.be
- In the future we are planning to digitise (imaging and databasing) in house (using a Pentax 645Z camera with a resolution of 450 DPI)
    - all new incoming material mounted on A3 sheets
    - all exceptions from the second mass digitisation project (specimens kept completely in envelopes, multi gatherings, over and undersized specimens)
    - our microscopic slide collection
    - our fruit and seed collection
    - our botanical curiosities collection (dried collection kept in jars)
    - our wood collection
    - herbarium material kept in books
    - some precious collections (too fragile to put on a conveyor belt)

For now we are digitising our collection on specimen level. Each specimen has its own barcode.

BGBase is used as the collection management system.

We do provide digitisation on demand (for those collections that aren't digitised yet)

Estonia

Estonian National Node, Estonia

The Estonian national node has four partner institutions and two custom built data management platforms for the digitization of specimens: **PlutoF** for biological (public data is

published via eElurikkus and GBIF) and **SARV** for geological data. Both data management platforms follow the major data standards like DwC, EML, GGBN, ABCD, MCL, etc.
We have a clear overview of the proportion of specimens digitized from those which have been assigned a unique id when accepted into the collections. 84% of specimens were made available to the public at specimen level via the platforms as of 31.12.2020. The number of specimens in the collections is monitored by the curators and the number of digitized specimens can be queried from the platforms when needed. The percentage does not take into account specimens which have not been assigned an id. These include some large legacy sets from the last century as well as sets donated by citizen scientists (mostly botanical and zoological sets).
Collection curators and other staff are digitizing specimens daily. The order depends on the needs of the scientists working in the universities. At the moment we are undertaking a project to digitize eDNA datasets and sample data from agricultural, forestry and water realms. Further down the line we would like to focus on continuous data gathering and to automate the publishing process. Today we accept specimens with minimum information like collection time, collecting and identifying agent(s), exact locality, taxon name, habitat information, etc. We also encourage collectors to upload data by themselves because then the data are most accurate.
Most collections have a digitisation plan and results are reported back to the network annually. Our collections share the digitisation equipment like herbarium scanners, cameras and microscopes.
For taxa where DNA sequences are important to make the exact identification, we recommend preparing specimens by following a specific protocol. Specimens with DNA sequences will pass validation analyses and new identification (taxon name) will be added if necessary. Georeferenced data are also validated with scripts written for this purpose.

Finland

LUOMUS, Finland

I am replying on behalf of Luomus as Digitation Manager in charge of the mass digitisation projects.
We don't have a detailed written digitisation strategy. It has been up to the collection teams and those responsible for a particular collection to make the prioritisation on what to digitise. A more coherent digitisation strategy is although in preparation at the moment. We currently digitise only in-house and we have mass digitisation (with images) processes and small scale databasing (some with images but most without). The workflow and prioritisation criteria differ somewhat between these processes. For the mass digitisation the main criteria have been the feasibility and promises to funders (mainly accelerating digitisation rate). For the small scale digitisation there is a variety of criteria; scientific importance (type specimens), out-going loans, accession on new specimens, DOD e.g Synth+ VA call, feasibility, and funding.

Our basic level of digitisation is specimen level. The amount of data that is transcribed depends on the collection, specimens and the digitisation process that is been used. Below is an example from the insect digiline:

- For the Insect digitizing line we do routinely:

  - Pinned insects, mainly Lepidoptera so far
  - Transcription is done at the time of digitization.
  - Methodology: not verbatim; data Is interpreted in the fly (if possible)
  - Data is read, if possible, from the label images off the preview screen of the digi line.
  - Minimum scope: ID, taxon, country (or higher geography), collection ID, record type
    - data created and editor are added automatically
  - Transcription past the minimum scope is a secondary task; it is done if time allows; if not, the record is flagged for later transcription.
  - Direct entry into Excel (which is later imported into our CMS)

- For the small scale digitisation the amount of data transcribed is project/collection specific and usually more comprehensive than in mass digitisation. The same minimum scope as in mass digitisation applies.

Our data goes to our collection management system, Kotka, which is only accessible by the museum staff. However the data is made available and open to everyone through our data portal Laji.fi. We can follow our digitisation status through these systems quite well. Some of the parameters are already generated automatically, mainly basic statistics e.g. Collections, which have their metadata or specimens in Kotka, specimen and transaction counts. For a more specific info we have to use different kinds of search commands. We can get quite detailed info out of the system this way, but it is a bit laborious and we are now developing the system to generate the needed parameters automatically, maybe also a dashboard to visualize them.
The parameters we mostly monitor are:

1. Number of digitised collection samples on the insect line
2. Number of digitised collection samples on the plant line
3. Number of digitised collection samples (other than mass digi lines) by collections
4. Number of digitised type specimens
5. Number of imaged specimens

France

MNHN, France

A single database for all collections sets would be a relief for system maintenance and would aid its evolution.
It is also important to plan from the outset for the security of digital data and to establish an archiving system in order to mitigate risks of data loss and to guarantee long-term conservation to collection data.

In Paris we do not have staff strictly dedicated to transcribing or cataloguing data from our collection specimens. A database manager(s) dedicated to the collections with a team strictly dedicated to computerisation or digitisation would definitely be an asset. This team would be in charge of coordinating the choice and implementation of metadata standards along with the dissemination of best practices through continuing education of staff. Such organisation would improve quality control and data consistency.
The advances in the computerisation and digitisation of our databases have been greatly facilitated and structured by the opportunities (grants) "developed" and seized by the various teams of the institution. This being said, we would need to secure additional, dedicated budgets and staff for digitisation and IT modernisation regardless of these grants so to maintain a sustainable workflow of production of digital collections.
It is mathematically more profitable to start with the smallest collections or well-defined corpus. This is something we did not do, as small collections are often not curated...
As far as digitisation is concerned, the choice we have made to favor type and figure specimens responds to the concerns of our research teams working with the collections.
For very large collections, the decision to favor simple digitisation (image) over computerisation (database) is questionable, at least until OCR systems are more efficient. The productivity of transcription by participatory science is limited.
On the contrary, batch digitisation can be interesting (e.g. insect boxes) as it allows for a wide distribution of specimens that are not necessarily already computerised.

