

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

DiSSCo Prepare WP3.2 – MS3.7 Pre-Digitisation Curation

Author(s)

Sofie De Smedt, Ann Bogaerts, Lisa French, Frederik Berger, Rob Cubey, Anne Koivunen, Krisztina Lohonya, Sabine von Mering, Tara Wainwright, Peter Wing, Laurence Livermore

Identifier of the author(s)

Sofie De Smedt: <https://orcid.org/0000-0001-7690-0468>
Ann Bogaerts: <https://orcid.org/0000-0003-3435-2605>
Lisa French: <https://orcid.org/0000-0001-7279-8582>
Frederik Berger: <https://orcid.org/0000-0001-8400-3337>
Rob Cubey: <https://orcid.org/0000-0001-7902-3843>
Anne Koivunen: <https://orcid.org/0000-0002-3475-7971>
Krisztina Lohonya: <https://orcid.org/0000-0002-5313-8674>
Sabine von Mering: <https://orcid.org/0000-0003-2982-7792>
Tara Wainwright: <https://orcid.org/0000-0001-5816-2529>
Peter Wing: <https://orcid.org/0000-0002-8634-8790>
Laurence Livermore: <https://orcid.org/0000-0002-7341-1842>

Affiliation

Meise Botanic Garden: Sofie De Smedt, Ann Bogaerts
Natural History Museum, London: Lisa French, Krisztina Lohonya, Tara Wainwright, Peter Wing, Laurence Livermore
Museum für Naturkunde: Frederik Berger, Sabine von Mering
Royal Botanic Garden, Edinburgh: Rob Cubey
Finnish Museum of Natural History (Luomus):

Contributors

Publisher

Identifier of the publisher

Resource ID

Publication year

2022

Related identifiers

Is it the first time you submit this outcome?

Yes

Creation date

16/05/2022

Version

Citation

De Smedt, S., Bogaerts, A., French, L. , Berger, F., Cubey, R., Koivunen, A., Lohonya, K., von Mering, S., Wainwright, T., Wing, P. & Livermore, L. (2022) Pre-Digitisation Curation Checklist. DiSSCo Prepare WP3 - MS3.7

Abstract

-

Content keywords

Project reference

DiSSCo Prepare (GA-871043)

WP number

WP3

Project output

Milestone report

Deliverable/milestone number

MS3.7

Dissemination level

Public

Rights

License

CC0 1.0 Universal (CC0 1.0)

Resource type

Text

Format

Funding Programme

H2020-INFRADEV-2019-2

Contact email

lisa.french@nhm.ac.uk



DiSSCo Prepare WP3.2 – MS3.7 Pre-Digitisation Curation

Sofie De Smedt, Ann Bogaerts, Lisa French, Frederik Berger, Rob Cubey,
Anne Koivunen, Krisztina Lohonya, Sabine von Mering, Tara Wainwright,
Peter Wing, Laurence Livermore

Index

Index	5
INTRODUCTION	7
Pre-Digitisation Curation Checklist	8
References	9
Pre-Digitisation Checklist	9
1. Detailed inventory of the collection (# of specimens)	9
1.1. Kind of collections	10
1.2 Estimation of the numbers of your (sub)collections	10
1.3 Classification system of the different collections	10
1.4 Assessment of the condition of your (sub)collections	10
1.4.1 Curation status	11
1.4.2 Specifying safeguards for handling specimens	11
1.4.3 Labels readable	12
2. Prioritisation list of (sub)collections	12
2.1. based on	12
3. Supplies	12
3.1 Restoration supplies	12
3.2 cleaning	13
3.3 storage	13
3.4 barcodes	13
3.5 Track and trace	14
3.6 IT infrastructure	14
3.7 digitisation station(s)	15
3.8. Space for infrastructure of external company	15
4. Pre-digitisation curation	15
4.1 Taxonomy	15
4.2 Specimen quality	15
4.3 Barcoding	16
4.4 Storage renewal	16
5. Staff availability	16
5.1 Preparation of the collection	16

5.2 Imaging	16
5.3 Transcription	16
5.4 Quality control of the images	16
5.5 Quality control of the data	16
5.6 Image archiving	16
5.7 Data publishing	16
5.8 IT developer	16
5.9 Project manager	16
6. Cost books	16
6.1 Staff	16
6.2 hardware and software	16
6.3 outsourcing	16
6.4 available resources	16
6.5 resource application	16
7. Decision to do in house or outsource	16
7.1 in house	17
7.2 outsource	17
8. Necessary workflows	18
8.1 Pre-digitisation curation	18
8.2 Imaging	18
8.3 databasing	18
8.4 QC imaging	18
8.5 QC databasing	18
8.6 Decontamination	18
8.7 tracking system	18
8.8 data storage	18
8.9 data publishing	19
8.10 data management plan	19
Pre-Digitisation: General Useful Links	19
Appendix 1: Digitisation Readiness Checklist for National Science Collections	19
Appendix 2: Case study: detailed inventory of the collections for DiSSCo Flanders	19
Appendix 3: Setting Natural Science Data Free: Scoping UK Collections	20
Summary of work	20
Inventory of collections	20
Estimation of your collections	20
Conclusion	20

