D1.3 Digitisation of natural history collections – criteria for prioritisation

Louise Isager Ahl (0000-0002-4727-1882)[1]
Luca Bellucci (0000-0002-8464-6153)[2]
Pip Brewer (0000-0001-7728-1021)[1]
Pierre-Yves Gagnier (0000-0002-7104-898X)[3]
Elspeth Haston (0000-0001-9144-2848)[4]
Sofie De Smedt (0000-0001-7690-0468)[5]
Henrik Enghoff (0000-0002-2764-8750)[1]

Affiliations
[1]Statens Naturhistoriske Museum/Natural History Museum of Denmark, Gothersgade 130, 1123 København K, Denmark
[2]Museo di Geologia e Paleonotologia, Sistema Museale di Ateneo, Università di Firence, via Giorgio LaPira 44, 50121 Firenze; Istituto Italiano di Paleontologia Umana, Università degli Studi di Firenze, Università degli Studi di Roma La Sapienza, Italy
[3]Muséum National d'Histoire Naturelle, Université Paris Diderot, France
[4]Royal Botanic Garden Edinburgh, Scotland, United Kingdom
[5]Agentschap Plantentuin Meise, Nieuwelaan 38, 1860 Meise, Belgium

Contributors
Lorenzo Cecchi (0000-0001-9706-0947)
Lisa French (0000-0001-7279-8582)
Helen Hardy (0000-0002-9206-8357)
Gianna Innocenti (0000-0002-4504-0765)
Anne Koivunen (0000-0002-3475-7071)
Laurence Livermore (0000-0002-7341-1842)

Table of Contents

EXECUTIVE SUMMARY ***3***

BACKGROUND ***5***

METHODOLOGY ***7***

Search for additional studies on digitisation criteria **7**

Surveys **8**

Survey 1 – Essay based questionnaire 8

Survey 2 – Multiple-choice questionnaire 8

Case studies on cost **9**

RESULTS ***10***

Literature review **10**

GBIF report 10

ICEDIG report 10

Surveys **11**

Survey 1 – Essay based questionnaire 12

Survey 2 – Multiple-choice questionnaire 12

Case studies on cost **14**

DISCUSSION ***16***

Introduction **16**

Relevance **17**

Data quality **18**

Information level 18

Reliability 19

Assessing and improving data quality 20

Cost **22**

Feasibility **23**

Implementing the criteria **24**

CONCLUSION AND RECOMMENDATIONS ***25***

Questions to be asked 26

A final word 26

Acknowledgements ***27***

References ***28***

EXECUTIVE SUMMARY

A core mission of the Distributed System of Scientific Collections (dissco.eu) is to unite the ~1.5 billion specimens kept in European Natural History Collections into a one-stop e-science infrastructure containing as many of these specimens as possible in the form of **digital specimens.** To achieve this, a massive digitisation effort is required, and to guide this effort, criteria for how to prioritise are needed.

This issue has been addressed in several previous publications, notably in a report from a GBIF taskforce (Krishtalka et al. 2016) and in a very comprehensive treatment resulting from the DiSSCo-related project ICEDIG (Bakker et al. 2016). We have reviewed these reports and conducted additional literature surveys in order to catch any relevant publications post-dating Bakker et al. (2018) and Krishtalka et al (2016). To address the issue in a different way, we have made surveys among DiSSCo partners asking for their digitisation plans and for the criteria they have been using for digitisation of their own collections. Furthermore, we have obtained information on the actual cost of digitisation projects, striving to include **all** costs associated with such projects, something that is lacking from available publications on this subject.

The general picture emerging from previous studies (Krishtalka et al. 2016, Bakker et al. 2018) is that scientific/research relevance is rated as the most important criterion, but apart from that, the signal is unclear. Relevance in relation to the collections themselves as well as funding opportunities are acknowledged as important criteria, whereas societal relevance is regarded as a less important criterion. As an attempt to provide some guidance through the complex landscape of criteria we suggest that an organization (e.g. DiSSCo, or an individual institution) planning to digitise natural history collections, considers four categories of criteria, viz.:

- Relevance
- Data quality
- Cost
- Feasibility

The four groups embrace all criteria, which have been brought forward previously, and we discuss each group more or less extensively.

Data quality is given particular attention since this aspect of digitization has been somewhat neglected in previous works. Data quality has two main components:

- How much information is there in each digital specimen? (Information level). This component has been addressed through the development of the MIDS concept (Minimum Information about a Digital Specimen, Hardisty et al. 2021)
- How reliable is that information? Reliability includes accuracy (the closeness of measured values, observations or estimates to the true value) and precision (e.g. of geographical information: latitude/longitude in degrees only, in degrees plus minutes, or in degrees plus minutes plus seconds, or of taxonomic information: identification to genus, species or subspecies level)

The quality of a data set also includes its potential for quality assessment and improvement, as well as its completeness in terms of taxonomic, geographical or collection coverage.

Cost is obviously a major consideration in any digitisation project. We emphasize that cost estimates should include **all** costs associated with the project, including pre-digitisation, digitisation *sensu strictu* and post-digitisation) as highlighted in two case-studies in which we have analysed all costs associated with the digitisation of a herbarium and a collection of fossils.

It has become obvious that there is no easy way to implement the multitude of criteria. The idea of an algorithm such as a "decision tree" seems unviable, and we suggest that projects be evaluated/prioritized by a combination of a scoring method and a panel discussion, similar to what has been done in the series of SYNTHESYS projects.

We strongly recommend collaboration, e.g., at DiSSCo level, in order to optimize resources and we want to underline the importance of bearing in mind that irrespective of which criteria are considered, there is no fit-all solution. Flexibility is essential, depending on the intended use of the digital specimens to be generated.

We provide a list of questions to be considered in connection with the drafting or evaluation of digitization projects.

Finally, we stress that digital specimens can never replace the physical specimens that exist in collections, and that ensuring the long-term preservation of the collections remains a top priority.

This report is a deliverable of the DiSSCo Prepare Project (dissco.eu/dissco-prepare)

BACKGROUND

Natural history collections are treasure troves for scientists, and in order to safeguard and expand the use of these collections for the future, digitisation is pivotal. Attempts to digitise natural history collections throughout the world have already started. Distributed System of Scientific Collections (DiSSCo) is a pan-European Research Infrastructure (RI) for natural science collections. The aim of the infrastructure initiative is to unify all European natural science assets under common access, curation, policies, and practices. This approach and set-up will ensure that all the data is easily Findable, Accessible, Interoperable and Reusable (FAIR principles).

Digitisation in this context spans the spectrum from making basic information on a specimen (name, collecting locality etc.) digitally available, to including (or linking to) digital images (photographs, X-rays, scanning electron micrographs etc.), DNA sequences, chemical information etc. in the digitised information.

Digitisation can be approached in different ways:

- Mass digitisation – large digitisation projects like the digitisation of an entire collection (usually of thousands up to hundreds of thousands of items).
    - Especially, but not exclusively for mass digitisation, a pilot phase testing new digitisation workflow and/or technology, is recommendable.
- Project-driven digitisation – smaller defined projects focusing on particular specimens like those collected on a specific expedition or for a particular purpose.
- Digitisation on demand – digitisation of a limited number of specimens for a particular scientific study or project by external researchers who approach the collection-holding institution.
- Business-As-Usual (BAU) digitisation – digitisation made in connection with everyday curation, e.g., digitisation of specimens going out on loan, coming back from a loan or selected for an exhibition

In Europe alone there are an estimated 1.5 billion specimens stored in collections, representing nearly 80% of described species worldwide (Bakker et al., 2018). Today, more than 41 million specimen related records have been uploaded by the DiSSCo network to GBIF (database accessed 4/11-2022). These specimens have become **digital specimens**, which means they are closer to the FAIR guiding principles. The DiSSCo RI[1], which is currently completing the preparatory phase and entering a transitory phase, aims to produce digitised specimens in a FAIR framework on a large scale.

Within institutions, prioritisation may need to take into account all of the four categories above, in a 'balanced portfolio' approach that for instance ensures mass digitisation projects are balanced against user-led services and the need for innovation or more bespoke pilots, or the need to make equipment available for business as usual. For DiSSCo, prioritisation of what to digitise is perhaps most critical in relation to the coordination of mass digitisation programs and/or larger project-based digitisation, as these will primarily drive critical mass of content creation through the DiSSCo infrastructure. It is also likely that central coordination of on demand approaches may be required,

[1] https://www.dissco.eu

however this is less a question of prioritisation - which by definition is user led in these services - and more one of service design, funding etc. Mass or larger project digitisation are therefore the main (but not only) focus of this report. Prioritisation of what to digitise is mostly a concern of mass digitisation programs and/or larger project-based digitisations. These are therefore the main (but not only) focus of this report. The technical approach to digitisation is a related and overlapping subject, but this will not explicitly be dealt with here unless it is of direct relevance to the discussion.

The crucial question can briefly be framed as Where to start? Another crucial consideration is, to what extent decisions should be made at a European or global level, rather than in individual collection-holding institutions? A coordinated approach would allow us to focus more efficiently on solving specific problems that have a wide and significant impact on all of us, for example by assembling critical mass of relevant data to address key societal challenges; or by enabling the most efficient and effective workflows to be deployed widely with maximum impact. Here, DiSSCo offers a unique opportunity for coordinating the prioritisation, though it should also be recognised that each institution will have their own drivers and stakeholder requirements that will impact the prioritisation process (not least in that different institutions hold different types of collections and objects, which they will naturally see as their priorities).

METHODOLOGY

There are few descriptions and models available for prioritisation of digitisation targeting natural history collections. Many potential factors may influence the decision-making process regarding prioritisation, and the present paper is to be seen as a help to "*establish relevant criteria to identify a prioritisation model for digitisation"* (DPP Description of Work). To get a better understanding of what has been done in past and what is included in current digitisation programs we did the following:

- Performed a comprehensive review of the literature
- Designed and conducted surveys of digitisation plans and criteria employed among all DiSSCo partners

Additionally, we obtained detailed information of all costs associated with two digitisation projects that have been carried out in recent years.

An explanation of the acronyms used in the present report is available in Appendix 1.

Search for additional studies on digitisation criteria

At the onset of this project two core studies were available on the topic of digitisation. The most recent work was carried out in the ICEDIG project and reported in the final deliverable "Inventory of criteria for prioritisation of digitisation of natural history collections" (Bakker et al., 2018). This work nicely complemented the study by Krishtalka et al. (2016) on how to accelerate the discovery of biocollections data. The most important points made in these studies have been summarised in Appendix 2, and they were the inspiration for our literature investigations. Two literature investigations were carried out in 2021 and 2022. Based on the results a corpus of previous studies on prioritisation of digitisation was compiled, covering the period from 2018 until June 2022. The list of relevant references found during the 2021 survey was included in milestone report M1.3 (Ahl & Enghoff 2021, see Appendix 4), and those found during the 2022 survey are listed in Appendix 3.

For the 2021 survey, works deemed to be relevant were scored (1-3) based on relevance for the investigation with 1 being most relevant. The searches were carried out in Google Scholar with the following search parameters:

1. Search: ”natural history collections” ”prioritisation” since 2017
2. Search: ”natural history collections” ”digitisation” since 2017
3. Search: ”digitisation” ”prioritisation” since 2017
4. Search: ”natural history collections” ”digitisation” ”prioritisation” since 2017

In comparison to the results presented by Bakker et al. (2018), a total of 12 new publications deemed to be relevant were identified from the 4 searches (April, 2021). In the additional analysis carried out in June 2022 a total of 14 new publications deemed to be relevant were identified from the 4 searches. See Table 1.

	April 2021		June 2022	
Search no.	**No. results**	**Relevant**	**No. results**	**Relevant**
1	143	4	223	6
2	775	4	1170	4
3	4640	2	4640	2
4	46	2	46	2

Table 1. Results of the 4 search compilations done April 2021 and June 2022.

The 2022 survey was carried out under much broader criteria and resulted in a large number of publications, see Appendix 3.

Surveys

As a necessary complement to the literature study, two surveys were carried among DiSSCo partners on their digitisation strategy (if existing), as well as on which prioritisation criteria they employed for digitisation which had already been done or was in progress.

Survey 1 – Essay based questionnaire

DiSSCo partners were asked to provide information, in free text and preferably no more than 2 A4 pages, on

1. Their digitisation strategy (if available, they were asked to provide a copy or link).
2. The prioritisation criteria employed for digitisation which has already been done or is in progress in the institution.

The following guiding questions were supplied to highlight 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 digitisation 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 was suggested that in their answers it could be useful to distinguish between

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

This study was carried out in the autumn and early winter of 2021.

In Appendix 4 a list of all countries in DiSSCo and the institutions from each country that have replied to our questionnaire has been compiled. Institutions marked with * are partners in DPP task 1.3. A complete compilation of replies was submitted in the DPP milestone report "Corpus of previous studies on prioritisation of digitisation compiled" (Ahl & Enghoff 2022), and this has been included in Appendix 4.

Survey 2 – Multiple-choice questionnaire

The multitude of thoughts, approaches and results described by respondents to the essay-based questionnaire provides answers for interesting reading although, as expected, the format makes it difficult to quantify or even to describe the results in a few paragraphs or diagrams. Therefore, we subsequently developed a short multiple-choice questionnaire focused on the digitisation activity, using a Google Form. The short questionnaire, after being reviewed by the task partners, was sent to all DiSSCo National Nodes who shared it with their own institutions in order to collect information from as many institutions as possible involved in DiSSCo. To facilitate the dissemination, the questionnaire was translated into different languages (English, Danish, French, Italian and Dutch). An overview of the questions and answers can be found in Appendix 4.

The structure of the questionnaire was as follows:

- Q1 – Q3 compiler's information (personal details, e-mail, role, country, institution)
- Q4 – Q5 information about collections (size and staff employed)
- Q6 – Q9 information about digitisation strategy (digitisation initiative, digitisation priorities classified in 5 main categories, Scientific Relevance, Institutional Relevance, Economic Relevance, Educational Relevance, Technical feasibility and subcategories for each one of them)
- Q10 – Q12 information about the management of collections (overview and monitoring of the digitisation status, use of CMS-Collection management system)
- Q13 – Q16 information about digitised items (procedure for data validation, standard used for databasing, digitisation levels for databased items, images and 3D models)
- Q17: further remarks about digitisation strategy

Of the 23 national nodes, only 10 answered, with a total of 79 answers. Most of the answers came from NH Museums or University Museums and Research Institutions. Thus, most respondents are curators, several are researchers or directors of the collections, and a few are digital collection managers or similar (Appendix 4, Q1-3).

More than 50% of the responding collections are very small to small, 23% are medium-size, and only 14% are large to very large. In general, a small number of persons is employed at very small to medium-size collections; with some exception: 5 large or very large collections have a small staff, 6 small collections have medium size staff, and one very small collection has a large staff (Appendix 4, Q4-5).

Case studies on cost

We asked all partners in the task for detailed and complete information on digitisation costs. Such information is not readily available for most projects, but we can present two detailed case studies obtained from NHMD and UniFi.