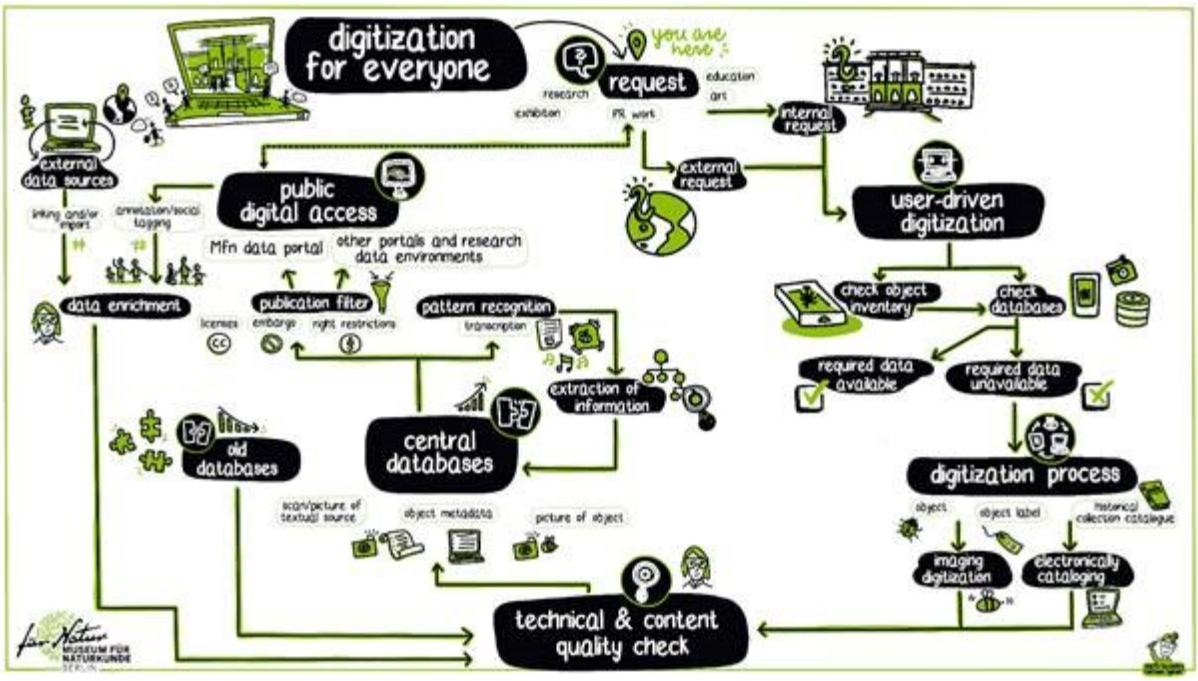
Germany

MfN, Berlin, Germany

Finally some feedback from us regarding your question on the prioritization of digitization at the MfN. We do not have a finalized documentation how we prioritize digitization projects, but here are a couple of lines which are explaining the current process and the idea behind:

- Around a year ago MfN started a huge digitization project (https://www.museumfuernaturkunde.berlin/en/future/collection-disclosure-and-development), embedded in the re-construction of our building
- Within the project the main goal is to set up a digital catalogue of our holdings
- For various object groups (if useful) images will also be taken (differing in technique, details, 3D vs 2D etc.)
- For any (research driven) project with specific requirements (CT Scan, specific images) or detailed information (e.g. referenced collections sites) we set up meetings with the researcher and the collection staff to investigate first of the feasibility of the digitization for this project
- The digitization might depend on various factors comprising but not limited to available funding (for staff but also material), the availability of the collection (access might be limited due to moving activities), collection management staff available, laboratory available (not only DNA, but also 3D digitization facilities, scanning electron microscope.), data management routines established, …
- This is not only true for internal projects but also other third party funded projects (e.g. Virtual access calls of Synthesys+)

- The results from this survey can then be taken into consideration and be included in case we need to make any priorities
- Currently, we are still setting up our internal workflows and bringing our data and media management infrastructures up to date, but the workflow in the illustration below shows the idealized work- and dataflow (its publically available under the DOI: https://doi.org/10.7479/8h2v-4040)



Berger, F.; Glöckler, F.; Hermann, E.; Hoffmann, A.; Hoffmann, J.; Petersen, M.; Quaisser, C.; Schuster, F.; Tata, N. (2021). Digitalisierung für alle / Digitization for everyone. [Dataset]. Data Publisher: Museum für Naturkunde Berlin (MfN) - Leibniz Institute for Evolution and Biodiversity Science. https://doi.org/10.7479/8h2v-4040.

- As soon as we do have our digital catalogue available we are expecting an increase in requests of detailed information and high quality imaging of collection objects, we are therefore in the process of making any decision towards the digitization of a single collection / depth of digitization as transparent as possible
- We are happy to share any documents explaining our digitization and the process behind in detail once they are available.

Italy

Florence, Italy

Digitisation strategy of the NHM of the University of Florence

The Natural History Museum of Florence is part of the broader Museum System of the University of Florence, therefore it falls under the University's administration. The Museum is made up of three main units, which are located in different sites within the city of

Florence. This has resulted in a partial inhomogeneity in the management of the various collections, which have different collection management systems, despite several efforts were made for many years to unify the CMS among the collections. The Museum does not have a specific department devoted to digitization activities, which are carried on by the curators. Each curator follows the protocols, software tools and workflows that best suit the specificity of the collection he/she takes care of. Despite this inhomogeneity, the Museum has always put great effort in cataloguing and digitising its collections (a first digitization plan was developed in the 1980s) and the criteria and policies guiding this are mostly shared across the museum's sections. The main criteria and circumstances driving us to digitisation, today as in the past years, are the following ones:-Availability of the funding and expertise for specific research/cataloguing activities; this, in turn, depends on: the scientific competencies and interests of curators and/or academic staff; the availability of volunteers and external experts; the occurrence of anniversaries, celebrations, etc. linked to specific specimens or collections; the availability of government/local funding on specific subjects.-Digitisation of the papery catalogues compiled in the past, which could be organized highlighting, e.g., the presence of Types, the geographic provenance of the specimens or data related to the acquisition.-Loans and exhibitions (e.g., digitisation of specimens determined by an external expert who provided updated and reliable information; digital tracing of temporarily outsourced specimens; etc.).-Accession of new specimens(especially when directly collected during Museum research activities).-Logistic feasibility of the digitisation and in particular: the accessibility of the collection and/or of the papery or printed catalogues; the possibility to digitise a complete collection, from one end to the other, with limited time and economic investment; and the possibility to digitise from remote (especially during the last year of pandemic) for example from papery catalogues.-Requests from scholars and other external users.-Specimens which particularly contribute to the economic heritage of the University of Florence. It must be considered that the listed criteria may have different relevance in the different museum's sections. We are often carrying on opportunistic digitisation and digitisation on demand. We have never performed mass digitisation (as defined in DiSSCo D8.11). Nevertheless, the Museum of Florence is leading the Italian national network of herbaria (CORIMBO), which has recently developed a detailed project (LEGIT) aimed at the mass digitisation of Italian herbaria; the project has not been funded yet. It is worthy of mention that in Italy there is an official national institute, named ICCD (Central Institute for Catalogue and Documentation, http://www.iccd.beniculturali.it/), which is devoted to the cataloguing and digitisation of cultural heritage, including naturalistic heritage. ICCD provides cataloguing standards and should collect in its general catalogue the records of all Italian cultural assets. It is among the duties of the Museum to send the digitised records of its collection to ICCD. Therefore, we must consider that the output data of our digitization activities should be compatible with ICCD standards and requirements. We are currently working to ensure interoperability between the ICCD platform and both internal DBs management tools and other repositories, such as GBIF and the forthcoming DiSSCo platform. As for the monitoring of the museum digitization activity, from 2018 the number of records sent to ICCD was included in the Annual Report of the museum. A much more detailed and complete dashboard was prepared in the last few months (triggered also by the DiSSCo-related activities) and is currently being finalised. The dashboard provides an overview of the digitisation status of all the collections of the museum and includes information about the level of digitisation, the presence of images in the digital

records and the public availability of the digitised data. About 17% of our Natural history collections are digitised but, of course, the situation is very variable from one collection to the other. For example, the Litho-mineralogy collection is almost completely digitised (85%), both the Geo-paleontology and Zoology collection are digitised at about 40%, while only 6% of the numerically much larger collection of Botany is digitised. Regarding the MIDS (as defined in the blueprint for DiSSCo), the majority of specimens are digitised at a MID1 level (Botany, general collection, and Zoology, invertebrate collections) or at MID2 level for Zoology (vertebrate collections), Litho-mineralogy, Botany (Types), and Geo-paleontology, the two latter also having a remarkable percentage of MID3 level digitised samples. Images are rarely present, except in the case of the herbarium. An option currently under evaluation in the planning of future digitisation is to implement the number of records with a low level of information instead of increasing the level of information of already digitised records or of new records entry (i.e more records at low MID level vs less records at higher MID level).The digitised data are not available to external users, excluding parts of the herbarium and of the litho-mineralogic collection. The digitisation trend for each collection will be included in the dashboard and this monitoring tool will be helpful to define a digitisation plan by assessing the field where to allocate available funds, improving the digitisation of collections and leading to a more selected and focused use of resources.