Appendix 4: Assessment of Microslide Collections at the Natural History Museum, Berlin	21
Appendix 5: Estimation of the numbers of the African and Belgian Herbarium Collection at Meise Botanic Garden	21
Appendix 6: Estimation of collection size at the NHM London	21
Appendix 7: Decontamination of parts of the Herbarium BR at Meise Botanic Garden	22
Appendix 8: Staff List for the Mass Digitisation Project DOE! At Meise Botanic Garden	25
Appendix 9: Tender written by Meise Botanic Garden for the outsourcing of the African and Belgian Herbarium Collections	25
Appendix 10: Quality control procedure of Meise Botanic Garden for the mass digitisation project DOE!	25
1. Method	25
2. Types of errors	25
3. Measuring standards that reflect the acceptance levels (ISO 2859)	26
Appendix 11: ordering supplies for pinned insect digitisation: Natural History Museum, London	26
Appendix 12: Tracking System (Collection Move: Naturalis)	27

INTRODUCTION

This project report was written as a formal Milestone (MS3.7) of the DiSSCo Prepare Project. The following text is the formal description (Task 3.2 and subtask 3.2.3) from Description of the Action (workplan)

Task 3.2. Collate, refine and implement best practices for data mobilisation at the institutional level to develop the DiSSCo plan for data mobilisation and curation pipelines.

How do you best prepare collections for digitisation, digitise them, curate the associated data, publish this information and measure the outputs? What are the options and rationale for different types and sizes of collections, when should this be outsourced and what different project management approaches are most appropriate in this range of circumstances?

This task seeks to address these questions, describing and refining best practices and building on a substantial investment from prior and current projects (MOBILISE COST Action, ICEDIG; SYNTHESYS+ - Table 4) and feeding these into DiSSCo Prepare WP8). Consolidating what is known into a community-edited manual (supported by WP5), and other relevant platforms, WP3 will streamline the reuse and implementation of these procedures and enhance digitisation capacity across the DiSSCo collection-holding organisations.

Subtask 3.2.3 Pre-digitisation curation

Many NSCs are not digitisation-ready, and in some cases, the curation necessary to support digitisation is a much bigger bottleneck to the digitisation process than the act of digitisation.

This subtask will develop a checklist of requirements that ensure a collection is fit for digitisation.

We will outline the steps necessary to improve digitisation readiness, including estimates of required resources that take into account the different scales and speed of operation typically encountered among DiSSCo facilities.

Pre-Digitisation Curation Checklist

Pre-digitisation curation and staging was defined by Nelson et al. (2012) as the first task cluster in the process of digitisation of biological and palaeontological collections. This step is usually viewed as essential to efficient digitisation but often has benefits that extend beyond the immediate needs of the digitisation itself. Nelson et al. (2012) provided a list of curatorial tasks that collection managers reported as opportunities in the pre-digitisation phase:

- inspect for and repair specimen damage and evaluate collection health,
- re-pin or remount specimens and replenish or replace preservatives in containers,
- treat specimens for pests,
- attach a unique identifier (most often a 1- or 2-D barcode) to a specimen, container, or cabinet,
- discover important but previously unknown, lost, or dislocated holdings (e.g. those owned by other institutions or the federal government),
- update nomenclature and taxonomic interpretation,
- reorganize the contents of cabinets, cases, trays, and containers, especially when these are the units of digitisation,
- vet type specimens, and
- select exemplars for digitisation, when that approach is appropriate.

On the website of iDigBio a list of tasks is stated for the pre-digitisation curation step:

- T1. apply storage locator barcodes
- T2. selecting what to digitise
- T3. apply barcodes at collection level
- T4. locate specimens (flag cabinets)
- T5. pull specimens from cabinet (optional: sort)
- T6. curate collection in place
- T7. transport specimens to imaging station
- T8. placeholder to flag pulled specimens
- T9. sort to remove any already imaged/barcoded specimens
- T10. separate specimens needing conservation work before imaging
- T11. apply barcodes
- T12. create skeletal database record

The ICEDIG deliverable 3.6: Best practice guidelines for bulk imaging of herbarium specimens (Guiraud *et al.*, 2019), also mentions a list of pre-digitisation curation tasks:

- Assessing the number of specimens
- Institutional or governmental policy
- Prioritisation
- Decontamination
- Barcoding
- Mounting/restoration/desleeving/unfolding envelopes
- Packaging and transport
- Metadata collection
- Freezing/pest treatment
- Quality control procedures

This milestone report outlines a pre-digitisation curation checklist which allows user to perform a self-assessment to understand if a collection is ready to proceed with digitization. It is divided into eight sections:

1. Detailed inventory of the collection (# of specimens)
2. Prioritisation list of (sub)collections
3. Supplies
4. Pre-digitisation curation
5. Staff availability
6. Cost book
7. Decision to do in house or outsource
8. Necessary workflows

The checklist can be found below. Each section contains check boxes for the user to indicate decisions and logical steps in preparing the collection. Where relevant, useful links and references have been provided. In some cases, where there was not much information available in the literature, case studies have been described and included in Appendix 2 - 11.

Some sections of the checklist are still to be developed, and link closely with other DiSSCo Prepare tasks. These sections will be added to once the relevant DiSSCo milestone and/or deliverable has been completed.

The checklist and case studies will be added to the [DiSSCo Digitisation Guides website](#) by the end of July 2022 (details will be included in the WP3.2 deliverable). MS3.5 describes the development of this site and the intended audience. The Digitisation Guides site is hosted in github, allowing for easy version control, and once migrated, the checklist will continue to be maintained there. This milestone contains the current version of the checklist as of May 2022.

References

Guiraud, M *et al.* (2019). Best practice guidelines for imaging of herbarium specimens. Zenodo. Available at: <https://doi.org/10.5281/zenodo.3524263>

Nelson, G., Paul, D., Riccardi, G. & Mast, A.R. (2012). Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys*: 209: 19-45. Available at: <https://doi.org/10.3897/zookeys.209.3135>

Pre-Digitisation Checklist

1. Detailed inventory of the collection (# of specimens)

Before you start digitising your collection, it is really important to know the content of your collection. If you do not have a clear overview of the type of collections, the numbers, the arrangement and the status of your collection, it is very hard to estimate the work and the costs of your future digitisation projects.