RESULTS

Literature review

The most significant results obtained through the literature review were reports carried out by GBIF (2016) and within the DiSSCo-related Project ICEDIG (2018). These two publications will therefore be summarised here (extended summary in Appendix 2); the additional relevant publications are listed in Appendix 3 and 4.

GBIF report

A task force was convened by GBIF "to help accelerate the discovery, digitisation and access to biocollections data". One of the task force's main objectives was to provide guidance on establishing priorities for digitising biocollections to serve institutional, national, and global needs and achieve the greatest economies of scale (Krishtalka et al., 2016). The GBIF task force undertook a large-scale, global survey among collection-holding institutions on the state and prioritisation of digitisation. A total of 519 respondents gave information on their priorities, and these are presented in Figure 1.

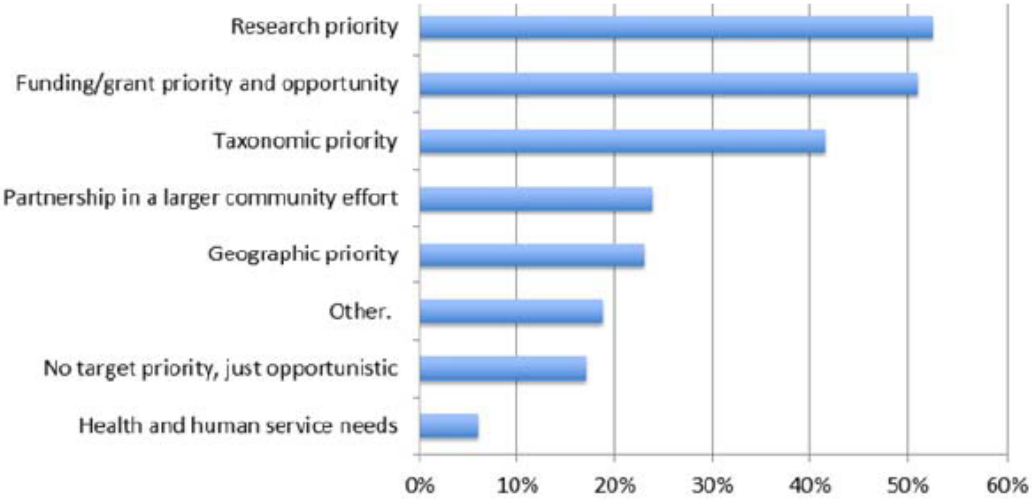


Figure 1. Percentages of collections surveyed by GBIF applying various criteria for prioritization of collections, from Krishtalka et al (2016).

The most important priorities identified by the GBIF task force were reported to be:
1) research
2) funding/grant opportunities
3) taxonomic priorities.

However, these findings are only in part compatible with the most important criteria found by ICEDIG, see below.

ICEDIG report

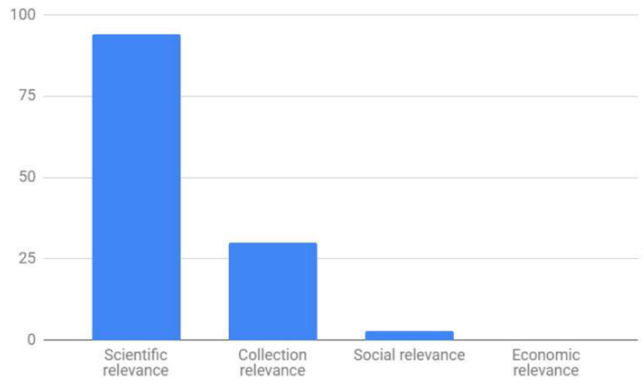
ICEDIG was an EU funded project under the Horizon 2020 Framework. In the Deliverable D2.1 "Inventory of criteria for prioritization of digitization of Natural History Collections" (Bakker et al. 2018) a corpus and analysis of digitisation criteria was presented. It forms a very substantial part of

the basis for the present deliverable of DPP Task 1.3. The aim of the ICEDIG deliverable was to contribute to an "easy and well-informed decision-making process in relation to prioritisation of digitisation of natural history collections". In ICEDIG it was decided to follow a multi-stage process to ensure that the solutions put forward were solid regarding the prioritisation of digitisation of natural history collections. Stages identified:

1. **A literature and reports inventory** was carried out to create an overview of the criteria of prioritisation of digitisation.
2. **Targeted survey.**

For the questions regarding prioritisation Bakker et al. (2018) obtained 68 completed responses that were included in the depictions of the data shown in figure 1 and included in Appendix 2. Figure 2 gives the overview of the ranking of the four areas of relevance identified: scientific, collection, social, and economic. Included in Appendix 2 are figures S1-S4 that show the ranking of the criteria used in the questionnaire identified for each of the four areas.

Figure 2. Overview of the relative importance of the relevance areas identified regarding digitisation.



Based on the additional information added in free text, an extensive and revised list of criteria was assembled on six overarching topics:

1. collection relevance
2. economic relevance
3. funding
4. practical criteria
5. scientific relevance
6. social relevance.

Due to the broad range of criteria that were identified to be of importance in the process of prioritising digitisation efforts, three possible methods to determine the strategy for a digitisation project were proposed: 1) Decision tree; 2) Scoring method, and 3) Panel review.
Although relevant publications were identified through the additional literature survey, they did not add anything substantial that had not already been covered by Krishtalka et al. (2016) and Bakker et al. (2018).

Surveys

Two surveys were carried among DiSSCo partners on their digitisation strategy (if existing), as well as on which prioritisation criteria they employed for digitisation which had already been done or was in progress. The main findings have been summarised here, and the complete responses can be found in Appendix 4 and 5.

Survey 1 – Essay based questionnaire

The natural history collections that replied to our questions are at different levels in their digitisation efforts. This means that the answers reflect whatever level they are at and are therefore hard to sum up in a coherent way as they varied from "all our collections have been digitised" to "we have no official document outlining our digitisation priorities". However, most seem to adhere to the criteria put forward by Bakker et al. (2018) by starting their digitisation process by capturing the data of their most important specimens (types, historic, fragile, cultural). Another strong driver of the collective digitisation efforts by DiSSCo members has been the opportunistic approach, i.e., a broad span of research and funding opportunities has determined the priorities. Finally, a lot of members are actively trying to digitise all new incoming specimens to some degree. Survey 1 was summarised and presented in a milestone report included here as Appendix 4.

In terms of prioritisation criteria employed for digitisation efforts many respondents had left this blank or indicated that internal work was in progress to define their approach. It is therefore not possible to extract general tendencies. Instead, we present, as a concrete example, the key criteria for digitisation efforts employed by NHMD.

- National collection strength
- Research and public relevance
- Digitisation cost and volume
- Established international policies and archival formats

Survey 2 – Multiple-choice questionnaire

Of the 79 institutions that replied to the questionnaire, 28 have a well-defined digitisation strategy (20 with small collections, 4 medium-size, 2 large and 2 very large collections), 13 were uncertain about this, but most (37) do not have any digitisation strategy. In general, the digitisation seems to be primarily driven by "Projects (e.g., E-Recolnat, national lists of flora or fauna etc.)" and "Opportunistic digitisation (e.g., moving the collection into a new site, out-going loans, new specimens entering the collection, exhibition and other contingent events)". The "Digitisation on demand (i.e., *ad hoc* digitisation for specific research, as requested by external researchers, e.g., through VA Synthesys+)" is the third choice. In any case, mass digitisation still occupies a small part in the digitisation activity, and digitisation mainly by manual data entry is most frequent (Appendix 5, Q6-8). Among the few institutions that mainly applied mass digitisation (50-75%, up to 90% of the digitisation activity), three own very large or large collections, one holds a medium-size and one a small collection.

The short questionnaire highlighted that almost all the institutions share the same digitisation priorities as follows (see Appendix 5, Q.9-9e):

1. Scientific relevance:
    1. “Focusing on taxonomic targets”,
    2. “Geographic targets”,
    3. “Museological targets”,
    4. “Global challenges activities”,

2. Institutional relevance:

    The two subcategories “Importance for the museum itself” and “Strategic for national and/or regional programs/projects/guidelines” have the same relevance,

3. Educational relevance:
    1. “Education and training young people”,
    2. “Citizen science initiatives”,
    3. “Other public engagement”,

4. Technical feasibility (resulted between the third and fourth place):
    1. “Ease in specimens handling”,
    2. “Remote digitisation (e.g., from paper catalogues)”,
    3. “Availability of dedicated technologies (e.g., conveyor belt for herbaria and pinned insects)”.

5. Economic Relevance:
    1. “Overall performance in respect to human resources and tools”,
    2. “Overall performance in respect to financial resources”,
    3. “Faster digitisation improving cost/volume rate”)

Therefore, the “Scientific relevance” of a collection is the key element that drives the digitisation, the taxonomic and the geographic relevance are the most important sub-criteria in this category; if the collection has an institutional importance (maybe for funding program) the priority for its digitisation is boosted.

70% of the respondents declared that their institution has a clear overview of the digitisation status (how many specimens are in the database, how many imaged, open access database, etc.) but for most, the database is not in open access. The digitisation status is monitored by automated means only in less than 20%, while the remaining 80% is divided between “no monitoring in place” or “monitoring by extracting the needed information through different databases or sources”. A single CMS is used by a small percentage (28%) whereas 50% do not have a CMS, but use traditional databases (e.g., Access, Excel files) (Appendix 5, Q10-12). This result suggests that, even if it is more appropriate to have a single CMS to better manage all the collections, it is still very difficult to apply a unique CMS for different types of collections, from the geological to the biological ones.

Regarding the information about digitised items (Appendix 5, Q 13-16), 70% of compilers answered that data are validated by the curator and/or by other specialists; of these, 50% answered that data is only partially validated, while the remaining 20% is totally validated. It is interesting that 23% declared they do not have a validation procedure in place. There are clearly needs and opportunities for, creating more links among the institutions to share the expertise for helping in the data validation.

As regards the Minimum Information about Digital Specimen, four levels were defined in the questionnaire:

MIDS0 - Bare: name + unique identifiers (inventory number).
MIDS1 - Basic: MIDS0 + higher taxonomy (to family level) + higher geography (to country level).
MIDS2 - Complete: MIDS1 + label information (collection locality, collector, date)
MIDS3 - Integrated: MIDS2 + external data, not directly available from labels (e.g., bibliography)

(These definitions of MIDS level differ from the more recent version mentioned on p. 18 of the present document)

The answers showed that MIDS3 level has the lowest percentage for almost all the collections (n=41); while MIDS2 is the best «compromise» since it provides considerable information, while not being too demanding. The expected decreasing trend from MIDS0 to MIDS3 was not clear in the replies, probably because some respondents did not answer by following the suggested logic "MIDS0 ≥ MIDS1 ≥ MIDS2 ≥ MIDS3" in the question; observing the single answers, they probably reported the values by subtracting the number of digitised specimens at a level from the total one digitised. Finally, there is a low percentage of imaged items and 3D models, this is probably due to lack of specific tools/technologies and a larger repository for data.

Finally, the replies have highlighted how funding, particularly for employed dedicated staff, is crucial for planning a digitisation strategy.

The multiple-choice questionnaire can be found in Appendix 5.

Case studies on cost

Cost is an important consideration in any digitisation project, it often constitutes a criterion ruling other considerations. We found that most of the published cost analyses of digitisation, including the in-depth analysis made in the context of the ICEDIG project (Hardisty *et al.* 2020) did not consider all the costs involved (pre-digitisation, digitisation *sensu strictu,* post-digitisation). In the two examples summarized below, we have tried to include all stages in the process, from the moment a sample has left the cabinet until it has been safely returned. Perhaps the most important function of the examples is to serve as a checklist of cost items to keep in mind. See also the list of questions to be considered in the recommendation sections below (p. 25).

Example 1. Costs associated with the digitisation of the Greenland herbarium at the Natural History Musem of Denmark

This mass-digitisation project at the Natural History Museum of Denmark (NHMD) was initiated in 2019 and still has not been entirely completed. The project was partly financed by a grant (2.2 million DKK ~ 295,000 euro) from the Aage V. Jensen Charity Foundation, and NHMD invested considerable additional resources from its internal collection budget.

Table 2 presents an overview of the various expenses, and Table 3 gives a detailed example of the data cleaning process.

item	cash cost (euro)	time cost (hs)	notes
Imaging of 147,500 sheets and 15.900 folders	109,150		done by external contractor, paid by grant
Transcription of 170,000 labels	103,700		done by external contractor, paid by grant

Transport of specimens, materials and professional freezing services,	12,500		done by external contractor, paid by grant
Project manager		960	800 hours paid by grant, rest by NHMD
Collection packer		160	paid by grant
Specify manager		303*	small part paid by grant, rest by NHMD
Collection managers		175	paid by NHMD
Student helper		158	partly paid by grant, rest by NHMD
Total	**225,350 euro**	**1581 hours**	**Total cost = cash (euro) plus time (hours)**

Table 2. Expenses associated with the digitisation of the Greenland herbarium at NHMD. Important: the cost for each item consists of cash costs plus time costs; conversion of time (hours) to cash (euro or other currency) has not been attempted. *71,879 out of 170,000 records had been entered into Specify as per August 2022; this required 128 hours. The figure in the table, 303 = 128 × 170,000/71,879

item	time spent	upscaled to 170,000 specimens	notes
cleaning collector names – clustering	60 min		
cleaning taxonomy – clustering	15 min		
cleaning author names	10 min		
cleaning infraspecific taxonomy - clustering	10 min		
cleaning locality – clustering	90 min		variable, depends on original data quality
uploading images	0		usually scheduled to happen during night
Total	3 hours 5 min	130 hours	

Table 3. Example of Specify manager's work on a batch of 4019 sheets

Example 2. Costs associated with the 3D digitisation of the fossil holotypes housed at Museum of Geology and Paleontology of the University of Florence (Italy)

This 3D digitisation was initiated in 2020 and finished in 2022 thanks to Tuscany Region Postdoc Grants in Cultural Heritage 2018 ("POR FSE 2014-2020 Asse A – Occupazione"). This project entitled "Virtual paleontology - a non-invasive approach for the fruition, diffusion and sharing of the paleontological heritage" (PalVirt) was carried out by Dr. Saverio Bartolini Lucenti and was the first example in Italy of the systematic and massive 3D digitisation of paleontological type-specimens, in particular 138 vertebrates (almost all) and 69 invertebrates and plants. Three partners were involved in the project: the Earth Science Dept. – Paleo[Fab]Lab, the Geology and Paleontology Museum, and Tbnet Soluzioni3d srl (Arezzo). For further information, see Bellucci et al. (in press). Table 4 presents an overview of the various expenses.

item	cash cost (€)	time cost (hs)	notes
3D models of 200 fossil specimens (acquisition and elaboration)	*56,000*	792	done by external contractor, paid by grant
Project coordinator		176	paid by NHM UniFi
Collection manager (Project Referent)		352	paid by NHM UniFi

Collection managers		176	paid by NHM UniFi
Total	**56,000**	**1496 hours**	**Total cost = cash (euro) plus time (hours)**

Table 4. Expenses associated with the PalVirt Project.