1"By mass digitization, we mean digitizing entire collections or their major distinct parts at industrial scale (i.e., millions of objects annually at low cost (e.g., < c.€0.50 per item), characterised by improved workflows, technological and procedural frameworks based on automation (both hardware and software) and enrichment (link-building)."

Luxemburg

MnhnL, Luxemburg

Up-to-date, no prioritisation criteria for the digitisation of natural history specimens or collections have been formulated or applied at the National Museum of Natural History, Luxembourg (MnhnL). The collections and specimens digitised so far have usually been chosen in a more or less opportunistic way, i.e. following the availability of funds, personnel and know-how or in an accompanying fashion to ongoing curation or research efforts at the MnhnL. As an example, a mass-digitisation project (2D images) of all herbarium samples was realised after funds became available on rather short notice (within several months). At that point the herbarium specimens were chosen due to them being easily manageable, packable and shippable. They were transferred to an external service provider able to image the specimens within a short period of time (weeks). For future mass-digitisation projects the prioritisation would follow an assessment of "curatorial readiness" of the specimens but also an evaluation of the risks of specimen damage during such a project (e.g. pinned material).
For other in-house digitisation projects (small imaging projects), easily manageable specimens (e.g. microscope slides, dried non-pinned insects or fossils) are currently prioritised due to their ease of handling and imaged by non-domain expert staff. Availability of qualified staff or staff training would shift the priority to more "vulnerable" specimens.

A unique collection management system is available at our institution (Recorder Collection Module). All data (occurrence data, collection management data and media files) on specimens kept at our institution are managed in this central database; thus, information on how many specimens is databased or imaged can be extracted. However, calculating the percentages of digitised specimens regarding the total number of specimens remains rather difficult. This holds true mainly for domains like invertebrate zoology, where the total number of specimens present in our collections can only be estimated, due to lack of a detailed inventory of the several million specimens stored. The progress of digitisation is currently not monitored in an automated way (dashboard or other). Once a year an estimation of newly databased specimens or number of images attached to specimen records is extracted from the database.
Digitisation is usually done on a specimen level, currently no systematic higher collection unit digitisation is done. Opportunistic efforts have been made to digitise bulk specimens (e.g. wet specimen jars) or boxes/drawers during accession of specimens. Higher collection unit level digitisation will be started in the coming years prioritising those units that can be considered having a high "curatorial readiness" e.g. butterfly collection.
Digitisation is not done according to a particular protocol.
In general, a priority has been the extraction and registration of metadata (occurrence data) rather than digitisation in terms of imaging. As a consequence, many botany and invertebrate zoology specimens are registered in our database with precise metadata but are lacking images (with the mass-digitisation of the herbarium this will be completed for botany).
Creation of digital records of specimens usually involves the import of as much and precise data as possible. Metadata related to specimens (determination, people, locations, curatorial information etc.) is usually captured as precisely as possible before being validated in collaboration with curators of the concerned specific domains. Creation of records with only very basic information, with the intention to complete the records at a later stage is usually not done.
Currently no clear plan has been defined on what to digitise next in terms of mass-digitisation projects. In general, we will prioritise digitisation of specimens easy to handle for in house digitisation and intensify efforts to digitise specimens in parallel to ongoing research or curation projects with the resources available to us.
Funding opportunities might lead to larger currently unforeseen digitisation projects in the future.
Clearer guidelines on how to handle digitisation/loan requests, new accessions or specimens used in publications need to be established and communicated.

Netherlands

NHMR, Netherlands

DiSSCo Digitisation Prioritisation Natural History Museum Rotterdam
Bram Langeveld curator Natural History Museum Rotterdam, Rotterdam, the Netherlands.

Current digitization level is estimated at just over 85% of the total collection of the Natural History Museum Rotterdam (NMR) on a sample level (which often is: specimen level in vertebratesand fossil vertebrates, Insecta and Plantae and lot level (more than one

specimen with identical data per sample) for most other groups, e.g. mollusks):c.357,000 digital records. Specimen data is digitized as much as possible (Table 1). Images are not routinely included, but still there are over 20,000 digital images associated with the specimens. The focus of digitization efforts is always on keeping up with current acquisitions (c. 150 annually, varying from individual specimens to entire private collections numbering thousands of specimens) of scientifically valuable specimens/collections as much as possible, and working on backlog whenever possible. Digitization strongly depends on a small number of volunteers as well as museum staff. C. 80% of the backlog consists of a collection of microscope slides with specimens from subclass Acari. These have not been prioritized in the past and must be suitable for some form of mass-digitization. Objects with rich associated data and thus a high scientific value are prioritized in digitization of the backlog; here, the work is often concentrated on the most scientifically valuable specimens of a certain taxonomic group before moving on to another group. Sometimes, (external) taxonomists are included to verify or perform identifications as part of the digitization effort. Specimens are digitized in a FileMaker database that was developed over the past decades in the museum. This database largely follows the Darwin Core standard and hence facilitates easy data exportation to GBIF and thus guarantees a broad audience and easy access to selected specimen data for all users from anywhere on earth. Digitization protocols are documented in an internal document/guide to the collection database. There are currently no standardized procedures for validating data (e.g., accuracy of identification and georeferencing); these are assessed irregularly manually from random samples of newly digitized specimens. Also, many suspicious coordinates are automatically flagged by GBIF after uploading the data, which can then be altered. There is no definite plan/schedule for future digitization efforts, mainly due to the unpredictable nature of scientifically valuable acquisitions of collections in the future, which always receive prioritization in digitization. The Algae collection as well as various invertebrate groups which have not been digitized at all yet, are however likely to receive attention whenever possible. Furthermore, researchers requesting access to specimens/use of specimens in exhibitions may lead to some digitisation on demand.

Table1:Data that are digitized by NMR per digital record whenever available.

Acquisition date	Event remarks	Organism remarks
Acquisition number	Ex collection	Other catalogue numbers
Additional notes	Habitat	Owner of sample
Basis of record	Habitat	Pelage/Plumage
Cause of death	Identification (en hogere taxonomie)	Preparation date
Collecting date	Identification remarks	Preparator
Collector	Identifier	Preservation
Condition	In exposition	Sampling
Count	Item/Object	Sex
Date identification	Lifestage	Storage
Date non-ISO	Locality (en hogere topografie)	Type
Date removed	Location remarks	Type status
Disposition	Measurements	

Donator	Occurrence remarks	

Naturalis, Netherlands

NATURALIS MASS DIGITIZATION (2010-2015)

In 2010, following the organizational consolidation of several Dutch museums (Natural History Museum Naturalis, the National Herbarium, and Zoological Museum Amsterdam) under Naturalis Biodiversity Center, Naturalis embarked on a massive effort to digitize its combined 37 million heterogeneous specimens. The €13M,5-year effort was funded bythe Dutch Ministry of Economic Affairs and Finances using money earmarked for strengthening Dutch infrastructure [Fonds Economische Structuurversterking (FES)], resulting in Naturalis' FES Collection Digitization (FCD) program. The paper documenting, he effort and its prioritization criteria is attached.