For case studies see [Appendix 2](#) (detailed inventory of the collections for DiSSCo Flanders), [Appendix 3](#) (Setting Natural Science Data Free: Scoping UK Collections) and [Appendix 4](#) (assessment of microslide collections at the Natural History Museum Berlin), respectively.

1.1. Kind of collections

The approach to digitising each collection can be very diverse. Therefore, it is very important to divide your collection into different sub-collections with a similar digitising approach. List the numbers for the different kinds of collections. It is also useful to know if these collections are kept separately or not, as it will influence the workflow.

- Microscope slides
- Skins and vertebrate material
- Liquid preserved specimens
- pinned insects
- herbarium sheets
- 3D objects
- Other

1.2 Estimation of the numbers of your (sub)collections

For each kind of collection type, it is useful to have a trustworthy estimate of your holdings. It will not only facilitate the project planning related to management, staff, equipment and working space but it is also important when you want to outsource your imaging and/or transcription. Mass digitisation of a collection is only cost effective above a minimum threshold of the number of specimens.

For case studies see [Appendix 5](#) (estimation of the numbers of the African and Belgian herbarium collection at Meise Botanic Garden) and [Appendix 6](#) (estimation of the numbers at the NHM London).

1.3 Classification system of the different collections

The classification of your collection is crucial in the way you can set up your digitisation workflow. Some information can only be found on the specimen itself but some information can be found on the folders/drawers/cupboards and is the same for all specimens in the same folder/ drawer/ cupboard. That information is very valuable as it can be directly linked to all specimens in it. This can save time and money.

For example vascular plant specimens in the herbarium BR of Meise Botanic Garden are stored in alphabetical order by family, genus and species. Therefore we could digitise the filing name from the folders (a QR code was added to a folder every time the filing name changed) instead of digitising it from the sheets. So all the folders with a QR code were imaged as well and based on these images, the filing name was transcribed. This QR code of the filing name was linked to all the specimens that followed. It saved us a lot of duplicate work as on average 1 folder was linked to 10 sheets.

1.4 Assessment of the condition of your (sub)collections

A crucial step in the pre-digitisation phase is to find out the curatorial status of your collection to know how much curatorial work is needed before your collection is ready for digitisation. As you will probably digitise your collections only once, you would want to do it properly.

The [SYNTHESES3 NA3 deliverable 3.3](#) report describes the outcome of a questionnaire that was sent around to 18 partners to review digitisation workflows and the used equipment. Two thirds of the institutions that completed the questionnaire performed at least some minimal curation or conservation steps prior to digitisation.

1.4.1 Curation status

- mounting needed
As mounting is much more time consuming than imaging (for herbarium specimens it is 50 to 5000 per day, Guiraud et al. 2020), it is best to decouple the mounting process from the digitisation workflow.
- desleeving/unfolding/opening needed
- cleaning needed
- restoration needed
- renewal of the folders/boxes/jars/drawers needed
- refiling needed
For example removing obviously incorrectly filed items
- no curation needed

While checking the curatorial state of your collection, you can already start making a list of necessary supplies (see section 3 for supply lists).

1.4.2 Specifying safeguards for handling specimens

- toxicity list & risk assessment
Many collections have been treated in the past with chemicals that are now considered as unsafe: herbarium collections could have been treated with mercury, nitrobenzene, naphthalene (in mothballs) or other chemicals; vertebrate skins with arsenic, geology collections store asbestos-containing material or can be radioactive. Liquid collections can contain formaldehyde or other harmful substances. Mould and biological hazards form other problems that can be found in collections.

List all the toxins that were used in your collection to make sure that you take the necessary precautions.

This is not only necessary for the people who work with the collections on a daily basis but it is also important when you plan to outsource the digitisation of your collection. It is recommended to inform the external company and to make a risk assessment.

For a case study see [Appendix 7](#) (decontamination of parts of the herbarium BR at Meise Botanic Garden).

Useful literature/links:

Hawks C & Makos K 2000. Inherent and Acquired Hazards in Museum Objects, implications for Care and use of Collections. Available at: <http://www.jorgealiaga.com.ar/documentos/gestion-SG2-Depositos/Inherent%20Hazards%20in%20Museum%20Collections.pdf>

Hawks C, McCann M, Makos K, Goldberg L, Hinkamp D, Ertel D & Silence P (Eds) 2010. [Health and safety for Museum professionals](#). - Society for the Preservation of Natural History Collections (SPNHC), 647 pp. Available at: <https://www.universityproducts.com/health-and-safety-for-museum-professionals.html>

Omstein L 2010 Poisonous heritage: pesticides in museum collections Theses. 253. <https://scholarship.shu.edu/theses/253>

Slocum N 2018. Toxins in the Collection: Museum Awareness and Protection Museum Studies Theses. 16. Available at: http://digitalcommons.buffalostate.edu/museumstudies_theses/16

Rae, A 2012. Hazards in museum collections: A Collections Care How to Guide. Available at: <https://collectionstrust.org.uk/wp-content/uploads/2016/11/SHARE-Museums-East-How-To-Guide-to-Hazards-in-Museum-Collections.pdf>

1.4.3 Labels readable

It is important to know is the readability of the specimen labels when digitised. Sometimes they are hidden underneath (part of) the specimen and will not be readable from the image. Will you reposition the label before digitisation? Will you transcribe the information from the specimen itself or from the image?