DISCUSSION

Introduction

The results from both the essay-based and the multiple-choice questionnaire, like the results from the literature studies, highlighted the extreme complexity of prioritisation. Fulfilling the ambition of DiSSCo, to digitise millions of specimens in all possible shapes, sizes, origins, ages, state and value, is indeed a daunting task. The very high number of prioritisation criteria that have been suggested may appear as a barrier to progress for many institutions. An organisation planning a digitisation project needs to consider whether, for example, scientific relevance should be a guiding principle (and what *is* scientific relevance in their specific case?), and/or what the funding opportunities are, and/or what data quality can be obtained with the resources at hand, and/or what the societal interest in the digital specimens to be created is, etc. etc.

With the aim to facilitate decisions about prioritisation of digitisation to be taken by DiSSCo or by individual institutions we here offer a classification of the multitude of possible criteria into four main categories. Based on our literature study and the results of our surveys, we propose the following four categories

- Relevance
- Data quality
- Cost
- Feasibility

All criteria that have been suggested previously fall into one (or more) of the four groups which are thus not new criteria but are meant as an aid to reduce the multi-dimensionality of the "criterion space" during the first steps in the prioritisation process.

The categories of criteria are not completely mutually exclusive. For example, "Cost" may be seen as a component of "Feasibility", and indeed, cost considerations often overrule other criteria. In spite of the somewhat simplistic classification of prioritisation criteria presented above, prioritisation remains a very complex task. It is important to bear in mind that considering just one criterion, or just one category of criteria in isolation will not result in a sound prioritisation. All categories need to be considered, as visualised in Figure 3. Also worth remembering is that prioritisation is not an exact science. Nor is prioritisation constant, but may vary over time, e.g., as policies or funding opportunities change.

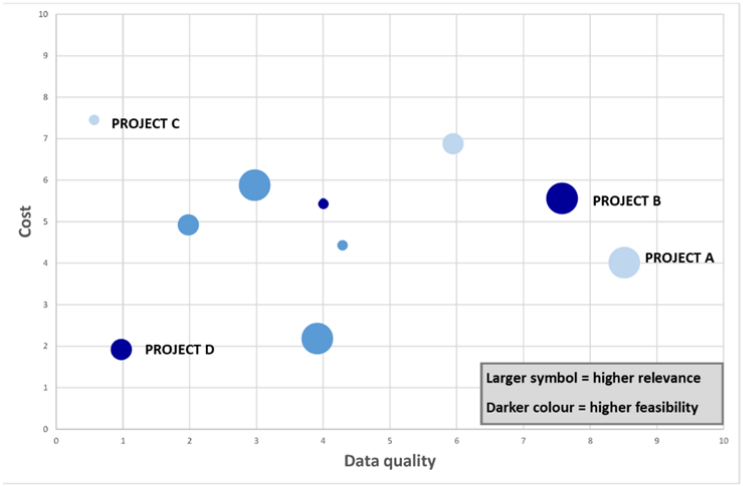


Figure 3. Interrelation of the four main categories of criteria. Data quality and cost are represented on the horizontal and vertical axes (axis values are arbitrary). Relevance is represented by the size of the circles, and feasibility by the intensity of their colour. Project A and B will both deliver data of high quality and high relevance. Although Project B data will be of slightly lower quality, and slightly higher cost, this project may be chosen because of higher feasibility. Project C has little to recommend it, whereas Project D (low data quality, medium relevance and feasibility, and low cost) might be prioritised depending on what the data will primarily be used for.

Relevance

Relevance may be seen as the primary criterion for prioritising digitisation. If the digitised specimens to be generated are of low relevance, i.e., will lead to no benefit or have no impact, other types of criteria (data quality, cost, feasibility) become almost irrelevant.

Different kinds of users have different needs: what is seen as most relevant for one may not be most relevant for another. According to the comprehensive ICEDIG study (Bakker et al. 2018), scientific relevance is deemed most important, at least among the respondents to the ICEDIG's survey, but collection relevance is also important, whereas social and economic relevance are less so. However, depending on the nature of the specimens to be digitised, on the funding possibilities etc., none of these categories of relevance can be neglected. Concerning social/societal relevance, see Deliverable 1-4 from DPP: "Report on socio-economic impact indicators of DiSSCo and DiSSCo-enabled research and research applications", as well as the "Discussion and outlook" chapter in Fitzgerald et al. (2021) and von Mehring et al. (2021). The GBIF study (Krishtalka et al. 2016) agreed with ICEDIG in finding research most important, but disagreed in finding funding/grant opportunities, and taxonomic priorities second and third,

Even "scientific relevance" is a complex thing. See Table 5 for an attempt to visualise the different needs of different scientific disciplines

PRIMARY USE OF DIGITISED SPECIMENS → / TYPES OF INFORMATION INCLUDED ↓	Taxonomic research s.l.	Other types of fundamental research (e.g., biogeographical, ecological)	Applied research (e.g., medical)	Conservation / land use	Outreach
Taxonomy	+	+	+	+	+
Georeference	+	+		+	
Images	+				+
Habitat info	+	+		+	
Sequence data	+	+	+		

Table 5. Type of information to be included in digital biological specimens depending on intended use.

There are two further complexities in relation to using scientific relevance as a guide to prioritisation in DiSSCo. Firstly, it is likely that almost all collections objects where sufficient data are present have scientific relevance against one or more of the types of research mentioned above. Deciding which of these purposes are 'most' important or relevant is extremely challenging. Secondly, this relies on our current understanding of what is important, relevant and useful - but a key benefit sought through digitisation is to unlock new avenues and paradigms of research, for example joining up collections data to other data sources in ways which have not previously been explored. Again, this makes judgements of scientific relevance based on today's evidence inherently flawed, although still worthwhile as one of the criteria to inform prioritisation. Irrespective of how carefully relevance criteria are analysed, nothing is cast in stone. Like prioritisation in general, relevance may change over time as institutions and researchers change their focus.

Data quality

Collection A has 100,000 digital specimens; collection B has 1,000,000. It would seem that collection B is more advanced in terms of digital specimens. But what is the quality of the digital specimens in the two collections? When planning and assessing digitisation, data quality needs to be taken into consideration although this aspect has not been very much considered in previous studies. See Chapman (2005a) for a thorough treatment of the data quality concept.

There are two main dimensions of data quality:

- How much information is there in each digital specimen? (Information level)
- How reliable is that information?

A third essential aspect of data quality is potential for validation and improvement.

- How can we know how reliable our data is, and how can we improve it?

Discussion of data quality is also not independent of the relevance criteria discussed above - the reason data quality is important has to do with whether data are 'research-ready' and impactful. There may be areas of data quality such as high quality geo-referencing that are relevant to widespread fields of research; but other areas of detail which are critical for particular studies but less valuable to widespread users.

Information level

A digitised specimen may be anything from a textual record with minimal information (e.g., species name) to an extended digital specimen represented by full collection information, illustrations in the form of photos and CT scans, morphometric data, DNA sequences, sound recordings, chemical profiles, and with links to related data and resources.

In order to quantify the information level of digital specimens, a digitisation standard has been developed. The Minimum Information about a Digital Specimen (MIDS) standard (Hardisty et al. 2021) comprises three main levels of digitisation plus an initial 'pre-digitisation' level. These levels provide a framework for prioritising, planning, costing, and monitoring a digitisation programme for collections. Using the MIDS standard, the digitisation level of a collection can be scored, and changes can be tracked. The four MIDS levels are shown in Table 6.

The level of information required varies significantly depending on what the data are being used for. Planning and costing a digitisation programme potentially requires a low level of information; some 'big data' analyses, including species distributions, require an additional set of data; whilst taxonomic research may require all the data that are available on the specimen. Mass digitisation programmes are commonly taking a staged approach to capturing information, starting at the basic level (MIDS Level 1) and using a range of options, including outsourcing and crowdsourcing, to transcribe additional data and reach a higher digitisation level. The extended record (MIDS Level 3) equates to the DiSSCo open Digital Specimen specification.

MIDS level	Record extent	Purpose
1	Basic	A basic record of specimen information.
2	Regular	Key information fields that have been agreed over time as essential for most scientific purposes.
3	Extended	Other data present or information known about the specimen, including links to third-party sources.
0 (Note)	*Bare*	*A bare or skeletal record making the association between an identifier of a physical specimen and its digital representation, allowing for unambiguous attachment of all other information.*

Table 6. Four levels of MIDS (Minimum Information about a digital Specimen). From Hardisty et al. (2021).

An example of a digitised specimen with a very high information level is a so-called cybertype. This concept was introduced by Godfray (2007) to denote a digital version of a type specimen which is available online. Considering that requests for access to type specimens constitute a significant fraction of requests for access to natural history specimens, a cybertype may save travel and shipment expenses, as well as time. For example, Akkari et al. (2015), described a new species of millipede. In addition to the physical type specimen, they published a cybertype including interactive CT scans of the same specimen (Figure 3). The scans have subsequently been used by Naumann et al. (2019) for a study on millipede feeding mechanisms.

However, while digitisation of type specimens to a high level of detail has many benefits, it does not enable 'big data' type analyses such as species distributions which are critical to understanding environmental change - it is likely that a balance is required in prioritisation between detailed data on some specimens and lower levels of data on many specimens.

Figure 3. Part of the millipede cybertype described by Akkari et al. (2015). The image shows the anterior part of the body with mouthparts and copulatory organs highlighted. The cybertype may be manipulated to show details important for, e.g., taxonomy.

Reliability

Reliability (data quality in the strict sense) was treated in detail by Chapman (2005a). The data that DiSSCo deals with to a high degree includes species-occurrence information, i.e., records of a particular species from a particular place. A typical species-occurrence data point includes taxonomic/nomenclatural information (which species, subspecies or other taxon), geographical information, collector and collecting date information and often also other descriptive data such as habitat, host plant etc.

For all these components of a data point, but especially obvious for spatial data, their accuracy and precision need to be considered. Accuracy and precision are often confused: Accuracy refers to the closeness of measured values, observations or estimates to the real or true value, whereas precision includes statistical precision (the closeness with which repeated observations conform to themselves) and numerical precision (the number of significant digits that, e.g., decimal latitude/longitude is recorded in (Chapman 2005a). The difference between accuracy and precision of species-occurrence data is shown in Figure 9. The accuracy and precision can also be applied to non-spatial data. For example, a collection may have an identification to subspecies level (i.e. have high precision), but be the wrong taxon (i.e. have low accuracy), or be [correctly] identified only to family level (high accuracy, but low precision) (Chapman 2005a).

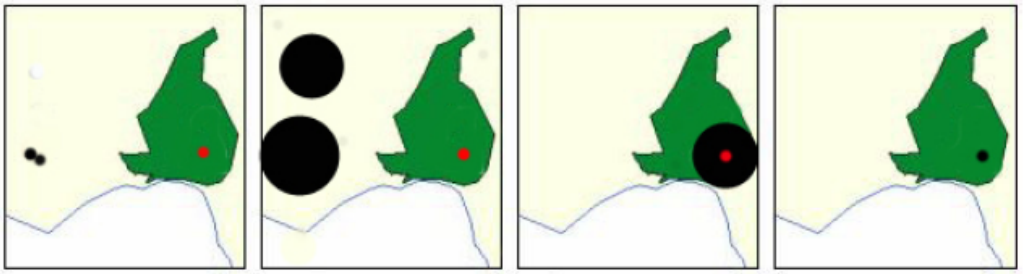


Figure 4. Shows the differences between accuracy and precision in a spatial context. The red spots show the true location, the black spots represent the locations as reported by a collector.

Far left. High precision, low accuracy.
Middle left. Low precision, low accuracy showing random error.
Middle right. Low precision, high accuracy.
Far right. High precision and high accuracy.
From Chapman (2005a)

Ideally, all data points would have high accuracy and high precision. However, for some purposes, high precision is not necessary for the data to be "fit for use". This is illustrated in Figure 5. The figure refers to spatial data, but "fitness for use" considerations also apply to other types of information. For example, for some purposes, identification to subspecies level is necessary, whereas for others, species level is sufficient. Also, for some purposes, year of collection is sufficient, whereas for others, the exact date, or at least month, is required.

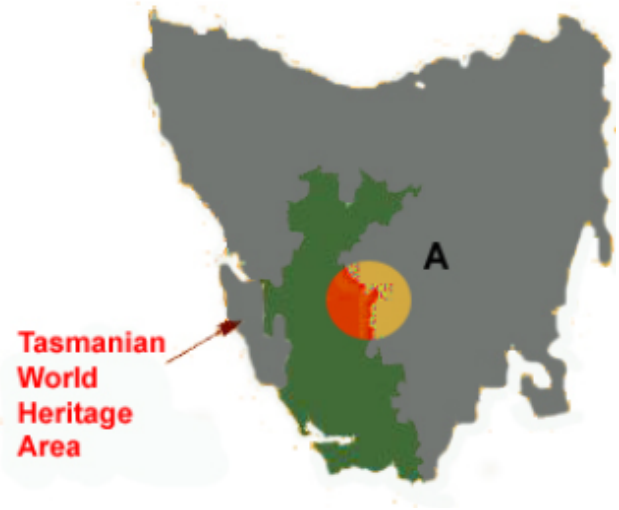


Figure 5. Map of Tasmania, Australia, showing a record (A) collected with a precision of 0.5º (ca. 50 km) as shown by circle. The footprint area of possible collection (determined using the precision value) overlaps the Tasmanian World Heritage Area. The record is suitable for answering the question whether the species in question occurs in Tasmania (fit for use), but it is not suitable for answering the question whether the species occurs in the Tasmanian World Heritage Area marked in green (unfit for use). From Chapman (2005a).

Assessing and improving data quality

Irrespective of how carefully a dataset has been prepared, very few datasets – if any at all – are guaranteed error-free. Therefore, quality assessment and data cleaning are important aspects of digitisation.

For DiSSCo, four types of information are particularly relevant, viz., 1) taxonomic and nomenclatural information, 2) spatial information (georeferencing), 3) collection date, 4) image quality. For fossils, 5) geological age is also essential. Concerning type 1–3, data cleaning was treated in detail by Chapman (2005b), with emphasis on 1) and 2). Just as the digitisation process itself needs prioritisation according to the four main categories of criteria the data validation and cleaning process needs to be prioritised according to criteria of relevance, cost, and feasibility.

Quality control should be done by experts with access to both the physical and digitised collections. When voucher specimens are kept in a collection, the accuracy and precision of the taxonomic/nomenclatural information can be checked by a specialist at any time, but this seldom applies to the accuracy and precision of data on location, date, collector, habitat etc. Hence a great responsibility for accuracy and precision in recording rests on the collectors themselves. An alternative approach is to use a range of online tools such as the data quality control checks within aggregators such as the Global Biodiversity Information Facility (GBIF) and SpeciesLink, which include checks on geocoordinates, taxon names and date formats. GBIF also provides a list of tools which include a number to support assessing and improving data quality

(https://www.gbif.org/resource/search?contentType=tool). Bionomia is an online resource which has automated the process of parsing and cleaning names of collectors and determiners and finding associated specimens, using integrations with GBIF, Wikidata, ORCID and Zenodo. This enables the discovery of errors or inconsistencies in specimen data relating to collectors and determiners (https://bionomia.net/) (Shorthouse, 2020).