OUTCOME Twenty-three percent of Naturalis' entire collection was made digitally available in detail, and the rest of it on a metalevel.
●eight million+ specimens were digitized at object level
●~30 million specimens were digitized at storage level(drawer, box, shelf)
●a permanent digital infrastructure (PDI) was established for the digitization, management and accessibility of the Naturalis collection including procedures and protocols [e.g., unique object identifier (QR code), uniform registration codes, registration equipment, uniform data entry policy, and development of a collection registration system (CRS)].
GOVERNANCE
●Steering committee, oversees scientific quality ofthe project
●Project leader, process owner, team leaders responsible for everyday work including project set up to hiring staff, from housing to planning of collections to operations control, from budgeting to decision preparation and execution.
●Total staff: ~70 people
●Partner institutions (Paris, London, Finland, Berlin) were visited to help define best practice

MODE OF OPERATION
Based on past experience, the average digitization cost was estimated to be approximately €5/object, while the FES budget (€13Mto digitize approximately 7 million objects) only allowed for € 1.86/object including overhead, permanent storage, and equipment costs. Therefore the following decisions were made:
●to digitize a large number of objects through an industrial approach that facilitated digitization of all types of natural history collections at specimen and storage levels. This approach manifested as nine specialized digitization streets or 'digi-streets' which were specialized digitization production lines for each collection type [herbarium, microscopic slides, entomology,2D (books, journals, etc.), dry vertebrates, geology, alcohol, mollusks, wood)].
●to collect only basic metadata associated with an object, which could be later amended.

●prior to the implementation of each digi-street, a pilot was carried out to explore and develop the technology and work processes to inform the goals and budget set for each street.

PRIORITIZATION CRITERIA

Available financial resources were critical in determining prioritization. Not all 37 million objects could be registered and digitized. It was also not possible to digitize all 7 million objects in-house. For these reasons, the FCD Steering Committee made broad choices based on commercial, quantitative calculations (what costs one million objects to register yourself), budget analyses, and on the consequences for the distribution of numbers and costs per year (and cumulatively) for the organization.

Starting Criteria

●Develop a digitization prioritization framework
●Develop digitization processes specific to collection type (alcohol, dry, microscopic slides, printed publications)
●Collections not extensively used at present, or for which mass digitisation technologies are not yet available or too expensive were digitized at a high-level(drawer, box, shelf) in a manner as descriptive and detailed as possible.
●Divide complicated and labor-intensive processes into several shorter tasks, each executed by an individual specialized in that task
●Standardize data entry by using one metadata standard and central data management system
●Digitization at object level included registration of label data, its storage unit, and storage location. Most digi-streets also made a photo.
●Register only metadata relevant for collection management and accessibility
●Photograph only specimens for which value is added
●Use (commercial) third parties for digitization where beneficial in terms of cost/benefit.

Outsourcing Digitization Criteria

●Value-for-money
●Industrial-scale digitisation technologies exist for the candidate collection
●Collection can be safely moved to the service location at a reasonable price.

The most obvious example of collections meeting these criteria is herbaria.

Prioritization Framework

●Phase 1: Scientists/collection managers submit proposals for digitization including a description and its benefit to current research and collection preservation. Department level proposals are discussed in a plenary session.

●Phase 2:
●FCD Core team evaluates proposals for feasibility, quality and consistency with institutional policy.
●Proposals are scored by panel considering policy and operational necessity:
    ●institutional research priorities
    ●institutional public / education programs
    ●national / international biodiversity projects
    ●European funded / co-funded projects

- ●economic importance of proposed collection
- ●availability of existing collection documentationand data
- ●physical state of proposed collection
- ●Proposals are rated by expert stakeholders.

An online survey of stakeholders rank each proposal with respect to professional and personal considerations in their field of expertise.

●Phase 3: Results of Phase 2 are compiled and prioritized and presented to the steering committee for final decision. The transparent working method, objective selection criteria, and the involvement of direct stakeholders in prioritizing the projects increased support for the choices made among the employees.

Naturalis prioritization and digitisation protocols

Prepared for: DiSSCo Prepare Project D1.3, Establish relevant criteria to identify a prioritization model for digitisation
3 jan 2022

Written by: Tina Loo, ISBI Senior Project Officer (tina.loo@naturalis.nl)
Interview: Steven van der Mije, Naturalis Department Head (in Collections Department), 15 dec 2021 Email: Myriam van Walsum, Naturalis Information Specialist (Collections Information Sector), 7 apr 2021

Prioritization

Naturalis Biodiversity Center currently uses a digitisation prioritization model originally developed by the Smithsonian National Museum of Natural History (Move the Dots), and later adapted and employed at the Natural HIstory Museum (NHM) London (Join the Dots). Naturalis uses the NHM model and calls it CollectieScan.

The model captures the qualities and needs of discrete, collection-manager-defined collection units whose condition and digitisation level are recorded, as well as their (scientific, socio-economic, cultural) importance and information value that is evaluated to the best estimation of the collection manager/curator. The CollectieScan process occurs annually and results in collection unit evaluations for Condition, Importance, and Information and Outreach scored across 16 criteria. The collection unit scores are compiled and prioritized in a report that is reviewed and approved by sector and department heads and the museum management team who further evaluate cost, feasibility, quality, and consistency with institutional policy. The final report is used as a basis for submitting collection sector project proposals, although proposals not derived from this process are also possible. The projects generally have a digitisation component which is often a significant part of the workload. Therefore, a dedicated Digitisation Team, currently consisting of three permanent staff members, has been developed to assist collection managers in project and ad hoc digitisation. The Digitisation Team is trained to handle the museum's disparate collections, associated tools and equipment, and digitisation workflows.

Prioritization is higher for those collection units of greater scientific, socio-economic, or cultural importance, however, assessing importance at the regional, national or global levels is the most challenging part of the prioritization process. Getting a high-level overview of

needs and initiatives is difficult as there is no systematic means of evaluation which would optimally include a core group of stakeholders, experts, specialists, standards and benchmarks. At the museum level, collection unit digitisation that's part of a project proposal that furthers the tangible objectives of

Naturalis' annual year and budget plan receives a higher priority, and useful and important externally funded projects also receive priority.
There is no specific funding allocation for digitisation. The Digitisation Team is paid from the operating budget, and digitisation that is part of an approved collection project is funded by project funds.

Digitisation
Numbers: As of October 2021, Naturalis' entire collection of approximately 43 million specimens had been registered in one form or another to at least MIDS-0. A majority of this specimen and storage level registration was accomplished in Naturalis' mass digitisation effort (2010-2015) following the organizational consolidation of Naturalis with another Dutch natural history museum and herbarium, and prioritization strategies associated with this effort were reported here earlier (8 million objects registered at specimen level and 30 million at storage level). Most digitisation now is done on a small scale, collection managers entering 10-20 records.
The current digitisation effort focuses on bringing the metadata up to object level. As of October 2021, Naturalis had approximately 8.6 million individually registered specimens or samples, 171,412 registered storage units (drawers, jars, etc.), and approximately 5.5 million multimedia records consisting not only of specimen images but also pdfs, audio files, etc. Specific numbers regarding specimens digitised per taxa, by Dutch province, by country, and other details can be found on Naturalis' Digitisation Dashboard.
Data Quality and Monitoring: Issues with data quality have arisen with respect to the reported digitisation numbers. During the mass digitisation initiative, registration methodology between different taxonomic registration efforts was not standardized, particularly with respect to counting specimens, resulting in inconsistent counts, e.g., Is a jar containing six fish one specimen or six? Are the separately registered skin and skull of one mammal one specimen or two?
The lack of standardization also led to inconsistencies in data entry. CRS was in its programming infancy then, and the use of fields was not yet clearly defined nor standardized with pull-down pick-lists. For example, the 'country' location may have only been input in the (Verbatim) Locality Text field or, in a separate Country field with thesaurus, or both; and (for a Dutch species), the country could have been input as NL, the Netherlands, Netherlands, Nederland, etc. These inconsistencies make current data access and retrieval highly unreliable and inconsistent. To optimize the utility and reliability of CRS, the standardization issues, as well as developing methods and systems ensuring data quality (including georeferencing) need to be addressed by someone with oversight responsibility. High level monitoring for performance and quality assurance is currently only performed at the project level.