2. Prioritisation list of (sub)collections

This prioritisation of collections is extensively treated in WP1 of DiSSCo Prepare, and this section of the checklist will be updated once this deliverable is complete.

Useful links:

Bakker, H, Willems, L, van Egmond, E, Casino, A, Gödderz, K., Vermeersch, X. 2018. Inventory of criteria for prioritisation of digitisation of collections focussed on scientific and societal needs. Available at: <https://doi.org/10.5281/zenodo.2579156>

Van Egmond, E, Willems, L, Runnel V, Saarenmaa, H, Koivunen, A, Lahti K, Livermoe L 2019. Prioritising scientific and societal needs for data using small and private collections. Available at: <https://doi.org/10.5281/zenodo.2582995>

2.1. based on

- size
- origin
- readiness
- scientific importance
- historical importance
- which collections can be mass digitised
- funding situation (opportunities, obligations to funders)

3. Supplies

Enough supplies are essential when you start digitising your collection. Bear in mind that it is most of the time cheaper to buy material in bulk. This however will possibly mean that you have to tender to get the necessary materials. For a case study, see [Appendix 11](#) (ordering supplies for pinned insect digitization, NHM London).

3.1 Restoration supplies

- Microscope slides:
 - cardboard slide folders
 - Glue
 - Very thin forceps
 - Temporary slide tags

- Skins and vertebrate materials
 - Boxes
 - Filling material e.g. cotton wool
 - Drawers
 - Unit trays
 - Thread, scissors, archival pens, forceps
- Liquid preserved specimens
 - Preservatives
 - Stoppers
 - Labels
 - Petri dishes
 - Flagging material
- Pinned insects
 - Pins
 - Drawers
 - Unit trays
 - Forceps
- Herbarium sheets
 - folders
 - sheets
 - gummed paper
 - needle
 - sewing thread
 - envelopes
 - Bags to store loose material/small parts
 - glue
 - pencil
 - pen with archival ink
 - scissors
- Fossils
 - Stickers

3.2 cleaning

- cloths, alcohol

3.3 storage

- boxes, drawers, folders, cupboards, jars, vials, racks, stoppers

3.4 barcodes

When buying barcodes you have to decide on the following:

- format (2D, QR, data-matrix)
- info on/in the barcode (machine and human readable)
- size
- label material
- paper quality (conservation grade)
- Pre-cut or pre-punched
- One or two-sided
- Fixation (wire, glue, self-adhesive, pierced...)
- Quality of the adhesive

Make sure that the adhesives used are of high quality as you don't want them to come off and end up in a pile at the bottom of a cabinet/drawer. Also make sure that the adhesive used is harmless for the specimens where you add them to.

Kind of barcodes you may need to use:

- specimen barcodes
- drawer barcodes
- barcodes for cupboards

Useful links:

Diazgranados, M & Funk, V 2013. Utility of QR code in biological collections. Available at:
<https://dx.doi.org/10.3897%2Fphytokeys.25.5175>

iDigBio Specimen Barcode and Labeling Guide Wiki. Available at:
https://www.idigbio.org/wiki/index.php/Specimen_Barcode_and_Labeling_Guide

3.5 Track and trace

If you want to keep track of what happened with your specimens and where they are in the digitisation process you can set up a track and trace system.

For example when you conduct the digitisation in house, you can add a sheet with a to-do list (preparation, adding barcodes, restoration, imaging, databasing, freezing,...) to each drawer/pigeon hole/storage unit where you can tick what has already been done.

When you are outsourcing the digitisation, you can ask the external company to describe a track and trace system so that it allows them to remove the specimens from the original spot and put them back correctly after digitisation. It also will allow them to trace the specimens back to the correct storage location in case anything goes wrong.

Once you have set up your track and trace system, you can start listing the material you need for it. This can be:

- paper, stickers, magnets, barcodes, trolleys, boxes

3.6 IT infrastructure

A reliable IT infrastructure will make or break your digitisation project and is one of the first things you will have to invest in. Especially hardware for storing your images safely.

- hardware for long term storage/ cooperation with an external institution specialised in long-term archiving.
- hardware for temporary/local storage
- backup storage
- servers
- stations for QC images
- stations for transcription
- stations for QC transcriptions
- Software
 - For image acquisition
 - Image display
 - Post processing

- Monitoring
- Quality control
- archiving
- cloud storage
- portal
- CMS

This section will link to the MS3.7 ETL Procedures best practice recommendations once these are added to the Digitisation Guides website.

3.7 digitisation station(s)

- camera
- lens
- lighting
- table
- computer
- software
- IT storage (local, disk, server, cloud, image transfer)
- Background
- Equipment for specimen mounting
- tripod/stand
- Set of scales and color charts
- Connection cables, batteries and accessory
- Logo of your institute

3.8. Space for infrastructure of external company

If you decide to outsource (part of) the digitisation but plan the digitisation itself in your own institute, make sure you have a spare room where the external company can install their digitisation infrastructure. It should be easily accessible and close to the collections. The space should be suitable to locate additional workplaces for the external project management and break rooms. Clarify access to existing infrastructure (restrooms, meeting rooms) of your institution and agree on facility management tasks of the space (e.g. cleaning). Foresee enough access points for electricity and a link to the server room of your institute.

At Meise Botanic Garden, we added a floor plan to the tender with (if possible) a few options of where the installation could be set up. We also added our preference location but this way the external company had the choice to see what the most suitable location was. For the installation of a conveyor belt, as it was the case in Meise, a room of 65 m² was needed.