Manual data cleaning, e.g., by taxonomic specialists or curators, will continue to be important. For example, the identification of collectors' itineraries ... allows for checking for possible error if, for example, the date of collection doesn't fit the particular pattern of that collector (Chapman 2005b).

In the framework of the SYNTHESYS+ project Walton et al. (2020a) made a "landscape analysis" for the Specimen Data Refinery that will become one of DiSSCo's e-services. See also Dillen et al. (2019), especially their chapter 3. This DiSSCo Prepare deliverable deals with the semantic enhancement of digital specimens, with emphasis on taxonomic names, geographical features of the specimen and names of persons (collectors, identifiers etc.) associated with the specimen.

Finally, as always, a balanced view is recommendable. It is better to release imperfect data than to hold data back in the pursuit of (impossible?) perfection. Releasing (imperfect) digital data can help to improve data quality, e.g., by opening it up to comment from international experts remotely.

Cost

Cost considerations including funding opportunities will have a big impact as to what is prioritised in a digitisation project. The cost of digitisation has been the subject of many analyses – recent examples are Tegelberg et al. (2017), Hardisty et al. (2020), Medina et al. (2020), Walton et al. (2020b), also the upcoming costbook of DiSSCo (DPP deliverable 4.1). A general lesson from these analyses is that it is impossible to give a simple figure for "What does it cost to digitise a specimen?" The desired data quality, the level of infrastructure already available, as well as salary levels for different categories of people in different countries, all play a role in cost considerations.

Hardisty et al. (2020) analysed the different types of costs based on information from seven natural history collection institutes in Europe and described the different types of costs to be considered:

- Capital costs, such as the purchase of equipment, buildings.
- Fixed operating costs (i.e., operating costs which are not dependent on the level of usage of the facility), such as maintenance contracts, some salaries, building/floor rental, heating and lighting, etc.
- Variable operating costs (i.e., operating costs which depend on the level of activity), such as per hour costs of staff carrying out digitisation tasks, barcode labels and other consumable materials.

Another useful classification described by Hardisty et al. (2020) divides costs into:

- Establishment costs, meaning the upfront costs of building and equipping a digitisation facility.
- Costs of digitising specimens.
- Costs of preserving the digitised data and making it findable, accessible, interoperable, and re-usable (i.e., 'FAIR').

Especially the cost of preserving the digitised data is often neglected or underestimated although it may constitute a very significant part of the digitisation costs. See, for example, the case studies of costs in the results section of the present report. While cost, including funding opportunities, is likely to be critical to any decision to undertake digitisation, focusing on this criterion alone is problematic DiSSCo prioritisation only of the cheapest specimens to digitise is unlikely to be desirable. Cost needs to be taken into account alongside the other criteria, and is perhaps better expressed and understood as 'value for money' - the most advantageous combination of cost and quality, or in other words whether it is cost-effective to digitise certain things, because there is a feasible workflow; scientific or other relevance that will make the data impactful; sufficient data available; and funding to meet the expected costs. Cost data will be added to some of the workflows in DiSSCo's digitisation guides website (https://dissco.github.io/) and to the "digit-key" (https://digit.naturalheritage.be/digit-key) being developed by the Royal Belgian Institute of Natural Sciences.

Feasibility

The feasibility of a digitisation project is, of course, dependent on available funds. In other words, cost might be seen as one aspect of feasibility. But cost considerations aside, there are other factors that determine a project's feasibility: Is the collection ready to be digitised? Are skilled staff available? Is the IT and other technical infrastructure geared to the task? Has a digitisation workflow been tested and established at a suitable scale?

De Smedt et al. (2022) provide a useful checklist for "pre-digitisation curation" as a contribution to the DiSSCo Digitisation Guides website (https://dissco.github.io/). "Skilled staff" not only refers to the people who do the digitisation. These people should, of course, know how to handle the sometimes fragile specimens; ideally, they would also possess some knowledge of the organisms they are digitising, and of the collection in which the specimens reside. In addition to the "hand-on" digitisation staff, it is important that people with extensive knowledge of the organisms to be digitised are available, in order to ensure a high quality of the digitised data. For historical collections, knowledge on the relevant collections, collectors, expeditions etc. is also necessary.

"IT and other technical infrastructure" includes such things as cameras/scanners, conveyor belts etc., but also computing power, appropriate software, storage space, back-up options.

The human and other resources necessary for a successful project vary according to the type of specimen. It has become well-known that digitisation, including mass digitisation of herbarium sheets, is relatively easy. For collections of dried insects (which in terms of sheer specimen numbers constitute a very large, if not the largest part of DiSSCo's collections) methods are being developed for efficient mass digitisation of the specimens and the associated labels (Tegelberg et al. 2017, Price et al. 2018, Wu et al. 2019). Also, an automated mass digitisation workflow for microscope slides has been prepared (Allan et al. 2019). Wet-preserved specimens such as invertebrates stored in jars with alcohol, or in glass tubes which are in turn stored in jars, pose a huge challenge in terms of human and other resources, but see Dupont et al. (2020).

The human and other resources necessary for a successful project also vary according to the desired level of data quality, including information level (e.g., MIDS), accuracy and precision.

Many especially smaller institutions will have difficulties with mustering the necessary resources to make a digitisation project feasible. Collaboration may ameliorate this situation. DiSSCo provides a unique opportunity, not only for sharing and learning from best practice workflows which can improve feasibility, but for direct collaboration on digitisation. The efficiency and potential impact of the digitisation of natural history collections will be immensely higher if DiSSCo-wide agreements can be made. At the DiSSCo level it may also be possible to apply for European funds to carry out large-scale digitisation projects. DiSSCo-wide digitisation targets could be of the following types (hypothetical examples):

- X % of all herbarium sheets in DiSSCo collections databased and imaged before 20XX.
- All primary types of insects in DiSSCo collections databased to MIDS level X before 20XX.
- All African birds in DiSSCo collections databased and imaged before 20XX
- ….

Implementing the criteria

Despite the complicated nature of the matter, the "academic" presentation of various types of criteria for prioritisation is relatively straightforward. In contrast, their practical implementation is anything but straightforward. All analyses show that there is no such thing as one primary criterion taking precedence over others. Bakker et al. (2018) outlined three methods to implement prioritisation criteria for digitisation, viz.

1. A decision tree (not a tree, but an electronic multi-entry key), focusing on practical (feasibility) and funding (cost) criteria.
2. A scoring method.
3. A panel review.

Concerning the decision tree Bakker et al. (2018) referred to an "Appendix 6" which, however, is not included in their report. We have had access to an incomplete draft of this appendix in the form of an extensive Excel sheet. It is obvious that constructing an operational decision tree or multi-entry key will be extremely complicated, if possible at all, even if the scope of the tree/key will be limited to feasibility and cost criteria. Therefore, we have focused on the scoring and panel methods. As pointed out by Bakker et al. (2018) these can be used one at a time, or in combination, and, based on the experience from the SYNTHESYS projects[2]) a combination does indeed look like the best solution.

[2] SYNTHESYS (https://www.synthesys.info/about-synthesys.html) has run successfully from 2004 to 2023 and as a core activity has funded short transnational research visits to a considerable number of European collections. In the latest version of the project, SYNTHESYS+, a virtual access grant scheme to fund smaller digitisation projects of the collections was included as well. Applications for transnational and virtual access in SYNTHESYS are prioritised and funded based on a combination of scoring and panel review. Applications are submitted to the newly developed ELViS platform (https://elvis.dissco.eu/welcome). The application is a structured form, and applications are evaluated and scored by a panel of experts. Importantly, prioritisation and funding are not decided on the basis of the panel scores alone but are discussed at a panel meeting where aspects that cannot easily be assigned a numerical score can also be discussed and considered

CONCLUSION AND RECOMMENDATIONS

- with a list of questions that might be included in a scoring sheet for digitisation projects

When DiSSCo RI becomes fully operational it is expected that prioritisation of digitisation will, at least in part, take place at DiSSCo level. Whereas it is beyond the scope of the present report to suggest which specimens to digitise first, the preceding sections provide a background for making optimal decisions.

When choosing what to digitise, and how to do it, consider:

- Where possible, collaboration on digitisation proposals, particularly within the DiSSCo framework. We support using the community itself and the rapid developments in approaches which are happening around the world as a solution in itself to help drive forward strategic prioritisation of digitisation activities. Communicating summaries of these and adding to these will have a dual role in helping others define or refine their strategies.
- Aiming to provide data that are sufficient for the use case within the project, whilst considering other likely use cases and paying attention to data quality. Biodiversity data quality is likely to affect downstream analyses, reports and decisions made based on the data, and a consistent approach to assess and manage data quality will be required.
- Using a combined approach of scoring and panel review, allowing for a balanced and nuanced implementation of the prioritisation criteria.

More specifically, consider:

- Relevance, including
  - scientific relevance
  - societal relevance

- Data quality, including
    - level of information
    - reliability
  - potential for validation
  - dataset completeness

- Cost, including
  - pre-digitisation
  - digitisation *s.s.*
  - post-digitisation

- Feasibility, including
  - possibilities for collaboration

Questions to be asked

To gather the information required for prioritisation, whether for evaluation or preparation of project proposals or for preparing an internal strategy, the following questions are recommended.

RELEVANCE:

- What is the scientific relevance of the project? (Which types of research will be facilitated by the generated digital data?)
- What is the socio-economic relevance of the project? (Which economic and social benefits will result from the project? Will the project support national/European/global political goals, including the 17 Sustainable Development Goals of the UN?

COST

- Is the cost/benefit ratio of the project reasonable? ("value for money")
- Are all steps in the digitisation process considered?
- Is sufficient funding available?
    - If not, is there a realistic plan for obtaining sufficient funding?

QUALITY

- Is the level of information (e.g., MIDS) of the generated digital data sufficient for the purpose of the project?
- Is the accuracy and precision of the generated digital data sufficient for the purpose of the project?
- Is long-term storage and FAIR availability of the digital data ensured?
- Is there a plan for data validation / quality control / data enhancement?

FEASIBILITY

- Is the necessary IT infrastructure available?
    - If not, is there a realistic plan for gaining access to the necessary IT infrastructure?
- Is the necessary technical infrastructure (e.g., cameras, scanners, conveyor belts) available?
    - If not, is there a realistic plan for gaining access to the necessary technical infrastructure?
- Is the necessary scientific (e.g., taxonomic experts, curators) and technical (e.g., IT) staff available?
    - If not, is there a realistic plan for making such staff available?
- Is there scope for joining forces with other projects?

A final word

Finally, whereas prioritisation of digitisation is the subject of the present report, it is important to remember that the digital specimens that have been and will be created, still need links to the physical specimens since physical specimens always will be the ultimate (potential) validators for digital ones. Irrespective of the "digital revolution" in which DiSSCo takes part, physical collections therefore will need continued funding, including funding for skilled curators. This priority for digitisation of natural history collections is as high as any other.9

Acknowledgements

We extend our thanks to all those persons who have provided information, either in the form of response to our surveys, or in a more informal way. Particular thanks to Tim Robertson from the GBIF secretariat, and to Arthur Chapman, Australian Biodiversity Information Services for permission to use illustrations from GBIF reports.

References

Ahl LI& Enghoff H 2022. Corpus of previous studies on prioritisation of digitisation compiled. DiSSCo Prepare Project Milestone Report.

Akkari N, Enghoff H & Metscher B 2015. A new dimension in documenting new species: High-detail imaging for myriapod taxonomy and first 3D cybertype of a new millipede species (Diplopoda, Julida, Julidae). *PLoS ONE* 10(8): e0135243. http://dx.doi.org/10.1371/journal.pone.0135243

Allan EL, Livermore L, Price BW, Shchedrina O & Smith VS 2019. A novel automated mass digitisation workflow for natural history microscope slides. *Biodiversity Data Journal* 2019. 7: e32342. doi: 10.3897/BDJ.7.e32342.

Bakker HPAJ, Willemse L, van Egmond E, Casino A., Gödderz K & Vermeersch X. 2018. Inventory of criteria for prioritization of digitization of collections focussed on scientific and societal needs. A deliverable (D2.1) of the ICEDIG project. https://doi.org/10.5281/zenodo.2579156

Bellucci L, Bartolini-Lucenti S. Dominici S, Rook L & Cioppi E, in press. Digitalizzazione 3D delle collezioni paleontologiche del Museo di Geologia e Paleontologia di Firenze. *Museologia Scientifica Memorie.*

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

Chapman AD 2005b. Principles and Methods of Data Cleaning – Primary Species and SpeciesOccurrence Data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. https://assets.ctfassets.net/uo17ejk9rkwj/46SfGRfOesU0IagMMAOIkk/1c03ea3e21fcd9025cc800d786890e72/Principles_20and_20Methods_20of_20Data_20Cleaning_20-_20ENGLISH.pdf

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

Dillen M, Groom Q, Cubey R, von Mering S & Hardisty A 2019. A best practice guide for semantic enhancement and improvement of semantic interoperability. DiSSCo Prepare deliverable 5.4 report. doi: 10.34960/ajxs-zr25

Dupont S, Humphries J, Butcher AJ, Baker E, Balcells L & Price BW 2020. Ahead of the curve: three approaches to mass digitisation of vials with a focus on label data capture. *Research Ideas and Outcomes* 6: e53606. https://doi.org/10.3897/rio.6.e53606

Fitzgerald H, Juslén A, von Mering S, Petersen M, Raes N, Islam S, Berger F von Bonsdorff T, Figueira R, Haston E, Häffner E, Livermore L, Runnel V, De Smedt S, Vincent S & Weiland C 2021. DiSSCo Prepare Deliverable D1.1 Report on Life sciences use cases and user stories. https://doi.org/10.34960/xhxw-cb79

GBIF (accessed 4/11-2022) https://www.gbif.org/occurrence/search?basis_of_record=MATERIAL_SAMPLE&basis_of_record=PRESERVED_SPECIMEN&basis_of_record=FOSSIL_SPECIMEN&basis_of_record=LIVING_SPECIMEN&advanced=1&network_key=17abcf75-2f1e-46dd-bf75-a5b21dd02655

Godfray HCF 2007. Linnaeus in the information age. *Nature* 446: 259-260. doi: 10.1038/446259a.

Groom Q, Dillen M, Hardy H, Phillips S, Willemse L & Wu Z 2019 Improved standardization of transcribed digital specimen data. Database 2019:baz129. https://doi.org/10.1093/database/baz129

Hardisty A, Livermore L, Walton S, Woodburn M, Hardy H 2020. Costbook of the digitisation infrastructure of DiSSCo. *Research Ideas and Outcomes* 6: e58915. https://doi.org/10.3897/rio.6.e58915

Hardisty A, Addink W, Dillen M, Groom Q, Haston E et al. (Draft) Minimum Information about a Digital Specimen (MIDS) v0.14, 29 Mar 2021. https://github.com/tdwg/mids/blob/working-draft/current-drafts/MIDS-definition-v0.16-28May2022.md

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.