Data Acquired: Specimen registration levels vary with project need but a minimum level of data entry (collection date, species, location, collector) is usually standard. Images are taken when the extra time needed is justified based on the required photographic quality and use

of the photo. Imaging is more likely with special collections, e.g., types or 17th century objects. However, the consistent inclusion of an image in the specimen record can depend on the workflow, that is, whether registration occurs from specimen label data or a photo, in which case, the specimen is consistently imaged.
CMS: Specimen registration data is captured in two parallel collection management systems at Naturalis: (i) a custom-developed Oracle (Centrale Registratie System (CRS)) for zoologic, paleontologic, mineralogic and petrologic specimens and (ii) Brahms (University of Oxford, May 2019) for botanic specimens. These two data sources, together with the Dutch Species Register (Nederlands Soortenregister) listing and describing all Dutch species, the Dutch Caribbean Species Register, and the Catalogue of Life, form the query basis for Naturalis' Document Store which is accessible via Naturalis' BioPortal. It is anticipated that both Brahms and CRS will be used for the foreseeable future, and that custom improvements will continue to be implemented in CRS.
Currently, registration is not outsourced but that option is always within the realm of possibility for some collections depending on feasibility, efficiency and costs.
Planning: Specimen registration generally occurs according to a planning calendar based upon the accepted prioritized collection project proposals discussed in Prioritization above. However, during the pandemic period, digitisation projects that are/were suitable for executing at home, e.g, registration from a card index or catalogue, are/were necessarily completed before higher priority projects requiring museum access.
Ad hoc registration generally occurs when collection objects are handled by staff, for example when material is loaned, studied in-house or handled for preservation tasks. However, since the number of these specimens is small, this protocol is more routinely followed when the object at hand belongs to a collection for which some registration has already been accomplished, so that registration effort can focus on aggregating meaningful resources of data in place of registering small, fragmented numbers here and there.

References

The enclosed references are taken from NHM document portal regarding its Join the Dots collections prioritization system ("Join the dots" collections assessment exercise).

- Giles Miller (2020). Dataset: "Join the dots" collections assessment exercise. Resource: Collection Unit Definition and Guidlines. Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/0043731.
- Retrieved: 12:14 03 Jan 2022 (GMT)
- Giles Miller (2020). Dataset: "Join the dots" collections assessment exercise. Resource: Join the Dots Collections Assessment Exercise Manual. Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/0043731
- Retrieved: 12:24 03 Jan 2022 (GMT)
- Giles Miller (2020). Dataset: "Join the dots" collections assessment exercise. Resource: Join the Dots Scoring Tool. Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/0043731
- Retrieved: 12:22 03 Jan 2022 (GMT)
- Giles Miller (2020). Dataset: "Join the dots" collections assessment exercise. Resource: Cabinetry Scoring Document. Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/0043731
- Retrieved: 12:27 03 Jan 2022 (GMT)

Spain

MACB, Spain

At MACB, digitization of the collection is the main priority in our current action plan. We estimate that we host around 130,000 specimens, from which 115,000 already have an accession number for the database (and have already entered the workflow described below) and more than 90,000 are publicly available at the GFIB Spain node.
https://www.gbif.es/coleccion/herbario-de-la-facultad-de-ciencias-biologicas-de-la-universidad-complutense-de-madrid-macb/

Specimens pending to be digitized come from different sources and may arrive at different stages of preparation, so they may need to join the workflow at different points, but in general, the protocol we follow works as follows:
- Specimen mounting. This might be necessary for specimens arriving in exchange from other institutions. Most specimens are ordinary dried and pressed plants that are mounted on herbarium sheet with archive-grade glued tape or diluted hemicellulose.
- Preparation of a batch. Specimens are digitized in batches of 20-40 specimens so that they are not lost in the middle of the process.
- Accessioning. Each specimen is assigned a unique accession number, that is printed on the label and in our registration books.
- Imaging. We photograph each specimen using a basic stand and digital camera. This imaging is not aimed to obtain high-quality images that can be shared on GBIF, but to accelerate the metadata digitization process. See below our policy to obtain high resolution imaging.
- Digitization. We digitize the label metadata of specimens on Elysia, the main GBIF-supported software for biological collections. This step is undertaken in sessions with several people entering data simultaneously in different terminals. We do not need to have the specimens with us for this process, since we only use the digital images obtained in the previous step to transcribe the labels.
- Freezing. As part of our protocol to prevent pests in the collection, batches that have been digitized are frozen at -20 for a week.
- Placement in the collection. The specimens are then transferred to the collection, where they are classified taxonomically and alphabetically.
- Data upload. We periodically upload the updates in our local databases to GBIF Spain.

Digitization is constant in our collection, since there is a significant backlog dating several years. There is not a defined priority in the specimens we digitize although eventually we might speed up the process of some specimens that might be needed for an upcoming publication.

We own an archive-grade scanner that we use to obtain high resolution images of specimens that need a special treatment. We scan this way specimens upon request of researchers that contact us for a particular project. We do not include this type of scanning as part of our

regular workflow because it would be too time consuming (5 minutes per specimen) to make a realistic digitization plan with this system. Certain specimens (for example, our type collection) that are particularly valuable have been already digitized and imaged with the scanner and are publicly available online.

Madrid, December 16th 2021
Dr. Francisco Cabezas Fuentes (Director)
Dr. Rafael Medina Bujalance (Curator)

MAFH, Spain

The herbarium of the Faculty of Pharmacy MAF contains a very valuable collection representative of the flora of the Iberian Peninsula and, to a lesser extent, of other countries in Europe, America, Asia, Oceania and Antarctica and being a herbarium with antiquity (1892) it has material from classic botanicals.

In 1970 it was incorporated into the *Index Herbariorum*, a worldwide index of 3,100 herbaria and 12,000 associated staff where a total of 390 million botanical specimens are permanently housed. https://www.nybg.org/science-project/index-herbariorum-upgrade/.
The registration of the specimens was carried out manually until 1990 and from that year the digitization of the new incorporations began using a standard database (DBASE). Since 1996, the MAF Herbarium adopted the HERBAR application as a specific herbarium management program (https://www.gbif.es/software/herbar-zoorbar/). This application was recommended by the AHIM (Ibero-Macaronesian Herbaria Association), http://www.ahim.org/html/ahim_marcos.htm) and is the one currently used by all herbaria in Spain and Portugal. The Herbar and Zoorbar apps - the latter used in the National Museum of Natural Sciences - are integrated into the ELYSIA program (https://www.gbif.es/software/elysia/). All data is currently integrated into the international platform for information on Biodiversity GBIF (Global Biodiversity Information Facility), https://www.gbif.es/.