4. Pre-digitisation curation

4.1 Taxonomy

Useful links:

This will link to workflows on the Digitisation Guides website

4.2 Specimen quality

Useful links:

This will link to workflows on the Digitisation Guides website

4.3 Barcoding

Useful links:

This will link to workflows on the Digitisation Guides website

4.4 Storage renewal

Useful links:

This will link to workflows on the Digitisation Guides website

5. Staff availability

Staff availability is extensively treated in WP2 of DiSSCo Prepare, and will be updated once this deliverable is complete.

5.1 Preparation of the collection

5.2 Imaging

5.3 Transcription

5.4 Quality control of the images

5.5 Quality control of the data

5.6 Image archiving

5.7 Data publishing

5.8 IT developer

5.9 Project manager

For a case study see Appendix 8 (Staff list for the mass digitisation project DOE! at Meise Botanic Garden)

6. Cost books

The costbook is extensively treated in WP4 of DiSSCo Prepare and will be updated once this deliverable is complete.

6.1 Staff

6.2 hardware and software

6.3 outsourcing

6.4 available resources

6.5 resource application

Useful links:

Hardisty, A, Livermore, L, Walton, S, Woodburn, M, Hardy, H 2020. Costbook of the digitisation infrastructure of DiSSCo. Available at: <https://riojournal.com/article/58915/>

7. Decision to do in house or outsource

Based on the assessment of the collection, the available funding and the staff you have, you can make decisions on the following steps if you want to outsource them or do them yourself.

If the imaging/transcription is outsourced with quality control (QC) conducted by the outsourcing partner, we highly recommend you also should have an internal quality control procedure in place.

For a case study see [appendix 9](#) (Tender written by Meise Botanic Garden for the outsourcing of the African and Belgian herbarium collections)

7.1 in house

- Restoration
- Barcoding
 - add barcodes to the specimens prior to imaging
 - at the imaging station itself
- Transportation
 - Do you bring your material to the digitisation station yourself?
- Decontamination
 - Do you include this into the transportation or not?
- Imaging
- QC imaging
 - Automated QC
 - Visual QC
- Transcription
- QC transcription
- Image archiving
- Data publishing

7.2 outsource

Write down detailed procedures for every step you will outsource so that there is no room for discussion.

- Tendering
 - use examples
 - specify your needs
 - describe your collection in detail
- Restoration
- Barcoding
- Packaging

Depending on where you need to transport your specimens to: if your collection is housed in the same building you need less packing then when you need to transport them to another building.
- Transportation
 - Do you bring your material to the digitisation station yourself?
 - Transportation from and to the digitisation station in house.
 - Transportation to another building
- Decontamination
- Imaging
- QC imaging
- Transcription
 - external company
 - Crowdsourcing
 - Write a transcription protocol
- QC transcription
- Image archiving

Work together with a company specialised in long term preservation
- Data publishing

Useful links:

European Commission. [Funding & tender opportunities](#). Available at:

<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/tenders>

MuseumInsider: Public Tenders. Available at: <https://museuminsider.co.uk/category/public-tenders/>

NPS Museum Handbook 1999. Chapter 6: Handling, Packing and Shipping. Available at:

<https://www.nps.gov/museum/publications/mhi/chap6.pdf>

Canadian Conservation Institute: Controlling Insect Pests with Low Temperature. Available at:

<https://www.canada.ca/en/conservation-institute/services/conservation-preservation-publications/canadian-conservation-institute-notes/controlling-insects-low-temperature.html>

8. Necessary workflows

8.1 Pre-digitisation curation

- barcoding
- restoration
- taxonomy
- renewal of the storage unit
- marking already digitised specimens

8.2 Imaging

- digitisation standards
- format
- derivatives

8.3 databasing

- What fields do you want to have transcribed?
- Will you make skeletal records/add QR codes for minimal data entry?
- Data Standards

8.4 QC imaging

Useful links:

Nieva de la Hidalga A, Rosin PL, Sun X, Bogaerts A, De Meeter N, De Smedt S, Strack van Schijndel M, Van Wambeke P, Groom Q (2020) **Designing an Herbarium Digitisation Workflow with Built-In Image Quality Management**. Biodiversity Data Journal 8: e47051.

<https://doi.org/10.3897/BDJ.8.e47051>

8.5 QC databasing

For a case study see [Appendix 10](#) (Quality control procedure of Meise Botanic Garden for the mass digitisation project DOE!)

8.6 Decontamination

8.7 tracking system

For a case study see [Appendix 12](#) (Tracking system (collection move Naturalis))

8.8 data storage

8.9 data publishing

- own portal
- GBIF
- Digital curation: link to possible websites you can use for updating the quality of your data (GBIF, Geonames, gazetteers, worldfloraonline...)
-

8.10 data management plan

Useful links:

Ghent University DMPonline.be: How do I write a Data Management Plan? Available at:

<https://onderzoektips.ugent.be/en/tips/00001281/>

Pre-Digitisation: General Useful Links

iDigBio Pre-digitization Curation and Staging Wiki. Available at:

https://www.idigbio.org/wiki/index.php/Pre-digitization_Curation_and_Staging

Kalms, B 2012. Guidance. Digitisation: A strategic approach for natural history collections. Available at: <https://www.ala.org.au/wp-content/uploads/2011/10/Digitisation-guide-120326.pdf>

SPNHC Wiki: Digitization. Available at: <https://spnhc.biowikifarm.net/wiki/Digitization>

Popov, D, Roychoudhury, P, Hardy, H, Livermore, L, Norris, K. The Value of Digitising Natural History Collections. Available at: <https://riojournal.com/article/78844/>

Appendix 1: Digitisation Readiness Checklist for National Science Collections

When the checklist is added to the digitization guides website, it will include a pdf checklist which can be printed and used. This will only contain the checklist items, rather than the useful links and explanatory texted. The current version of this checklist can be found [here](#).