Medina JJ, Maley JM, Sannapareddy S, Medina NN, Gilman CM & McCormack JE 2020. A rapid and cost-effective pipeline for digitisation of museum specimens with 3D photogrammetry. *PLoS ONE* 15(8): e0236417. https://doi.org/10.1371/journal.pone.0236417

Naumann B, Reip H, Akkari N, Neubert D & Hammel JU 2019. Inside the head of a cybertype – three-dimensional reconstruction of the head muscles of Ommatoiulus avatar (Diplopoda: Juliformia: Julidae) reveals insights into the feeding movements of Juliformia. *Zoological Journal of the Linnean Society*,2019 , zlz109 https://doi.org/10.1093/zoolinnean/zlz109

Price B, Dupont S, Allan EL, Blagoderov V, Butcher A, Durrant J, Holtzhausen P, Kokkini P, Livermore L, Hardy H & Smith V 2018. ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation. Preprint from the ALICE project: https://www.doi.org/10.17605/OSF.IO/UVWRN

Shorthouse DP 2020. Slinging With Four Giants on a Quest to Credit Natural Historians for our Museums and Collections. *Biodiversity Information Science and Standards* 4: e59167. https://doi.org/10.3897/biss.4.59167

Societal Challenges, Horizon 2020. European Commission. Available online at: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges#Article

Tegelberg R, Kahanpää J, Karppinen J, Mononen T, Wu Z & Saarenmaa H 2017. Mass Digitisation of Individual Pinned Insects Using Conveyor-Driven Imaging. IEEE 13th International Conference on eScience. https://doi.org/10.1109/eScience.2017.85

van Egmond E, Willemse L, Paul D, Woodburn M, Casino A, Gödderz K, Vermeersch X, Bloothoofd J, Wijers A & Raes N. 2019. Design of a Collection Digitisation Dashboard. Deliverable 2.3 from ICEDIG. https://doi.org/10.5281/zenodo.2621055

Veiga AK, Saraiva AM, Chapman AD, Morris PJ, Gendreau C, Schiegel D & Robersen TJ 2017. A conceptual framework for quality assessment and management of biodiversity data. *PLoS ONE* 12(6) eo178731. https://doi.org/10.137/journal.pone0178731

von Mering S, Petersen M, Fitzgerald H, Juslén A, Raes N, Islam S, Berger F von Bonsdorff T, Figueira R, Haston E, Häffner E, Livermore L, Runnel V, De Smedt S, Vincent S. & Weiland, C 2021. DiSSCo Prepare Deliverable D1.1 Report on Life sciences use cases and user stories. https://doi.org/10.34960/n3dk-ds60

Walton S, Livermore L, Bánki O, Cubey RW, Drinkwater R, Englund M, Goble C, Groom Q, Kermorvant C, Rey I, Santos CM, Scott B, Williams AR & Wu Z 2020. Landscape analysis for specimen data refinery. *Research Ideas and Outcomes* 6: e57602. https://doi.org/10.3897/rio.6.e57602

Walton S, Livermore L, Dillen M, De Smedt S, Groom Q, Koivunen A, Phillips S 2020. A cost analysis of transcription systems. *Research Ideas and Outcomes* 6: e56211. https://doi.org/10.3897/rio.6.e56211

Wu Z, Koivunen A, Saarenmaa H, van Walsum M, Wijers A, Willemse L & Ylinampa T 2019. State of the art and perspectives on mass imaging of pinned insects. Deliverable D3.5 from ICEDIG, https://know.dissco.eu/bitstream/item/86/1/Deliverable%20D3.5_ICEDIG_StateOfArt_digitisation_of_pinned_insects.pdf

Overview of appendices

Appendix 1. List of acronyms

Appendix 2. Short descriptions of the ICEDIG and GBIF report

Appendix 3. Additional literature searches

Appendix 4. Survey 1, M1.3 – DPP milestone report

Appendix 5. Survey 2, Multiple-choice questionnaire

Appendix 1. Acronyms

DPP	– DiSSCo Prepare Project, https://www.dissco.eu/dissco-prepare/
DiSSCo	– Distributed System of Scientific Collections, https://dissco.eu
FAIR	– Findable, Accessible, Interoperable and Reusable, https://www.go-fair.org/fair-principles/
GBIF	– Global Biodiversity Information Facility, https://gbif.org
ICEDIG	– Innovation and Consolidation for large scale Digitisation of Natural Heritage, https://icedig.eu
NHMD	– Natural History Museum of Denmark
NSC	– National Science Consortium
RI	– Research Infrastructure
SYNTHESYS	– Synthesis of Systematic Resources, https://www.synthesys.info/
UniFi	– University of Florence

Appendix 2 - ICEDIG and GBIF report

ICEDIG Report

ICEDIG was an EU funded project under the Horizon 2020 Framework. Deliverable D2.1 of the ICEDIG project (Bakker et al. 2018) provides an impressive corpus and analysis of digitisation criteria. It forms a very substantial part of the basis for this final deliverable of DPP Task 1.3.

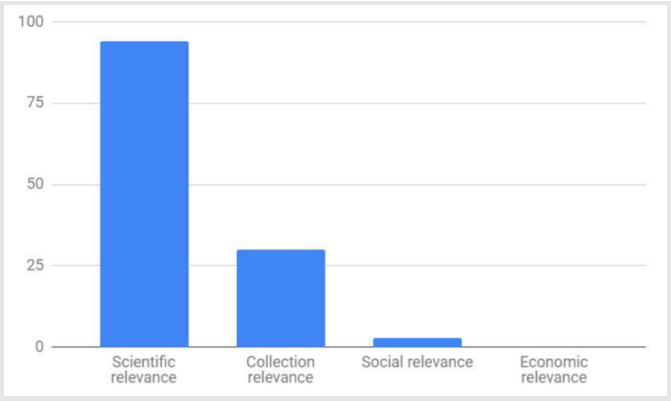
The ICEDIG project started January 1, 2018, and the project deliverable D2.1 was the report "Inventory of criteria for prioritization of digitization of Natural History Collections" (Bakker et al. 2016). The aim of this deliverable was to contribute to an "easy and well-informed decision-making process in relation to prioritisation of digitisation of natural history collections". In ICEDIG it was decided to follow a multi-stage process to ensure that the solutions put forward were solid regarding the prioritisation of digitisation of natural history collections. Stages identified:

3. **A literature and reports inventory** was carried out to create an overview of the criteria of prioritisation of digitisation.
    - Specific internet searches were aimed at obtaining relevant scientific literature and publications. The results of the searches were compiled.
    - Input from relevant organizations were sought and obtained from E-ReColNat, iDigBio and ICEDIG partners.
    - All data was compiled and presented in appendices to the report.

4. Targeted survey.
    - A survey with 20 questions was widely distributed to relevant institutions. Replies came from a range of natural history collections across Europe (including Russia, Turkey, Israel, and the Caucasus Republics). A total of 197 responses were received from 160 different institutions representing 34 European countries. Unfortunately, not all had completed or correctly filled out the survey, why only 68 responses could be used for the statistical analyses at the end.
    - The survey was divided into three overall themes covering: "general topics", "list of criteria" and "additional information in free text".

For the questions regarding prioritisation 68 completed responses were included in the depictions of the data shown in figures 1 to 5. Figure 1 gives the overview of the ranking of the four areas of relevance identified: scientific, collection, social, and economic. Figures 2 to 5 show the ranking of the criteria used in the questionnaire identified for each of the four areas.

Figure 1. Overview of the relative importance of the relevance areas identified regarding digitisation.



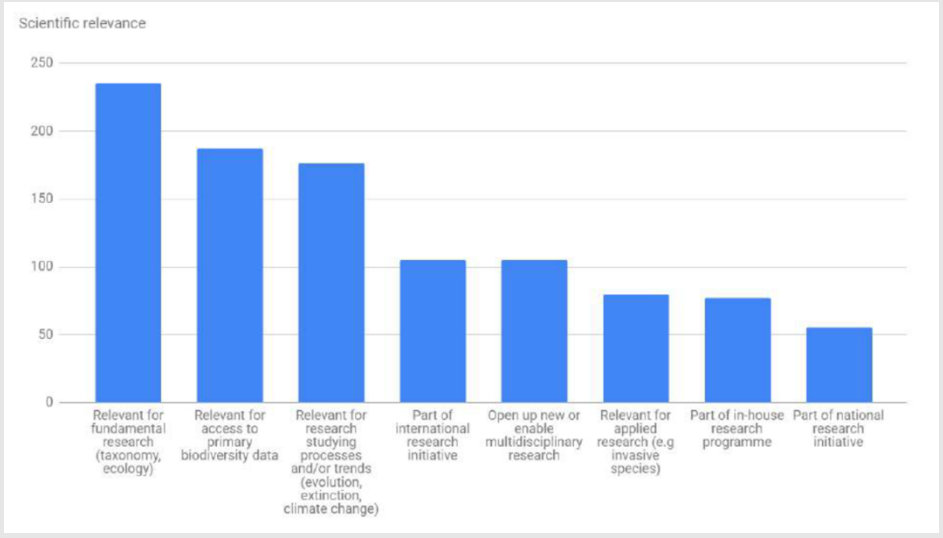


Figure 2. Science-relevant criteria arranged after relevance.

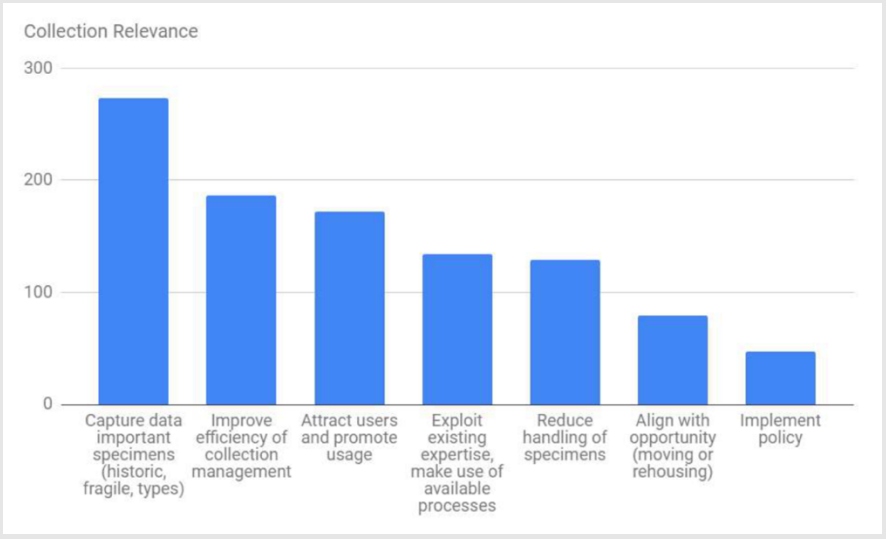


Figure 3. Collection-relevant criteria arranged after relevance.

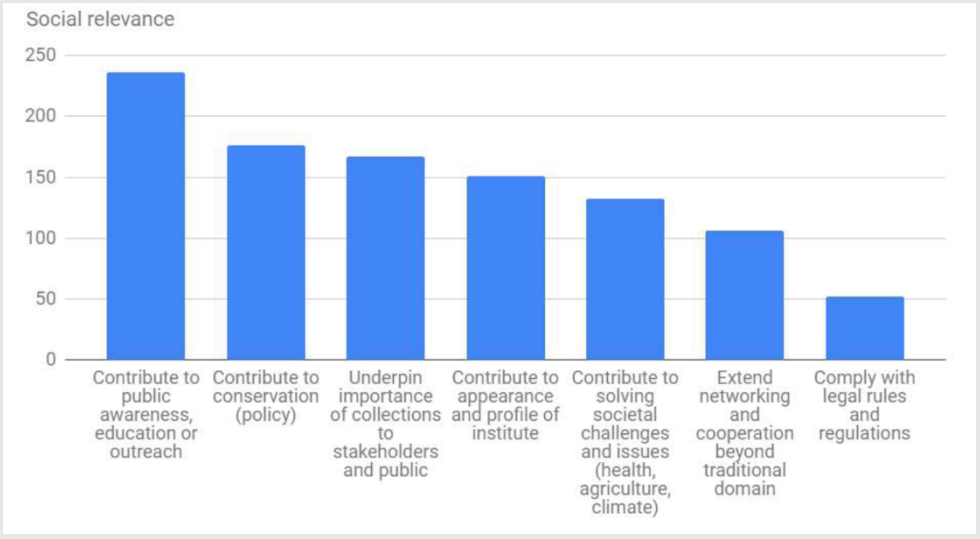


Figure 4. Society-relevant criteria arranged after relevance.

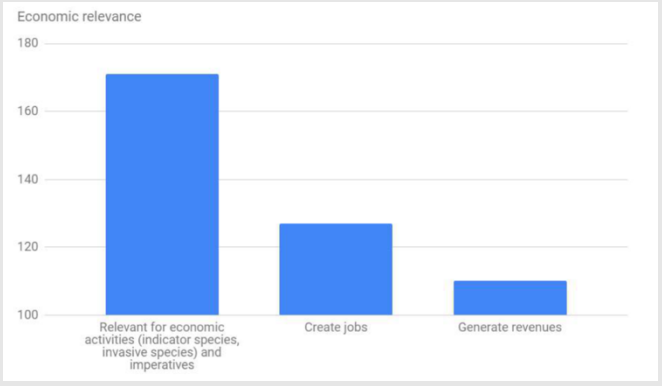


Figure 5. Economy-relevant criteria arranged after relevance.

A brief summary of the three most important aspects for each of the relevance areas:

- Scientific relevance
    - Fundamental research (taxonomy and ecology)
    - Access to primary biodiversity data
    - Relevant for research studying processes and/or trends (evolution, extinction, climate change)
- Collection relevance
    - Capture data important specimens (historic, fragile, types)
    - Improve efficiency of collection management
    - Attract users and promote usage
- Social relevance
    - Contribute to public awareness, education, or outreach
    - Contribute to conservation (policy)
    - Underpin importance of collections to stakeholders and public
- Economy
    - Relevant for economic activities (indicator species, invasive species) and imperatives
    - Create jobs
    - Generate revenues

5. Results revised

Based on the additional information added in free text, an extensive and revised list of criteria was assembled on six overarching topics:

7. collection relevance
8. economic relevance
9. funding
10. practical criteria
11. scientific relevance
12. social relevance.

6. Advisory methods identified

Due to the broad range of criteria that were identified to be of importance in the process of prioritising digitisation efforts, three possible methods to determine the strategy for a digitisation project were proposed: 1) Decision tree; 2) Scoring method, and 3) Panel review.

Concluding remarks and recommendations

The concluding remarks and recommendations by the ICEDIG team emphasize the need to work from a very clear set of criteria. This approach will enable the institution to set the priorities of which and what parts of natural history collections need to be digitised first. It is highlighted that in the prioritisation of digitisation the initial focus should be on stimulating fundamental research (e.g., taxonomy), followed by research on processes (e.g., evolution) and finally enhancing access to primary biodiversity data. It is also emphasized that in the process of deciding which specimens to digitise first in a scientific context it will be essential to further prioritise digitisation. The recommendation here is to set up a panel of biodiversity scientists representing different fields.