2. Prioritisation criteria employed for digitisation which has already been done or is in progress.
The registration of the specimens is carried out in chronological order, as they are acquired, but it is prioritized in special cases - nomenclatural types, new citations - and at the request of the researchers.
Second, the process of digitising historical collections that have not been previously computerized continues. Finally, the digitisation of the MAF General Herbarium continues.

3. Level of digitisation.
The total number of MAF records at present is 216,914 of which the data of the labels of 39,148 specimens of the general herbarium (MAF-Vascular Plants) have been digitised; 19,289 records of MAF-Lich (lichens), 600 of algae and 7,700 of the Historical Herbarium of P.A. POURRET (18th century). All this represents a total of 31% digitized. We also have more than 350 image files obtained using an EPSON A3 GT15000 scanner.

Digitisation has been carried out opportunistically and in some cases on demand. Only the data originally provided by the authors who provide (collect - with or without coordinates - and identify) the specimens are georeferenced. But we also carry out georeferencing on demand.

4. What are you planning to digitize next and what projects are planned for later and why?
Currently, a digitization and scanning project is being carried out on the existing nomenclatural Types in the MAF Herbarium, both in historical and more recent elements.
We also continue with general digitization, as explained in point 2.

MNCN, Spain

Here is the answer regarding your request for input on digitisation strategies and prioritisation criteria. At the moment MNCN has no mass or large scale digitisation strategy, although we do digitise with the following criteria:

- Digitisation through Synthesys project [databasing and high quality imaging]. This is a priority in MNCN, because we consider the new Virtual Access as the best way to synchronise European Natural Science Collections digitisation and promote open access of European collections data as a whole.
- During the normal collections activities.
    - New accessions [databasing and low-quality images], types are prioritised
    - Digitisation on demand: specimens requested for loan, for research, exhibition and other cultural purposes [databasing and high quality images], if the specimen is not yet digitised.
- As a background process: historical collections are digitised in the long term when the other activities allow it, and there is staff enough [databasing and low-quality images] eg. in Palaeontology less represented taxonomical groups are prioritised.

Every curator monitors digitisation data in their collections, and reports yearly to the Head of Collections.

Sweden

Herbarium GB, Sweden

I have put together some information regarding the framework of the DiSSCo Prepare Projekt (DPP).

1. As digitization strategy our intention is to have an image of ca 70% of all collections within three years from now (herbarium sheets). Remaining 30% are being registered in a database without image (material mostly kept in convolutes). About one third of those are completed.
2. Prioritization of taxa to be registered at moment is due to demands from external financiering.

Our collections and status of digitization.

Herbarium GB has a large collection of Plants, Fungi, Lichens, Bryophytes, Algae and Myxomycetes. Approximately 1 000 000 specimens. Digitization has been going on since 2006 with external financiering for one person working full time. Whenever possible other persons have participated in the digitization as well. To the most part the digitized collections are registered in a database. That is the information on the labels are transcribed to the database with separated information of geography, coordinates, and collector data. Original text is also included. A separate, external financiered project (Global Plants) gave us the opportunity to fully digitize our type collection. That is a high solution image was also made. Since then, we digitize all types this way. In total we now have 23% of our specimens registered (230 000 specimens).

We use the software FileMaker. All specimens have a unique accession number. The FileMaker server and the image storage is secure and fully backed up at IT University of Gothenburg. The data is published at Sweden's Virtual Herbarium. The Virtual Herbarium adds full taxonomy and further georeferenced data. For instance, if only the name of a district is mentioned on the label (which is normal for old collections) a center coordinate will be given and the collection can be located on a map. Different coordinates are transcribed to decimal degrees with reference to original data. Available images are linked to each specimen. All data is transcribed from the Virtual Herbarium to GBIF so that the records should be the same. At moment updates to the Virtual Herbarium are being made manually but we are working towards a system that automatically will update once every week. When that is in function, we will see to that GBIF do the same.

Geographical names are picked from a premade scroll list with international names of countries and each countries name of their provinces and districts. Collector and further notes have been entered in various ways. Monitoring the registration process was difficult before we had a server. When several persons are doing the registration there will always be some 'personal' ways to do so. Even after basic instructions. For example, the thought was that collector was entered with 'original text' as on the label, and then as a standard in a separate field with surname first. Now it is not always complete. Data cleansing can sometimes be very labor intensive. Changes and adds are continuously being made to the records when needed. An image of the label would clarify many things. In all, our data is quite good.

Pilot projects for mass digitalization, that is imaging and optical character recognition (OCR) have been done. Regarding OCR it will not work with old, handwritten specimens. Imaging of specimens mounted on herbarium sheets work well. Both scanning and photography will in most cases create good enough images. Imaging of specimens in convolutes is very time consuming and in many cases not yielding enough information for taxon recognition. Separate projects for certain taxa are a better approach. In place is our own scanner able to do minor batches on demand. We will also soon be able to do macroscopic and microscopic images for close ups.

The Herbarium will move to a new building in the fall 2023. We intend to image all our vascular plant specimens (ca 700 000) in conjunction with this move. Procurement is in progress.

We are thinking about releasing images for opportunistic registration, but that must be put into the future. The next two years will involve a lot of planning in conjunction with the move and mass digitalization. As far as possible we will continue with the ongoing registration process.

United Kingdom

NHM, London, England

I am replying on behalf of NHM London as Digital Collections Programme Manager in charge of mass (and some pilot / smaller) digitisation projects and broader digitisation strategy.

I attach a document which summarises our prioritisation approach from 2018. Prior to this, we had tried running open calls for ideas among our colleagues and using a scoring matrix to assess these, but in practice those scored highly were not always feasible and we moved to a more flexible approach of trying to have a balanced portfolio of mass projects with smaller projects to develop new workflows or services.

More recently, we have included more digitisation in response to user needs, both through SYNTHESYS+ Virtual Access projects and through internal pilots to centralise imaging for digital 'loans' in entomology - we hope to expand this approach to other divisions moving forward and it is proving more efficient than curators taking their own images provided more than 2 images are needed, as well as ensuring that data records are created consistently and released on our data portal.

We don't have a detailed digitisation strategy as we have found it better to have flexibility - for example at present much of our plans depend on which collections the NHM decide to move to the new site that we will build at Harwell Science Campus in Oxfordshire (https://www.nhm.ac.uk/about-us/harwell.html) . Other key drivers of our projects at any given time tend to be funding and partnership opportunities; and which curators have collections that are digitisation-ready, as well as the factors in the document attached.

We do have three recent digital strategy principles which have come out of thinking around the Harwell programme - these are aspirational and intended to set a direction of travel rather than currently fully implemented:

1. Digital from day one - this relates to the Harwell site. It does not mean that every specimen must be digitised before the site opens, but does mean we intend to have a digitisation plan for everything that moves (doing as much as possible during the move or within a defined timeframe thereafter, subject to resourcing), to avoid future backlogs.
2. Digital on acquisition - all acquisitions are given a digital record but this is often above specimen level, e.g. a single record for a large collection of insects. Our aim in future (subject to resourcing) is to move towards centralising these processes so that a digitisation team assesses and captures data and maybe images for new acquisitions before they are put into the collections
3. Digital by default - a direction of travel towards having all our collection digitally discoverable via either Collections Descriptions or specimen level

records. Further access with then be requested digitally in the first instance with digitisation and analysis services available on demand within defined service levels. Physical access will respond to demonstrable need and will therefore be better targeted with less time spent e.g. finding material and assessing its suitability. Our digital loan pilots are the very first steps in this direction - this will also require a step change in the pace of our mass digitisation for discovery, via our Harwell programme and other fundraising.