Appendix 2: Case study: detailed inventory of the collections for DiSSCo Flanders

Van Baelen, Ann, Poot, Nathalie, Beirinckx, Lise, Bogaerts, Ann, Bellefroid, Elke, Claerhout, Tim, De Smedt, Sofie, Dugardin, Chantal, Engledow, Henry, Leliaert, Frederik, Ossaer, Joke, Pereboom, Zjef, Semal, Patrick, Slos, Dieter, Smirnova, Larissa, Vandepitte, Leen, Veltjen, Emily, & Trekels, Maarten. (2022). DiSSCo-Flanders WP2 - task 2.1, Detailed inventory of the collections: Report (Version 1, May 2022). Zenodo. Available at: <https://doi.org/10.5281/zenodo.6511351>

Appendix 3: Setting Natural Science Data Free: Scoping UK Collections

Summary of work

As a first step towards improving natural science digitisation, we sought to gain insight into the breadth and depth of UK natural science collections, and the extent to which these collections have been digitised. The initial challenge of this scoping exercise was identifying all natural science collections in the UK. Using regional museum development groups, existing contact lists, and online searches, we collated a list of over 150 institutions with public natural science collections, consisting of museums, herbaria, university collections, and research societies. While not every natural science collection was accounted for, and some did respond to our request for collections data, we received survey responses from 84 institutions.

Inventory of collections

The scoping exercise was based on the SYNTHESYS+ survey to maintain standardisation across the similar DiSSCo led projects. The key difference with our national survey was the range of participating organisations, all with differing capacities for completing the survey. Large institutions with dedicated natural science curators and a digitisation team will have greater capacity to provide a detailed summary of their collections when compared to small institutions with no dedicated digitisation team or scientific expertise. To obtain as much detail as possible while not deterring participation from smaller collections, we made the survey graded, allowing different levels of granularity. All institutions were required to complete the collection overview which asked for specimen count and digitisation level estimates for 9 key natural science disciplines (Anthropology, Botany, Extraterrestrial Objects, Geology, Microorganisms, Palaeontology, Zoology Invertebrates, Zoology Vertebrates, and Other Geo/Biodiversity). Where possible, we also asked institutions to provide a finer level of detail for their collections. There were options to provide specimen quantity and digitisation level estimates broken down by taxonomic group (45 taxonomic groups listed), preservation type (57 preservation types listed), and stratigraphy. This is particularly useful in identifying the areas to focus resources and create training materials for.

Estimation of your collections

When asking for estimates of the number of specimens within a collection, the accuracy of this estimate will vary considerably across institutions, depending on the size of the collection, staff experience and expertise, and the digital infrastructure available to the institution. For instance, some institutions surveyed had no natural history curator, no online database, and described their collection estimates as 'best guesses'. To account for estimate uncertainty, all institutions were asked to provide a confidence interval for every estimate. This was recorded as a percentage to reflect the true number of specimens within the collection. For example, a 10% confidence interval for a 1000 specimen estimate indicated that the true number of specimens lies between 900 and 1100.

Conclusion

The scoping survey produced our most up-to-date and accurate understanding of what UK collections hold. It revealed that most organisations lack support in digitising their natural science collections and are unable to mobilise their data to be utilised by the scientific community. The results of the survey have been used to create a blueprint for a national digitisation programme, to improve national digitisation and unlock the full scientific potential of UK natural science collections.

Appendix 4: Assessment of Microslide Collections at the Natural History Museum, Berlin

This case study will be added when the checklist is added to the DiSSCo Digitisation Guides website.

Appendix 5: Estimation of the numbers of the African and Belgian Herbarium Collection at Meise Botanic Garden

In 2015, Meise Botanic Garden received a grant from the Flemish Government to digitise all the central African (Congo DR, Rwanda and Burundi) and Belgian herbarium specimens within 3 years. The first step in this mass digitisation project called DOE! (Digitally unlocking the heritage collection) was a 10% count of the whole African vascular plant collection, which is kept as a separate subcollection. In the African herbarium of BR, brown folders are used to mark the specimens collected in Congo DR, Rwanda and Burundi. All specimens collected in other African countries are stored in green folders. We wanted to know the percentage of central African collections and see if it was worthwhile to only digitise these specimens or digitise the whole African herbarium at once. For every row of cupboards in the herbarium, the first cabinet was completely counted. A division was made between specimens kept in green or in brown folders and all the type specimens for each colour of folders was noted as well. This was necessary because all the types were digitised in previous projects and it was decided not to digitise them again.

When we extrapolated the numbers for the whole African collection, we arrived at a number of around 900 000 sheets, 100 000 less than presumed before the count. We have also found that 57% of all the specimens were collected in central Africa, 43% was non central African material - Note that in BR the African herbarium only holds specimens collected South of the Sahara- . Based on these results, we decided to go for the digitisation of the whole African collection because it would cost us too much time to only extract the central African specimens instead of digitising them all. The number of digitised specimens at the end of the project was very similar to the 900 000 of the estimate.

As the Belgian herbarium is kept separately and was almost completely barcoded with a numbering system that allowed us to know how many holdings we have, we didn't have to conduct a 10% count.