However, before any prioritisation process is set up it is crucial to determine the practical aspects that it entails in terms of resources and funding. It is recognized that institutions will have to balance their final decisions on “competing interests at an international, national and institutional scale”. It is further recommended that within DiSSCo it will be important to make a guideline on how to

determine what to prioritise in terms of digitisation of collection specimens – a digital research agenda.

An obvious lesson from the ICEDIG report is that scientific relevance was deemed most important by the respondents to the targeted survey, followed at a distance by collection relevance. Social relevance and economic relevance were not regarded as important in this survey. See part 4 of the present report for a discussion of relevance.

GBIF task Force report

Although predating the ICEDIG report the paper "Accelerating the discovery of biocollections data." (Krishtalka et al., 2016) is briefly summarized here because its scope is so similar to ours. A task force was convened by GBIF "to help accelerate the discovery, digitisation and access to biocollections data". One of the task force's main objectives was to provide guidance on establishing priorities for digitising biocollections to serve institutional, national, and global needs and achieve the greatest economies of scale (Krishtalka et al., 2016). This is equivalent to the objective that the present document addresses – *Establish relevant criteria to identify a prioritisation model for digitisation.*

The GBIF task force undertook a large-scale, global survey among collection-holding institutions on the state and prioritisation of digitisation. A total of 519 respondents gave information on their priorities, and these are presented in Figure 6. The most important priorities identified by the GBIF task force were reported to be 1) research and 2) funding/grant opportunities, followed by 3) taxonomic priorities.

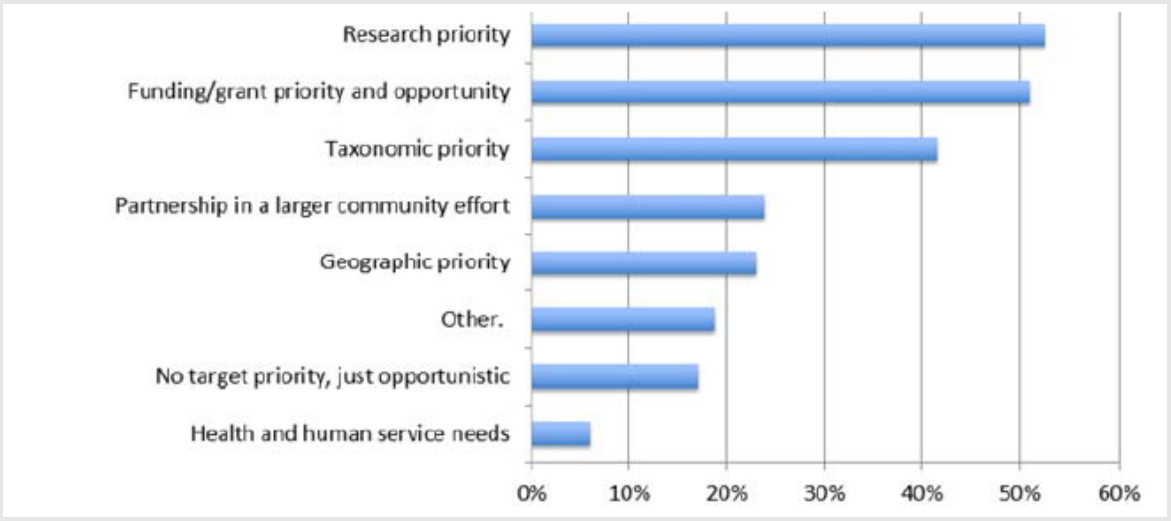


Figure 6. Percentages of collections surveyed by GBIF applying various criteria for prioritization of collections, from Krishtalka et al (2016).

The quintessence of the GBIF study: that 1) research, 2) funding/grant opportunities, and 3) taxonomic priorities are most imoportant are only in part compatible with the most important criteria found by ICEDIG. This serves to underline the extreme complexity of the prioritisation. See section 4 for more discussion of this.

Appendix 3 – Additional literature searches

The combined list of new and relevant studies found through these two searches were:

- Daru et al (2017) Widespread sampling biases in herbaria revealed from large-scale digitisation
- Veiga et al (2017) A conceptual framework for quality assessment and management of biodiversity data
- Cantrill (2018) The Australasian Virtual Herbarium: Tracking data usage and benefits for biological collections.
- Nelson et al (2018) The history and impact of digitisation and digital data mobilization on biodiversity research
- Schindel et al (2018) The next generation of natural history collections
- Wetzel et al (2018) Unlocking biodiversity data: Prioritisation and filling the gaps in biodiversity observation data in Europe
- Hereld et al (2019) LightningBug ONE: An experiment in high- throughput digitisation of pinned insects
- Nekola et al (2019) Caveat consumptor notitia museo: Let the museum data user beware
- Willemse et al (2019) Future Challenges in Digitisation of Private Natural History Collections
- Hedrick et al. (2020) Digitisation and the future of natural history collections.
- Miller et al (2020) Building Natural History Collections for the Twenty-First Century and Beyond
- Paton et al (2020) Plant and fungal collections: Current status, future perspectives

Conclusions

Analysis of the twelve newer studies highlighted by our literature search added little to the previous reports by ICEDIG (Bakker et al. 2018) and the GBIF taskforce (Krishtalka et al. 2016) in terms of prioritisation criteria. Disappointing as this may seem, it testifies to a high quality of the mentioned reports. However, two general recommendations may be distilled from the new literature.

Firstly, coordination and integration at the international level will be essential if the ambitious goal of a global collection of digital specimens is ever to be reached.

Secondly, more attention needs to be paid to data quality. Biodiversity data quality is likely to affect downstream analyses, reports and decisions made based on the data, and a consistent approach to assess and manage data quality will be required. This subject is addressed in Part 4 of the present report.

Relevant articles from search June 2021

Digitisation experience report

These reports are mainly based on the herbarium mass digitisation. If rarely the process is explicitly described, it often present institutional interest in some guideline or biases produced by the process.

1. Berger, F. (2019). Describing the German Research Infrastructure DCOLL Based on the Criteria Defined by the One World Collection Group-a test case. Biodiversity Information Science and Standards.

2. Bishop, A. M., Whalert, G. A., & Seltmann, K. C. (2021). Digitisation of the UCSB Herbarium's Seaweed Collection Provides Vital Data to Better Understand the Changing Marine Environment. (2nd Quarter).

3. Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J., ... & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitisation. New Phytologist, 217(2), 939-955. https://doi.org/10.1111/nph.14855

4. Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J., ... & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitisation. New Phytologist, 217(2), 939-955. https://doi.org/10.1111/nph.14855

5. De la Porte, B., & Higgs, R. (2019). Challenges in digitisation of cultural heritage material in the Western Cape, South Africa. South African journal of information management, 21(1), 1-11. https://journals.co.za/doi/epdf/10.4102/sajim.v21i1.1104

6. Dillen, M., Groom, Q., Chagnoux, S., Güntsch, A., Hardisty, A., Haston, E., Livermore, L., Runnel, V., Schulman, L., Willemse, L., Wu, Z. & Phillips, S. (2019) A benchmark dataset of herbarium specimen images with label data. Biodiversity Data Journal. In press.

7. Federal Agencies Digital Guidelines Initiative (2018) OpenDICE and AutoSFR. From: http://www.digitisationguidelines.gov/guidelines/digitise-OpenDice.html

8. Le Bras, G., Pignal, M., Jeanson, M. et al. The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. Sci Data 4, 170016 (2017). https://doi.org/10.1038/sdata.2017.16

9. Nekola, J. C., Hutchins, B. T., Schofield, A., Najev, B., & Perez, K. E. (2019). Caveat consumptor notitia museo: Let the museum data user beware. Global Ecology and Biogeography, 28(12), 1722-1734. https://doi.org/10.1111/geb.12995

10. Nelson, G., & Ellis, S. (2018). The history and impact of digitisation and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B, 374(1763), https://doi.org/10.1098/rstb.2017.0391

11. Ranatunga D, Milne P, Birch J (2018) Digitise This! Innovation in Digitisation Initiatives within Australasian Herbaria. Biodiversity Information Science and Standards 2 (e26077). https://doi.Org/https://doi.org/10.3897/biss.2.26077

12. Seregin, A. P. (2018). The largest digital herbarium in Russia is now available online! Taxon, 67(2), 465–467. https://www.jstor.org/stable/26824752

13. Seregin, A. P., & Stepanova, N. Y. (2020). MHA Herbarium: Eastern European collections of vascular plants. Biodiversity data journal, 8, e57512. https://doi.org/10.3897/BDJ.8.e57512

14. van Walsum M, Wijers A, Kamminga P, van der Mije S, van Dorp K (2019) State of the art and perspectives on mass imaging of liquid samples. https://doi.org/10.5281/zenodo.3469547

15. von Baeyer, M., & Marston, J. M. (2021). Best practices for digitising a wood slide collection: The Bailey-Wetmore Wood Collection of the Harvard University Herbaria. *Quaternary International*, *593*, 50-59

16. Wetzel, F. T., Bingham, H. C., Groom, Q., Haase, P., Kõljalg, U., Kuhlmann, M., ... & Häuser, C. L. (2018). Unlocking biodiversity data: Prioritisation and filling the gaps in biodiversity observation data in Europe. Biological conservation, 221, 78-85. https://doi.org/10.1016/j.biocon.2017.12.024

17. Whitaker, A. F., & Kimmig, J. (2020). Anthropologically introduced biases in natural history collections, with a case study on the invertebrate paleontology collections from the middle Cambrian Spence Shale Lagerstätte. Palaeontologia Electronica, 23, a58. https://doi.org/10.26879/1106

18. Willemse, L., van Egmond, E., Runnel, V., Saarenmaa, H., Rubio, A., Gödderz, K., & Vermeersch, X. (2019). Future challenges in digitisation of private natural history collections. Biodiversity Information Science and Standards. https://doi.org/10.3897/biss.3.37640

Criteria for digitisation priority (Research)

Again, there is really few papers establishing objective criteria for the prioritisation process. The selected literature mainly focused on the needs of digitisation and the future of natural history collections.

1. Ahl, L. I., & Enghoff, H. (2022). MS1. 3 Corpus of previous studies on prioritisation of digitisation compiled.

2. Albani Rocchetti, G., Armstrong, C. G., Abeli, T., Orsenigo, S., Jasper, C., Joly, S., ... & Vamosi, J. C. (2021). Reversing extinction trends: new uses of (old) herbarium specimens to accelerate conservation action on threatened species. New Phytologist, 230(2), 433-450. https://doi.org/10.1098/rstb.2017.0389

3. Ariño, A. (2018). Putting your finger upon the simplest data. Biodiversity Information Science and Standards.

4. Asase, A., & Schwinger, G. O. (2018). Assessment of biodiversity data holdings and user data needs for Ghana. Biodiversity Informatics, 13.

5. Bakker,H., Willemse, L., van Egmond, E., Casino, A., Gödderz, K., & Vermeersch, X. (2018). D2. 1 Inventory of current criteria for prioritisation of digitisation of collections focussed on scientific and societal needs. https://doi.org/10.5281/zenodo.2579156

6. Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., ... & Guralnick, R. P. (2019). Research applications of primary biodiversity databases in the digital age. PloS one, 14(9), e0215794.

7. Cantrill, D. J. (2018). The Australasian Virtual Herbarium: Tracking data usage and benefits for biological collections. Applications in Plant Sciences, 6(2), e1026. https://doi.org/10.1002/aps3.1026

8. Cobb, N. S., Gall, L. F., Zaspel, J. M., Dowdy, N. J., McCabe, L. M., & Kawahara, A. Y. (2019). Assessment of North American arthropod collections: Prospects and challenges for addressing biodiversity research. PeerJ, 7, 10.7717/peerj.8086

9. Davis, C. C., Kennedy, J. A., & Grassa, C. J. (2021). Back to the future: A refined single-user photostation for massively scaling herbarium digitisation. Taxon, 70(3), 635-643.

10. De Almeida, M., Pinto, Â., & Carvalho, A. (2021). Digitising Primary Data on Biodiversity to Protect Natural History Collections Against Catastrophic Events: The type material of dragonflies (Insecta: Odonata) from Museu Nacional of Brazil. Biodiversity Information Science and Standards, 5, e75284.

11. Gnanasekaran, R. K., & Marciano, R. (2021, October). Piloting Data Science Learning Platforms through the Development of Cloud-based interactive Digital Computational Notebooks. In Proceedings of International Symposium on Grids & Clouds 2021—PoS (ISGC2021) (Vol. 378, p. 018).

12. Griffin, E. (2019). Getting necessary historical data out of deep freeze. Polar Science, 21, 238-239. https://doi.org/10.1093/biosci/biz140

14. Hardisty, A., Saarenmaa, H., Casino, A., Dillen, M., Görddерz, K., Groom, Q., ... & Willemse, L. (2020). Conceptual design blueprint for the DiSSCo digitisation infrastructure-DELIVERABLE D8. 1. Research Ideas and Outcomes, 6.

15. Heberling, J. M. (2022). Herbaria as big data sources of plant traits. International Journal of Plant Sciences, 183(2), 87-118.

16. Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., ... & Davis, C. C. (2020). Digitisation and the future of natural history collections. BioScience, 70(3), 243-251. https://doi.org/10.1093/biosci/biz163

17. Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2019). Biological collections for understanding biodiversity in the Anthropocene. Philosophical Transactions of the Royal Society B, 374(1763), https://doi.org/10.1098/rstb.2017.0386

18. Miller, S. E., Barrow, L. N., Ehlman, S. M., Goodheart, J. A., Greiman, S. E., Lutz, H. L., ... & Light, J. E. (2020). Building natural history collections for the twenty-first century and Piebeyond. BioScience, 70(8), 674-687. https://doi.org/10.1093/biosci/biaa069

19. Mondejar, M. E., Avtar, R., Diaz, H. L. B., Dubey, R. K., Esteban, J., Gómez-Morales, A., ... & Garcia-Segura, S. (2021). Digitalization to achieve sustainable development goals: Steps towards a Smart Green Planet. Science of the Total Environment, 794, 148539.

20. Nic Lughadha, E. M., Graziele Staggemeier, V., Vasconcelos, T. N., Walker, B. E., Canteiro, C., & Lucas, E. J. (2019). Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology, 33(3), 511-522.

21. Norris, C. A. (2017). The future of natural history collections. In The future of natural history museums (pp. 13-28). Routledge.

22. Paton, A., Antonelli, A., Carine, M., Forzza, R. C., Davies, N., Demissew, S., ... & Dickie, J. (2020). Plant and fungal collections: Current status, future perspectives. Plants, People, Planet, 2(5), 499-514. https://doi.org/10.1002/ppp3.10141

23. Pellens, R. (Ed.). (2021). Natural History Collections in the Science of the 21st Century: A Sustainable Resource for Open Science. John Wiley & Sons.