In terms of measurement, we do have a clear overview although we would always like this to go further - we use this dashboard (this is not published but the link should be open to anyone we share it
with): https://app.powerbi.com/view?r=eyJrIjoiODBkOTBmY2YtMWM1My00Zjc1LTgwMWEtMjUyYzcwOGIyODI0IiwidCI6IjczYTI5YzAxLTRlNzgtNDM3Zi1hMGQ0LWM4NTUzZTE5NjBjMSIsImMiOjh9
Key measures include the number of specimen records on our data portal (visible on the home page www.data.nhn.ac.uk); records downloaded and download events via the data portal and GBIF; and citations of our digital records via GBIF plus onward citations of those papers. These are visible on the first page of the dashboard. Other pages have more team/project measures and indications of our comms activity etc. Please let me know if you want to discuss the dashboard to understand it further.

In addition, to push our understanding further we are currently tendering for some economic consultancy to help us understand the return on investment from digitisation and data mobilisation, and in future also hope to do more research to understand what makes our data more or less usable to a variety of end users.

All our data goes into our CMS (Axiell's EMu - we currently have a live Programme looking at our strategic CMS requirements and tendering for our next CMS), and from there 4 times a week to our public data portal.

Besides the project that I oversee, there is also 'business as usual' databasing and some imaging that takes place continuously among our collections teams or via research - this tends to be driven by team resources and priorities but we now have 'Digi Groups' for our Earth and Life Science departments which work to establish departmental priorities across my team, curatorial teams and occasional internal or smaller philanthropic funding opportunities.

Our Library and Archives undertake a separate digitisation programme that is primarily outsourced and driven by the partnership with the Biodiversity Heritage Library.

If you need further information on any aspects of this, please do let me know, but I thought it was better to keep it simple in the first instance.

DCP Tranche 3 digitisation projects – prioritisation framework

Note: It has never been possible to limit DCP projects to defined criteria e.g. fixed minimum specimen numbers. This document outlines different categories of project and the considerations applied to all of them. We expect that Tranche 3 will continue to support an evolving balance of all project types.

Already has external funding

Project will be prioritised unless it actively detracts from other priorities

i.e. if funding allows for hiring or backfill, project likely to proceed.

If funding is less, prioritisation will depend on digitiser resource plans but likely to be fitted in if possible

Good candidate for external funding

e.g. high impact, has a clear 'selling point', is a clear fit with a funding opportunity - often 'bespoke' rather than mass (may be small numbers of items)

DCP work with initiator to discuss with Development or prepare other funding bids

Prioritisation likely if funding achieved within an agreed time period, otherwise project not likely to proceed

Digitisation 'on demand' / as a service

SYNTHESYS+ 'on demand' funding (when open) will require third party requestor as well as NHM – DCP will work with requestors

DCP are an increasingly efficient provider of e.g. digitisation on acquisition – open to discussion on Science priorities but this is likely to be part of balanced T3 project mix

Pilot activity

Prioritised according to dialogue between Science priorities e.g. Divisional plans and DCP needs & opportunities e.g. to test new kit.

A pilot will be short term – usually below 6 months, sometimes much shorter, and needs to test something new – typically a new or substantially changed imaging workflow, but could also be e.g. new engagement approach (see criteria to right)

Mass digitisation and mass databasing

Prioritisation based on strategic criteria BUT with a heavier emphasis on feasibility – mass digitisation can only be done using proven workflows – which may be applied to tens or more thousands of specimens OR a smaller project that fits an in-progress workflow (e.g. one of many slide collections).

Collections moves are likely to require prioritisation of mass databasing e.g. locations.

Ongoing strategic criteria

Programme seeks a balanced portfolio of these, informed by: Join the Dots; Strategies/plans; standards; governance including Digi Groups; Collections Programme; DiSSCo etc.

1. **Scientific benefit** – may include research; curatorial or conservation benefits
2. **Public / cultural benefit** – e.g. historically unique/significant collections/collectors; link to public programme; link to Library & archives
3. **Feasibility** – project outcomes can be achieved at proportionate time/cost

RBGE Plan template 2020-25

Name of Plan: Herbarium Digitisation Plan 2020-2025

Author (inc Job title):	Elspeth Haston & David Harris
Date:	10 January 2020
Endorsed by Director:	
Date:	

Outcomes	*Please state the organisational outcomes that the work outlined in this plan will contribute to (as per the draft RBGE Corporate Plan)* This plan will assist in the delivery of two of the four RBGE Strategic Objectives*:* *1)* Maintaining/developing our internationally important collections in order to maximise their value as a research, education and heritage resource 2) Providing learning/training in horticulture, plant science and biodiversity conservation to stimulate people to appreciate, understand, and to contribute to the conservation of plants and our natural environment This plan will also contribute significantly to the following Key Pillars of the Science Biodiversity Strategy: Pillar 1: Unlock knowledge and understanding of plants for the benefit of society (a) Discovery Science: Understanding plant and fungal diversity; key priorities: • Technological innovation including large-scale use of genomic data for biodiversity characterisation and monitoring, and establishment of data-portals and data flows to support large scale analyses of biodiversity data and trends. (b) Global Environmental Change: Understanding biodiversity and ecosystem change; key priorities: • Understanding, quantifying and predicting drivers of change leading to biodiversity loss, at scales ranging from individual species to major biomes • Development of rapid-pass threat assessments to prioritise conservation actions and interventions to minimise biodiversity loss

	Pillar 2: Protect and develop the National Botanical Collection as a global resource Collections Custodians: Maintaining, enriching, & mobilising our botanical collections as a scientific and cultural heritage resource; key priorities: • Increasing the number of threatened plant species in ex situ conservation collections to protect against extinction • Digitisation of RBGE's collections to repatriate data and enable global access to the collections to support scientific and cultural research and to underpin conservation planning Pillar 3: Enrich and empower individuals and communities through learning and engagement Skills and Training: Building global capacity in plant biodiversity science, conservation and horticulture; key priorities: • Establishment of a Biodiversity skills centre, mobilising knowledge on biodiversity science, horticulture, practical conservation and sustainability
Primary Objective	*This should be the main objective / focal area of work* To transform a globally significant herbarium collection into a research-ready resource for next generation research
Rationale	*This should include reference to strategic priorities, drivers, outcomes etc* The Herbarium of RBGE holds 3 million specimens which represent nearly 2/3 of the world's plants and fungi, collected since 1697. Their preservation is a legal obligation under the Scotland Heritage Act. However, our role is not only to preserve the collections but to ensure that they are accessible and being used to build the foundation of biodiversity knowledge that underlies so much critical research. Digitised collections are accessible to, and discoverable by, taxonomists, biological scientists, cultural and social scientists and artists around the world. In particular, the specimens are accessible to people within the countries of origin, opening up historical and current data about their country's biodiversity. It is estimated that half of all undiscovered plant species have already been discovered and are held within existing collections. Digitisation will open up these specimens for species discovery. Digitised specimens cannot replace the need for the physical specimens. Destructive sampling of herbarium specimens, particularly for DNA extraction, is an increasing part of biodiversity research.