Appendix 6: Estimation of collection size at the NHM London

It is quite hard to accurately estimate the size of a collection. The success of estimation depends on many factors, among them:

- Is there previous experience digitising part of that collection or a collection that is similar?
- Age (of the collection)
- Origin (of the collection)
- Type (of the collection)
-

For example, in the NHM, we have had several projects digitising entomological slide collections, therefore we have good estimates on how many slides can fit into a full drawer. Knowing this, before

each new slide digitisation project, we audit the collections - meaning we go in, look at the drawers and using an eyeball estimate, we estimate the fullness of the drawers. Using that information and the existing data we have we can then make quite accurate estimations on the size of the collection.

It is more difficult if we have no pre-existing experience working with the collection and using a similar collection does not always work.

Estimating the size of the herbarium for digitisation purposes is a more complex task. We can start building the estimate on how many cabinets we have, how many sheets can fit into a cabinet and how full the cabinets are. But we are disregarding factors that are affecting the estimation, such as a) bulkiness of the specimens, b) multi-specimen sheets.

Multi-specimen sheets are quite tricky as without looking at the actual specimens, we can't determine how many specimens are on one single sheet, it can be one, two, three or twenty even. If we have a lot of these sheets in the collection. If we estimate the numbers based on the number of sheets, we will underestimate the size of the collection (and therefore the time taken to digitise).

The first mass digitisation project in the NHM herbarium was digitising the Brassicales order. The actual size of the collection was twice that of the original estimation. An accurate estimate requires a good knowledge of the collection. There are certain factors that can help us in the estimation process that comes from understanding the history of the collection. It is useful to have knowledge of when the collection was acquired, where it was collected and in what era. If we have a huge collection from relatively recent times, e.g. 1980s, we can safely estimate the number of sheets, as the multi-specimen sheet practice was not in use at this time. Information that can help includes the collector (e.g. are their collections often mounted together with someone else's specimens?), the region collected, and whether the paper was in short supply or expensive.

Knowing the collection, its history and origins can help us estimate the size better. But it is also a good practice to leave around a 20% variation if we are talking of a project larger than 40-50,000 estimated records.

Appendix 7: Decontamination of parts of the Herbarium BR at Meise Botanic Garden

The vascular plant collection of the herbarium BR at Meise Botanic Garden was treated in the past with mercury. The AWH collection, incorporated in the BR collection in 2006, is poisoned with nitrobenzene.

For the second mass digitisation project DOE12 when these collections were going to be digitised, a risk analysis on the use of these chemicals was added to the tender to make sure that the external company was aware of the risks so they could take the necessary precautions.

Before we outsourced the digitisation of this AWH collection, we removed the jars with nitrobenzene out of the metal boxes which contained the specimens. A protocol was written for this as well:

Removing jars of nitrobenzene and airing the Van Heurck collection (AWH)

1. Install ventilation and make it operational
2. Safety Precautions
 - Full face mask with filter A2B2P3
 - Yellow disposable pack (pesticide)

Disposable gloves (polyvinyl alcohol)

3. Supplies

Safety clothing (see point two)
 Jar with double closure
 Container for chemical waste
 Mobile scaffold
 Free workbench

4. Working method

- Requires a minimum of 2 persons who can pass along boxes that are at a higher height.
- Take box by box off the rack. Use mobile scaffolding for boxes at a higher height.
- Open boxes and place the lid under the box (for faster ventilation).
- Remove the jar of nitrobenzene from the box and place it in the double-closing jar.
- Replace boxes in the same order. Slide the bottom box all the way to the back, the box that comes on top is slightly slanted and stepped in the other places in such a way that both boxes can air sufficiently.
- Dispose the closed jar containing jars of nitrobenzene in the chemical waste container.
- The safety officer will take care of the disposal of this container of chemical waste.



Figure: Jar containing nitrobenzene

Here you can find what was added in the tender:

The following measures should be applied when working with herbarium specimens inside and outside the collection areas:

- Wear a lab coat and gloves (polyvinyl alcohol);
- Wash your hands after working with herbarium specimens;
- Do not eat or drink in the collection;
- Keep the doors of the collection areas closed;
- Walk away from the cabinet doors after opening and wait at least 1 minute before starting to work in the cabinet.

Pregnant women and women who wish to become pregnant are advised NOT to enter the collection areas and to avoid contact with herbarium specimens. Breastfeeding women are also NOT allowed to enter the collection areas and must avoid contact with the herbarium specimens.

Measures to work in the collection:

Herbarium material is susceptible to an attack by pests, especially various species of beetles and silverfish. Today, pest damage is prevented by regular freezing.

In order to keep the risk of 'contamination' (= damage by pests) as small as possible, a number of measures should be taken with regard to the collection areas

In the collection areas (storage and working spaces) it is prohibited to

- eat or drink (only a bottle of water with a 'drinking cap' is allowed)
- bring food
- bring objects or persons without the prior consent of the collection manager or his replacement;
- open windows without consultation of the collection manager or his replacement
- leave herbarium specimens unprotected, put them back in the herbarium cabinets or in closed boxes as soon as possible;
- leave room doors, cupboard doors and boxes open unnecessarily;
- move herbarium material between the different collection rooms.