24. Popov D, Roychoudhury P, Hardy H, Livermore L, Norris K (2021) The Value of Digitising Natural History Collections. Research Ideas and Outcomes 7: e78844. https://doi.org/10.3897/rio.7.e78844

25. Powell, C. (2020). In support of natural history digitisation: assistive tools for the mobilization of biodiversity data. Master Thesis of the University of Tennessee at Chattanooga. https://scholar.utc.edu/cgi/viewcontent.cgi?article=1814&context=theses

26. Powell, C., Krakowiak, A., Fuller, R., Rylander, E., Gillespie, E., Krosnick, S., Ruhfel, B., Morris, A. B., & Shaw, J. (2021). Estimating herbarium specimen digitisation rates: Accounting for human experience. Applications in plant sciences, 9(4), e11415. https://doi.org/10.1002/aps3.11415

27. Raes, N., Casino, A., Goodson, H., Islam, S., Koureas, D., K Schiller, E., ... & Robertson, T. (2020). White paper on the alignment and interoperability between the Distributed System of Scientific Collections (DiSSCo) and EU infrastructures-The case of the European Environment Agency (EEA). https://doi.org/10.3897/rio.6.e62361

28. Schindel DE, Cook JA (2018) The next generation of natural history collections. PLoS Biol 16:e2006125. https://doi.org/10.1371/journal.pbio.2006125

29. Semal, P., Adam, M., Van den Spiegel, D., Theeten, F., Engledow, H., Mergen, P., ... & Rubio, A. (2019). CETAF Collection Dashboard: Mapping natural history collections diversity. Biodiversity Information Science and Standards.

30. Shea, E. K., Sierwald, P., Bieler, R., & Rosenberg, G. (2018). Priorities and opportunities for digitising mollusk collections. American Malacological Bulletin, 36(2), 171-176.

31. von Baeyer, M., & Marston, J. M. (2021). Best practices for digitising a wood slide collection: The Bailey-Wetmore Wood Collection of the Harvard University Herbaria. Quaternary International, 593, 50-59.

32. Webster, M. S., Buschbom, J., Hardisty, A., & Bentley, A. (2021). The Digital Extended Specimen will Enable New Science and Applications. Biodiversity Information Science and Standards, (1).

33. Wetzel, F. T., Bingham, H. C., Groom, Q., Haase, P., Kõljalg, U., Kuhlmann, M., ... & Häuser, C. L. (2018). Unlocking biodiversity data: Prioritisation and filling the gaps in biodiversity observation data in Europe. Biological conservation, 221, 78-85. https://doi.org/10.1016/j.biocon.2017.12.024

34. Willemse, L., van Egmond, E., Runnel, V., Saarenmaa, H., Rubio, A., Gödderz, K., & Vermeersch, X. (2019). Future challenges in digitisation of private natural history collections. Biodiversity Information Science and Standards. https://doi.org/10.3897/biss.3.37640

Criteria for digitisation priority (Education)

Not in fact criteria, these papers predominantly focus on the attention of digital collections for the education systems and public awareness based on experiences.

1. Astrin, J. J., & Schubert, H. C. (2017). Community perception of natural history collections–an online survey. Bonn Zoological Bulletin, 66, 61-72.

2. Bakker, F. T., Antonelli, A., Clarke, J. A., Cook, J. A., Edwards, S. V., Ericson, P. G., ... & Källersjö, M. (2020). The Global Museum: natural history collections and the future of evolutionary science and public education. PeerJ, 8, e8225.

3. Ellwood, E. R., Douglas, N., Linton, D., Phillips, M., White, L., & Monfils, A. K. (2018, August). Building standards and resources for using biodiversity data in the classroom through Biodiversity Literacy in Undergraduate Education (BLUE). In 2018 ESA Annual Meeting (August 5--10). ESA.

4. Flemming, A., Phillips, M., Shea, E. K., Bolton, A., Lincoln, C., Green, K., ... & Cubeta, M. A. (2020). Using Digital Natural History Collections in K-12 STEM Education. Journal of Museum Education, 45(4), 450-461.

5. Gerdes, C., Harris, K. M., Beas-Moix, M., & Marsico, T. D. (2017). The transformative power of student-led natural history collections clubs. In Collection Forum (Vol. 31, No. 1-2, pp. 70-83). Soc. for the Pres. of Natural History Collections

6. Hogue, G., Phillips, M., & Cubeta, M. (2018). A Model for Creating Connections and Building Collections-Based Curricula for Pre-College Educators. Biodiversity Information Science and Standards.

7. Linton, D. L., Ellwood, E., White, L. D., Douglas, N. F., & Monfils, A. K. (2022). Experiments in Data Mining: Using Digitised Natural History Collections to Introduce Biology Students to Data Science. Trends in Teaching Experimentation in the Life Sciences, 123-141.

8. Monfils, A. K., Powers, K. E., Marshall, C. J., Martine, C. T., Smith, J. F., & Prather, L. A. (2017). Natural history collections: teaching about biodiversity across time, space, and digital platforms. Southeastern Naturalist, 16(sp10), 47-57.

9. Motz, G. J., Cook, K. J., Zimmerman, A., & Sturgeon, P. R. (2018, January). Building a case for institutional repositories: digitisation as an augmentation of the impact of natural history collections at home and abroad. In Geological Society of America Abstracts with Programs.

10. Seltmann, K., Allen, J., Brown, B. V., Carper, A., Engel, M. S., Franz, N., ... & Tuker, E. (2021). Announcing Big-Bee: An initiative to promote understanding of bees through image and trait digitisation. Biodiversity Information Science and Standards, 5(e74037).

Cost of digitisation

DISSCO program is working on the cost of digitisation which could be an important criterion for prioritisation. Beside strategy to work with limited resources or experiment for other collections than herbarium, it is more from the side of the socio-economic impact that we found in the literature.

1. Hardisty, A., Livermore, L., Walton, S., Woodburn, M., & Hardy, H. (2020). Costbook of the digitisation infrastructure of DiSSCo. Research Ideas and Outcomes, 6, e58915. doi: 10.3897/rio.6.e58915

2. Hereld, M. (2019) LightningBug ONE: An experiment in high- throughput digitisation of pinned insects. https://doi.org/10.3897/biss.3.37228

3. Medina, J. J., Maley, J. M., Sannapareddy, S., Medina, N. N., Gilman, C. M., & McCormack, J. E. (2020). A rapid and cost-effective pipeline for digitisation of museum specimens with 3D photogrammetry. PLoS One, 15(8), e0236417.

4. Santos, J., da Cunha, P. R., & Sales, F. (2020). A strategy to digitise natural history collections with limited resources. Biodiversity Data Journal, 8.

Socio-economic and citizen sciences

If this grouping could easily be associated with the education digitisation priority, we have chosen to place it more closely to the cost of digitisation mainly because of the economy of the crowdsourcing of the citizen sciences.

1. Alexander, R., Ruddock, K., & Summers, M. (2021). A collaboration in creating digital natural history collections: A case study of Alberta native bees.

2. Andersen, D., & Lau, R. (2018) Pay Rates and Subject Performance in Social Science Experiments Using Crowdsourced Online Samples. Journal of Experimental Political Science, 5(3), 217-229. doi:10.1017/XPS.2018.7

3. Best, J. (2018). DemoCamp: BRIT Digitisation Appliance. Biodiversity Information Science and Standards.

4. Cohen, L. R. (2021). Field Notes: An Exploration of Crowdsourcing Platforms for Natural History Collections.

5. Ellwood, E. R., Kimberly, P., Guralnick, R., Flemons, P., Love, K., Ellis, S., ... & Mast, A. R. (2018). Worldwide engagement for digitising biocollections (WeDigBio): The biocollections community's citizen-science space on the calendar. BioScience, 68(2), 112-124. https://doi.org/10.1093/biosci/bix143

6. Ellwood, E. R., Sessa, J. A., Abraham, J. K., Budden, A. E., Douglas, N., Guralnick, R., ... & Monfils, A. K. (2020). Biodiversity science and the twenty-first century workforce. BioScience, 70(2), 119-121. https://doi.org/10.1093/biosci/biz147

7. Gerdes, C., Harris, K. M., Beas-Moix, M., & Marsico, T. D. (2017). The transformative power of student-led natural history collections clubs. In Collection Forum (Vol. 31, No. 1-2, pp. 70-83). Soc. for the Pres. of Natural History Collections

8. James, S.A., P. S. Soltis, L. Belbin, A. D. Chapman, G.Nelson, D. L. Paul, and Collins, M., (2018) Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences 6(2): e1024. doi:10.1002/aps3.1024

9. Lang, P. L., Willems, F. M., Scheepens, J. F., Burbano, H. A., & Bossdorf, O. (2019). Using herbaria to study global environmental change. New Phytologist, 221(1), 110-122. https://doi.org/10.1111/nph.15401

10. Lughadha, E. M., Graziele Staggemeier, V., Vasconcelos, T. N., Walker, B. E., Canteiro, C., & Lucas, E. J. (2019). Harnessing the potential of integrated systematics for conservation of taxonomically complex, megadiverse plant groups. Conservation Biology, 33(3), 511-522. https://doi.org/10.1111/cobi.13289

11. Mankar, Akash & J. Shah, Riddhi & Lease, Matthew. (2017) Design Activism for Minimum Wage Crowd Work. arXiv:1706.10097v3

12. Nelson, G., & Ellis, S. (2018). The history and impact of digitisation and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B, 374(1763), 20170391.. https://doi.org/10.1098/rstb.2017.0391

13. Oswald, E. C. (2020). Getting to know other ways of knowing: Boundary experiences in citizen science. Citizen Science: Theory and Practice, 5(1).

14. Paton, A., Antonelli, A., Carine, M., Forzza, R. C., Davies, N., Demissew, S., ... & Dickie, J. (2020). Plant and fungal collections: Current status, future perspectives. Plants, People, Planet, 2(5), 499-514. https://doi.org/10.1002/ppp3.10141

15. Santaoja, M. (2021). The intertwined past, present, and future of local natural history museum and amateur naturalism in Tampere, Finland. Curator: The museum journal, 64(1), 57-77. https://doi.org/10.1111/cura.12398

16. Schindel, D. E., & Cook, J. A. (2018). The next generation of natural history collections. PLoS Biology, 16(7), e2006125.

17. Sheehan, K.B. (2018) Crowdsourcing research: Data collection with Amazon's Mechanical Turk, Communication Monographs, 85:1, 140-156, DOI: 10.1080/03637751.2017.1342043

18. Soteropoulos, D. L., De Bellis, C. R., & Witsell, T. (2021). Citizen Science Contributions to Address Biodiversity Loss and Conservation Planning in a Rapidly Developing Region. Diversity, 13(6), 255. https://doi.org/10.3390/d13060255

19. Willemse, L., van Egmond, E., Runnel, V., Saarenmaa, H., Rubio, A., Gödderz, K., & Vermeersch, X. (2019). Future challenges in digitisation of private natural history collections. Biodiversity Information Science and Standards.

20. Willis, C.G., Law, E. , Williams, A.C., Franzone, B.F., Bernardos, R. , Bruno, L. , Hopkins, C., Schorn, C., Weber, E., Park, D.S. and Davis, C.C. (2017) CrowdCurio: an online crowdsourcing platform to facilitate climate change studies using herbarium specimens. New Phytol, 215: 479-488. doi:10.1111/nph.14535

Impact of digitisation

Under this characterization of the literature, we grouped research papers that point the interest in the natural history collections data to apprehend the actual environmental crisis, societal changes or the potential to enhance research diversity.

1. Bakker, F. T., Antonelli, A., Clarke, J. A., Cook, J. A., Edwards, S. V., Ericson, P. G., ... & Källersjö, M. (2020). The Global Museum: natural history collections and the future of evolutionary science and public education. PeerJ, 8, e8225.

2. Bartomeus, I., Stavert, J. R., Ward, D., & Aguado, O. (2019). Historical collections as a tool for assessing the global pollination crisis. Philosophical Transactions of the Royal Society B, 374(1763), https://doi.org/10.1098/rstb.2017.0389.

3. Batke, S. P., Dallimore, T., Reyes-Chávez, J., Maradiaga, R. F. D., Somers, E., Jones, I., ... & Reid, G. (2022). The largest digital database of fern and lycopod records from Honduras: spatial, temporal and collector biases. Botanical Journal of the Linnean Society. https://doi.org/10.1093/botlinnean/boac019

4. Butcher, K. R., Power, M. J., Larson, M., Orr, M. P., Velásquez-Franco, S., Hudson, M. A., & Bailey, V. J. (2021). Museum Leadership for Engaging, Equitable Education: The Transformative Potential

of Digitised Collections for Authentic Learning Experiences. Curator: The Museum Journal, 64(2), 383-402.

5. Card, D. C., Shapiro, B., Giribet, G., Moritz, C., & Edwards, S. V. (2021). Museum genomics. Annual Review of Genetics, 55, 633-659.

6. Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J., ... & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitisation. New Phytologist, 217(2), 939-955. https://doi.org/10.1111/nph.14855

7. Davis, C. L., Guralnick, R. P., & Zipkin, E. F. (2022). Challenges and opportunities for using natural history collections to estimate insect population trends. Journal of Animal Ecology.

8. De Almeida, M., Pinto, Â., & Carvalho, A. (2021). Digitising Primary Data on Biodiversity to Protect Natural History Collections Against Catastrophic Events: The type material of dragonflies (Insecta: Odonata) from Museu Nacional of Brazil. Biodiversity Information Science and Standards, 5, e75284.

9. Denney, D. A., & Anderson, J. T. (2020). Natural history collections document biological responses to climate change: A commentary on DeLeo et al. (2019), Effects of two centuries of global environmental variation on phenology and physiology of Arabidopsis thaliana. Global change biology, 26(2), 340-342.

10. Drew, J. A., Moreau, C. S., & Stiassny, M. L. (2017). Digitisation of museum collections holds the potential to enhance researcher diversity. Nature ecology & evolution, 1(12), 1789-1790.

11. Ellwood, E. R., Sessa, J. A., Abraham, J. K., Budden, A. E., Douglas, N., Guralnick, R., ... & Monfils, A. K. (2020). Biodiversity science and the twenty-first century workforce. BioScience, 70(2), 119-121. https://doi.org/10.1093/biosci/biz147

12. Eversole, C. B., Powell, R. L., Lizarro, D. E., Moreno, F., Calderon Vaca, G., Aparicio, J., & Crocker, A. V. (2019). Introduction of a novel natural history collection: a model for global scientific collaboration and enhancement of biodiversity infrastructure with a focus on developing countries. Biodiversity and Conservation, 28(7), 1921-1931.

13. Ewers-Saucedo, C., Allspach, A., Barilaro, C., Bick, A., Brandt, A., Fiege, D., ... & Brandis, D. (2021). Natural history collections recapitulate 200 years of faunal change. Royal Society open science, 8(4), 201983

14. Gropp ER (2018b) What are the rules of international biology? Bioscience 68:731

15. Gropp RE (2018a) Specimens, collections, and tools for future biodiversity-related research. Bioscience 68:3

16. Hilton, E. J., Watkins-Colwell, G. J., & Huber, S. K. (2021). The expanding role of natural history collections. Ichthyology & Herpetology, 109(2), 379-391.

17. Kharouba, H. M., Lewthwaite, J. M., Guralnick, R., Kerr, J. T., & Vellend, M. (2019). Using insect natural history collections to study global change impacts: challenges and opportunities. Philosophical Transactions of the Royal Society B, 374(1763), 20170405.