	Digitisation enables more selective loan requests, and any destructive sampling requests can be based on an informed decision. The best specimen, or even the best specific part of a specimen, can be selected based on either manual visualisation or by image analysis using machine learning to ensure that the most effective material is removed. This is vital when sampling historical specimens for which very little material is available and any sampling cannot be repeated. An image of the specimen enables researchers and citizen scientists to see and record the phenology (flowering, fruiting, etc) state. This enables research into the impact of climate change on biodiversity and also helps researchers plan the most effective time to carry out fieldwork in remote parts of the world. A critical part of a taxonomist's work involves the identification of material. We are now starting to see developments in automated species recognition based on the increasing number of digital specimen images available for the machine learning algorithms. Image analysis is also being used to discover and identify additional species on existing specimens, including leaf miners and fungi. This can help identify the origin and spread of serious outbreaks, such as the Horse Chestnut Leaf-mining Moth. As the specimens are digitised, virtual collections can be created, providing new ways to explore the collections and giving us the opportunity to recreate historical collections. As we link our digitised collections to other online collections and resources within a Linked Open Data framework, we can collaborate in cross-discipline research and gain new insights into a wealth of linked information. A digitised collection will also provide us with accurate information about the specimens, enabling us to build a more informed collections accessioning strategy with more targetted collecting aiming to fill gaps in knowledge. All physical collections face some risk of catastrophic disasters and a fully digitised collection, whilst it cannot replace the physical specimens, can provide a level of data which would otherwise be irretrievably lost. We have identified the 5 key internal drivers for the digitisation of the Herbarium collection: 1. RBGE Scientific Research 2. Global Challenges activities 3. Scotland's biodiversity

	4. Education and training young people 5. Social engagement
Main Objectives	*Provide a broad overview of the main objectives of work to be undertaken over the five year plan linking these to pillars / strategic objectives* 1. To provide digital access to enhanced, curated and linked collections 2. To ensure that a digital representation exists for each specimen in the Herbarium in case of a catastrophic disaster

Page Break

Main Deliverables	*Please list the main deliverables to be undertaken – linking these to our organisational pillars and outcomes* With additional resources: 1. All angiosperm and cryptogam herbarium sheets digitised to MIDS Level 1 (minimal data) with data and high resolution images accessible online 2. Prioritised specimens digitised to MIDS Level 2 (partial data) 3. A digitisation protocol developed for the digitisation of plot vouchers, collections in spirit, carpological and microscope slide collections 4. All specimens from focus areas georeferenced where appropriate 5. All relevant field images linked to herbarium specimens and accessible online 6. Semi-automated workflows developed for ensuring alignment of physical and digital Herbarium collections 7. All relevant collections linked to records of permit documentation 8. All molecular data from RBGE specimens submitted to NCBI linked to an online voucher With no additional resources: 1. A total of 700,000 angiosperm and cryptogam herbarium sheets digitised to MIDS

	Level 1 (minimal data) with data and high resolution images accessible online 2. Of these, prioritised specimens digitised to MIDS Level 2 (partial data) 3. A digitisation protocol developed for the digitisation of plot vouchers, collections in spirit, carpological and microscope slide collections 4. Semi-automated workflows developed for ensuring alignment of physical and digital Herbarium collections 5. A proportion of relevant collections linked to records of permit documentation 6. A proportion of molecular data from RBGE specimens submitted to NCBI linked to an online voucher

Page Break

Year by year breakdown of key activities

Provide a year by year breakdown of the key activities to be undertaken and link these to specified outcomes

With additional resources:

Activity	*Year one*	*Year two*	*Year three*	*Year four*	*Year five*
Migration or core BG-BASE data to Specify	X				
Migration of additional data to Specify		X	X		
Develop & test digitisation workflows	X	X	X	X	X
Set up an improved mechanism for prioritisation of digitisation	X	X			
Digitise prioritised collections (Nepal, cultivated, SE Asia, West S America, SW Asia & Middle East, Scottish)	X	X	X		
Digitise specimens on demand	X	X	X	X	X
Digitise remaining specimens systematically and by request	X	X	X	X	X

Develop semi-automated georeferencing workflows		X	X		
Georeference prioritised collections by geographical area		X	X	X	X
Identify and process existing field images relevant to herbarium specimens	X	X	X		
Ensure future field images are uploaded as part of researcher workflow	X	X	X	X	X
Develop a digitisation protocol for carpological collections			X		
Develop a digitisation protocol for collections in alcohol					X
Develop a digitisation protocol for microscope slide collections					X
Develop a digitisation protocol for plot vouchers			X		
Develop semi-automated workflows for aligning physical and digital specimens	X	X	X	X	X
Develop semi-automated workflows for linking vouchers to molecular data in NCBI	X	X			

With no additional resources:

Activity	*Year one*	*Year two*	*Year three*	*Year four*	*Year five*

	Migration of core BG-BASE data to Specify	X				
	Develop & test digitisation workflows	X	X	X	X	X
	Set up an improved mechanism for prioritisation of digitisation	X	X			
	Digitise specimens on demand	X	X	X	X	X
	Identify existing field images relevant to herbarium specimens	X	X	X		
	Ensure future field images are uploaded as part of researcher workflow	X	X	X	X	X
	Develop a digitisation protocol for carpological collections			X		
	Develop a digitisation protocol for plot vouchers		X			
	Develop semi-automated workflows for aligning physical and digital specimens	X	X	X	X	X
	Develop semi-automated workflows for linking vouchers to molecular data in NCBI	X	X	X		
Measuring success:	*List what we will measure to demonstrate the efficacy of our activities in achieving our outcomes and KPIs (ensure all key activities have a corresponding KPI and all KPIs should include external benchmarking and baseline statistics)* Number of specimens digitised to MIDS Level 1 Number of specimens digitised to MIDS Level 2 Number of specimens with a latitude and longitude					

	Number of field images linked to herbarium specimens Number of herbarium specimens with field image attached Number of specimen records linked to appropriate permit documentation
Specific supporting infrastructure resource:	*List specific infrastructural needs e.g. ICT (link to stated activities and indicate whether in budget or not)* ICT requirements: Digital preservation management Server space and configuration Digital storage Bioinformatics requirements: Data processing pipelines development and maintenance Data portal development and maintenance API development and maintenance Data and image repository development and maintenance
Specific supporting staffing / skills resource:	*List specific staff training / knowledge requirements (linking the needs with stated activities and indicate whether in budget or not)* Technical photography training SQL training Specify training Project management training Data management / data science training
Internal stakeholders	*Input required from other RBGE departments (agreed or not)* Horticultural staff input required: Maintain alignment between Herbarium and Living Collections Preparation of collection data for voucher specimens from living collections Researcher input required: Field image processing Collection data preparation Permit data preparation
External stakeholders	*Key collaborators / dependencies etc* Specify software

	Picturae DiSSCo project collaborators
Risks:	*Identify major risks related to this plan* Loss of digital data & images Separation of Herbarium and Living Collection collection management systems resulting in misaligned collections Misaligned physical and digital Herbarium collections Lack of interoperability between research data management systems and Specify Specify not able to adequately manage Herbarium specimen, silica-dried, microscope slide and molecular collection data Research staff not supplying electronic collection and permit data
Financial return on investment:	*State financial return on investment to RBGE.* Financial return will be in the form of project funding, eg Virtual Access funding from SYNTHESYS+
Social return on investment:	*State social return on investment. Include details of how the plan will make a positive contribution to RBGE and RBGE objectives* Social return will include: Public engagement with online citizen science missions Opportunities for volunteers to engage with elements of digitisation in Herbarium Availability of online resources for education and exhibition programmes Accessibility of biodiversity data for researchers living in areas with more limited access to biodiversity collections, including within Scotland
Future plans	*Provide an overview of longer terms goals and priorities not covered in this plan (and any challenges that have not been met by this plan but would be desirable to be included in time and/ or if resources allowed)* DiSSCo UK