Wear a lab coat and gloves (polyvinyl alcohol) when working with herbarium material				
		exposure risk - operation		
		low (non contaminated material)	high (material contaminated or has been in possible contact with contaminated material)	very high (heavily contaminated or suspect of heavily contaminated) (1)
exposure risk - room	low (no use of chemicals in the room)	mounting new incoming material	(re)mounting old material, imaging intercalation, collection consultation, transcription	all actions
	offices			
	high (use of chemicals in the room)	mounting new incoming material	(re)mounting old material, imaging intercalation, collection consultation, transcription	all actions
	herbarium rooms			
				+ use fume cupboard
Wear a lab coat and gloves				
not necessary				
recommended (2)				
Obliged				

Table: Safety equipment required in the AWH collection

(1) Crépin roses collection (mercury), Van Heurck collection (nitrobenzene), material with 'poisoned', 'vergif'

(2) as long as no occupational exposure standards or limits are exceeded (for mercury: in urine: 20 µg/g creatinine; in blood 15µg/l; exposure limit : 20µg/m³)

Appendix 8: Staff List for the Mass Digitisation Project DOE! At Meise Botanic Garden

- Project manager (0,8 FTE)
- IT specialists (hardware, software, storage) (1,5 FTE)
- Collection manager (daily management and follow up of the restoration/preparation) (0,5 FTE)
- Collection technicians (restoration, preparation, imaging in house, transcription in house, QC external transcribed label data, pest management) (8FTE)
- Database manager (for daily management and QC) (0,5 FTE)
- QC manager images (for automated and visual checks) (0,6 FTE)
- Data publisher (publishing images and data to different portals, maintenance) (0,5 FTE)
- IT Developer (external)

Appendix 9: Tender written by Meise Botanic Garden for the outsourcing of the African and Belgian Herbarium Collections

This case study will be added when the checklist is added to the DiSSCo Digitisation Guides website.

Appendix 10: Quality control procedure of Meise Botanic Garden for the mass digitisation project DOE!

To determine the extent to which label transcription meets quality requirements the following will be examined:

1. the method that will be used for quality control;
2. the common mistakes, to which an error weight is assigned, ranging from 0,1 to 0,5 penalty points (error calculation);
3. the measuring standards that reflect the acceptance levels.

1. Method

The quality will be measured using a sub-sample of the data file. The sub-sample size depends on the size of the data file. The sub-sample size is determined using the table under point 3.

2. Types of errors

Two types of errors are distinguished:

Identification and Transcription errors

- 1) Identification errors occur when:
 - Data is entered into the wrong field or incorrect data is entered in a field;
 - Data has not been entered despite it being present on the label.
- 2) Transcription errors: when data have not been correctly transcribed from the label (typos).

Further detail on the error calculation method, including a table, will be included in the Digitisation Guides Website.

3. Measuring standards that reflect the acceptance levels (ISO 2859)

The acceptance or rejection of a file is determined with reference to the table below. Acceptance table: when a file has a batch size of 450 records, for example, the sub-sample batch size of 500 records will be used. We will use the test level II-Normal, which has the identification letter H. For this code letter a sample size of 50, where penalty points <2 are approved (G1) and ≥ 2 are rejected (A1). In other words, 1,9 penalty points are approved and 2 penalty points are rejected.

Suppose a batch comprises 500 records, thus a sub-sample of 50 records. A single record may not have a penalty point greater than 1. For example, if a single record has 10 mistakes giving it a total of 3,8 penalty points this still counts as 1 penalty point. If only one error was counted at 0,5 penalty points then this counts 0,5. The sum of all penalty points determines whether a batch is accepted or rejected.

Further detail measuring standards, including a table, will be included in the Digitisation Guides Website.

Appendix 11: ordering supplies for pinned insect digitisation: Natural History Museum, London

There are a wide variety of supplies required to ensure the smooth running of any pinned insect digitisation project. This can range from more substantial items - cabinetry, drawers, cameras etc. to consumables - pins, UID barcodes, EVA foam etc. and the ability to suitably plan to have these available for any project is contingent on several factors:

- What is currently already available to be utilised?
- What is the accuracy of the estimate of specimen numbers for a specific project?
 - It is useful to build in contingency to any order of regularly used materials but potential future issues and delays can be alleviated the more confident you are in any estimate
- Is there a budget available for required items? Is this ring fenced for the project or more general?
- Are certain items known to have long lead times?
- Are there any items that are difficult to source/no longer available and will suitable substitutes need to be found?

A recent pinned insect digitisation projects at the NHMUK shows a variety of issues that may be encountered when ordering supplies.

One large digitisation project involved rehousing the collection from old, cork-lined drawers to unit trays in new drawers prior to imaging. At the beginning of the project, there was a supply of both unit trays and new drawers to be used and it was known that these would likely need to be reordered before the culmination of the work.

This project was externally funded but the terms of the funding did not extend to consumables so provision for these became the responsibility of the collections budget. New drawers and unit trays are regularly ordered for the entomology collections to allow for rehousing/recreation and expansion and is normally done in bulk to make benefit of related savings to orders at scale.

Unfortunately, this bulk ordering meant that there was a period when suitable, new collection drawers ran out as the latest outstanding order was yet to be fulfilled (it appears that the drawer manufacturer had scaled back their workforce due to a downturn in business during the pandemic causing increased lead times).

In order to be able to continue with the project, it was necessary to source a temporary storage alternative until the arrival of the new drawers. Fortunately, there were drawers of a different size available that could be used as a stop gap to store the newly rehoused specimens, in unit trays, in the collections.

Appendix 12: Tracking System (Collection Move: Naturalis)

Labelling containers with future storage location

- Efficient tracking system for objects and containers of objects (location, condition, quantity)
- Use of barcodes or RFID tags. Barcodes don't need to be physically attached to the objects themselves but placed in move trays and supports. rolls of double barcodes were produced - one to put on the worksheet and another to place on the container.
- items/crates can be scanned at a number of points e.g. when an item is taken off a shelf, when it is packed, when it is placed in a crate, when the crate is put into/and taken out of a lorry, and when the item is placed in store or at its final destination. Barcodes may be stuck directly onto boxes or packing materials, or onto slips of paper which can be inserted into collection items
- Knowledge of drawer contents
- Current image