18. Lendemer, J., Thiers, B., Monfils, A. K., Zaspel, J., Ellwood, E. R., Bentley, A., ... & Aime, M. C. (2020). The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. BioScience, 70(1), 23-30.

19. Linton, D. L., Ellwood, E., White, L. D., Douglas, N. F., & Monfils, A. K. (2022). Experiments in Data Mining: Using Digitised Natural History Collections to Introduce Biology Students to Data Science. Trends in Teaching Experimentation in the Life Sciences, 123-141.

20. Marsico, T. D., Krimmel, E. R., Carter, J. R., Gillespie, E. L., Lowe, P. D., McCauley, R., ... & Monfils, A. K. (2020). Small herbaria contribute unique biogeographic records to county, locality, and temporal scales. American journal of botany, 107(11), 1577-1587.

21. National Academies of Sciences, Engineering, and Medicine. (2021). Biological collections: Ensuring critical research and education for the 21st century. National Academies Press.

22. Nelson, G., & Ellis, S. (2018). The history and impact of digitisation and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B, 374(1763), 20170391. https://doi.org/10.1098/rstb.2017.0391

23. Reddy, M. M., Jennings, L., & Thomas, O. P. (2021). Marine Biodiscovery in a Changing World. Progress in the Chemistry of Organic Natural Products 116, 1-36.

24. Shorthouse, D. P., & Page, R. (2019). Quantifying institutional reach through the human network in natural history collections. Biodiversity Information Science and Standards, (4).

25. Soltis, P. S. (2017). Digitisation of herbaria enables novel research.

26. Watanabe, M. E. (2019). The Evolution of Natural History Collections: New research tools move specimens, data to center stage. BioScience, 69(3), 163-169.

Uses for research

This combination of papers slightly differs from the impact of digitisation by pointing more specifically their uses for research programs and development.

1. Fornari, A. (2018). Employing Natural History Collections in the Aid of Conservation: Streamlining an Approach to Model Species Distributions En Masse for the Preservation of Biodiversity.

2. Gropp, R. E. (2018). Specimens, collections, and tools for future biodiversity-related research. BioScience, 68(1), 3-4.

3. Harmon, A., Littlewood, D. T. J., & Wood, C. L. (2019). Parasites lost: Using natural history collections to track disease change across deep time. Frontiers in Ecology and the Environment, 17(3), 157-166.

4. Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences, 118(6), e2018093118.

5. Hedrick, B. P., Heberling, J. M., Meineke, E. K., Turner, K. G., Grassa, C. J., Park, D. S., ... & Davis, C. C. (2020). Digitisation and the future of natural history collections. BioScience, 70(3), 243-251.

6. Lamichhaney, S., Card, D. C., Grayson, P., Tonini, J. F., Bravo, G. A., Näpflin, K., ... & Edwards, S. V. (2019). Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. Philosophical Transactions of the Royal Society B, 374(1777), 20180248.

7. Lopez, L., Turner, K. G., Bellis, E. S., & Lasky, J. R. (2020). Genomics of natural history collections for understanding evolution in the wild. Molecular ecology resources, 20(5), 1153-1160.

8. Marcer, A., Chapman, A. D., Wieczorek, J. R., Xavier Picó, F., Uribe, F., Waller, J., & Ariño, A. H. (2022). Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. Ecography, e06025.

9. Meineke, E. K., Davies, T. J., Daru, B. H., & Davis, C. C. (2019). Biological collections for understanding biodiversity in the Anthropocene. Philosophical Transactions of the Royal Society B, 374(1763), https://doi.org/10.1098/rstb.2017.0386

10. Nekola, J. C., Hutchins, B. T., Schofield, A., Najev, B., & Perez, K. E. (2019). Caveat consumptor notitia museo: Let the museum data user beware. Global Ecology and Biogeography, 28(12), 1722-1734. https://doi.org/10.1111/geb.12995

11. Nelson, G., Phillips, M., Monfils, A., MacFadden, B., & Hogue, G. (2018). Completing the Data Pipeline: Collections Data Use in Research, Education and Outreach. Biodiversity Information Science and Standards.

12. Phillips, M., Basham, A., Cubeta, M., Harris, K., Hendricks, J., Hogue, G., ... & White, L. (2018). Engaging K-12 Audiences with Biodiversity Data through Advancing Digitisation for Biodiversity Collections. Biodiversity Information Science and Standards.

13. Seltmann, K. C., Cobb, N. S., Gall, L. F., Bartlett, C. R., Basham, M. A., Betancourt, I., ... & Zolnerowich, G. (2017). LepNet: The Lepidoptera of North America Network. Zootaxa, 4247(1), 73-77.

14. Shirey, V. (2018). Visualizing natural history collection data provides insight into collection development and bias. Biodiversity Data Journal, (6).

15. Sikes, Bowser M, Daly K, Høye TT, Meierotto S, Mullen L, Slowik J, Stockbridge J. (2017) The value of museums in the production, sharing and use of entomological data to document hyperdiversity of the changing North. Arctic Science 33, 498– 514. (doi:10.1139/as-2016-0038)

16. Sire, L., Gey, D., Debruyne, R., Noblecourt, T., Soldati, F., Barnouin, T., ... & Rougerie, R. (2019). The challenge of DNA barcoding saproxylic beetles in natural history collections—exploring the potential of parallel multiplex sequencing with Illumina MiSeq. Frontiers in Ecology and Evolution, 7, 495.

17. Van Allen, A. (2017). Bird Skin to Biorepository: Making Materials Matter in the Afterlives of Natural History Collections. Knowledge Organization, 44(7).

18. Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., Gendreau, C., Schigel, D., & Robertson, T. J. (2017). A conceptual framework for quality assessment and management of biodiversity data. PloS one, 12(6), e0178731. https://doi.org/10.1371/journal.pone.0178731

Tools and technnology

Technology is also, at some points an important criterion to prioritise the digitisation process. This compilation beside the technology itself like 3D contains tolls which are of interest in the digitisation and data treatments.

1. Allan, E. L., Livermore, L., Price, B. W., Shchedrina, O., & Smith, V. S. (2019) A Novel Automated Mass Digitisation Workflow for Natural History Microscope Slides. Biodiversity Data Journal 7 (e32342). https://doi.Org/https://doi.org/10.3897/BDJ.7.e32342

2. Allan, E. L., Price, B. W., Shchedrina, O., Dupont, S., Livermore, L., & Smith, V. (2018). A low cost approach to specimen level imaging of natural history microscope slides using a DSLR system.

3. American Museum of Natural History (2018) Fluid Preserved Specimens. https://www.amnh.org/our-research/natural-science-collections-conservation/generalconservation/health-safety/fluid-preserved-specimens.

4. Bertrand, L., Bernard, S., Marone, F., Thoury, M., Reiche, I., Gourrier, A., ... & Bergmann, U. (2017). Emerging approaches in synchrotron studies of materials from cultural and natural history collections. Analytical Chemistry for Cultural Heritage, 1-39.

5. Best, J. (2018). DemoCamp: BRIT Digitisation Appliance. Biodiversity Information Science and Standards.

6. Daru, B. H., Park, D. S., Primack, R. B., Willis, C. G., Barrington, D. S., Whitfeld, T. J., ... & Davis, C. C. (2018). Widespread sampling biases in herbaria revealed from large-scale digitisation. New Phytologist, 217(2), 939-955. https://doi.org/10.1111/nph.14855

7. de la Hidalga, A. N., Rosin, P. L., Sun, X., Livermore, L., Durrant, J., Turner, J., ... & Hardisty, A. (2022). Cross-validation of a semantic segmentation network for natural history collection specimens. Machine Vision and Applications, 33(3), 1-31.

8. EDICO SK (2018) Witikon Photogrammetry System, Retrieved April 17, 2018: http://witikon.eu/

9. Engledow, H., De Smedt, S., Groom, Q., Bogaerts, A., Stoffelen, P., Sosef, M., & Van Wambeke, P. (2018). Managing a Mass Digitisation Project at Meise Botanic Garden: From Start to Finish. Biodiversity Information Science and Standards. doi: 10.3897/biss.2.25912

10. Fetyukova, Y., Tegelberg, R., Karppinen, J., Mononen, T., Wu, Z., & Saarenmaa, H. (2017). INDUSTRIAL METHODS FOR LARGE SCALE DIGITISATION OF BIOLOGICAL COLLECTIONS IN FINLAND. In Международная научно-практическая конференция" Использование современных информационных технологий в ботанических исследованиях" (pp. 136-137).

11. Folk, R. A., Kates, H. R., LaFrance, R., Soltis, D. E., Soltis, P. S., & Guralnick, R. P. (2021). High-throughput methods for efficiently building massive phylogenies from natural history collections. Applications in Plant Sciences, 9(2), e11410.

12. Fraunhoffer IGD (2017) 3D Scanning Technologies, Fraunhofer Institute for Computer Graphics research IGD. Brochure. https://www.cultlab3d.de/wp-content/uploads/2018/05/2017-02-27_CHDBrosch%C3%BCre_EN_web.pdf

13. Garrouste, R. (2021). Digital Photography In Natura in Zoology: More Biology in Natural History Collections?. Natural History Collections in the Science of the 21st Century: A Sustainable Resource for Open Science, 271-287

14. Glöckler, F., Macklin, J., Shorthouse, D., Bölling, C., Bilkhu, S., & Gendreau, C. (2020). DINA—Development of open source and open services for natural history collections & research. Biodiversity Information Science and Standards, 4, e59070.

15. Greeff, M., Caspers, M., Kalkman, V., Willemse, L., Sunderland, B. D., Bánki, O., & Hogeweg, L. (2022). Sharing taxonomic expertise between natural history collections using image recognition. Research Ideas and Outcomes, 8, e79187.

16. Green, D. W., Watson, J. A., Jung, H. S., & Watson, G. S. (2019). Natural history collections as inspiration for technology. BioEssays, 41(2), 1700238.

17. Hardisty, A., Brack, P., Goble, C., Livermore, L., Scott, B., Groom, Q., ... & Soiland-Reyes, S. (2022). The specimen data refinery: a canonical workflow framework and FAIR digital object approach to speeding up digital mobilisation of natural history collections. Data Intelligence, 4(2), 320-341.

18. Hereld, M., & Ferrier, N. (2019). LightningBug ONE: An experiment in high-throughput digitisation of pinned insects. Biodiversity Information Science and Standards, 3, e37228.

19. ICEDIG Project (2019). MS 17 Interim report on rapid 3D capture methods in biological collections and related fields. Work in progress, unpublished (Due 2019/09/30).

20. Jackowiak, B., Lawenda, M., Nowak, M. M., Wolniewicz, P., Błoszyk, J., Urbaniak, M., ... & Meyer, N. (2022). Open Access to the Digital Biodiversity Database: A Comprehensive Functional Model of the Natural History Collections. Diversity, 14(8), 596.

21. JSTOR (2018). JSTOR Global Plants: Guidelines for Scanning Specimens. From: https://guides.jstor.org/ld.php?content_id=31764146

22. Kaiser, C., Kaiser, H., Rickerl, K. J., & O'Shea, M. (2018). A portable, low-cost approach for photographing fluid preserved snake specimens-Recommendations with comments on optimizing specimen photography in Natural History collections.

23. Khidas, K., & Tessier, S. (2018). Building Next-Generation Collections: Natural History Specimens, Just One Click Away!. Biodiversity Information Science and Standards.

24. Levesque-Beaudin, V., Miller, M., Dikow, T., Miller, S., Prosser, S., Zakharov, E., ... & deWaard, J. (2022). A workflow for expanding DNA barcode reference libraries through 'museum harvesting'of natural history collections. ARPHA Preprints, 3, e84304.

25. Levesque-Beaudin, V., Rosati, M. E., Silverson, N., Warne, C. P., Brown, A., Telfer, A. C., ... & Dewaard, J. R. (2017). Museum harvesting in major natural history collections. Genome.

26. Library of Congress (2018). Technical Standards for Digital Conversion of Text and Graphic Materials From: https://memory.loc.gov/ammem/about/techStandards.pdf

27. Little, H., Paul, D., & Strotman, J. (2017). SPNHC 2017 Natural History Collections Biodiversity Informatics 101 Short Course Insights. Biodiversity Information Science and Standards.

28. Livermore, L., et.al. (2017) Digitising Louse Slides. NERC / Natural History Museum pilot project. http://www.nhm.ac.uk/our-science/our-work/digital-museum/digital-collectionsprogramme/digitising-slide-collections.html

29. Mailhot, J. (2020). Seeing the story: Making data visualization accessible to natural history collections through the creation of custom-made dashboards and resources (Doctoral dissertation, University of Colorado at Boulder).

30. Marcer, A., Haston, E., Groom, Q., Ariño, A. H., Chapman, A. D., Bakken, T., ... & Wieczorek, J. R. (2021). Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. Diversity and Distributions, 27(3), 564-567.

31. Medina, J. J., Maley, J. M., Sannapareddy, S., Medina, N. N., Gilman, C. M., & McCormack, J. E. (2020). A rapid and cost-effective pipeline for digitisation of museum specimens with 3D photogrammetry. PLoS One, 15(8), e0236417.

32. Meeus, S., Augus, T., Reyserhove, L., Trekels, M., & Groom, Q. (2021). Implementation of the IIIF for Natural History Collections.

33. Mendez, P. K., Lee, S., & Venter, C. E. (2018). Imaging natural history museum collections from the bottom up: 3D print technology facilitates imaging of fluid-stored arthropods with flatbed scanners. ZooKeys, (795), 49.

34. Miller, S. E., Barrow, L. N., Ehlman, S. M., Goodheart, J. A., Greiman, S. E., Lutz, H. L., ... & Light, J. E. (2020). Building natural history collections for the twenty-first century and beyond. BioScience, 70(8), 674-687. https://doi.org/10.1093/biosci/biaa069

35. Motz, G., Zimmerman, A., Cook, K., & Bancroft, A. (2018). Collections Management and High-Throughput Digitisation using Distributed Cyberinfrastructure Resources. Biodiversity Information Science and Standards.

36. Nguyen, C. et. al. (2017) Towards high-throughput 3D insect capture for species discovery and diagnostics. eScience 2017, BigDig 24 Oct - High Throughput Digitisation for Natural History Collections from: https://press3.mcs.anl.gov//bigdig/files/2017/09/Nguyen-Shared-BigDig2017-Towards-high-throughput-3D-insect-capture.pdf

37. Nieva de la Hidalga A, Owen D, Spasic I, Rosin P, Sun X (2019) Use of Semantic Segmentation for Increasing the Throughput of Digitisation Workflows for Natural History Collections. Biodiversity Information Science and Standards 3: e37161. https://doi.org/10.3897/biss.3.37161

38. Novikov, A., & Sup-Novikova, M. (2021). Simple and cheap photosystem for herbarium digitisation. Plant Introduction, (91/92), 50-53.

39. Nowak, M. M., Słupecka, K., & Jackowiak, B. (2021). Geotagging of natural history collections for reuse in environmental research. Ecological Indicators, 131, 108131.

40. Powell, C., Motley, J., Qin, H., & Shaw, J. (2019). A born-digital field-to-database solution for collections-based research using collNotes and collBook. Applications in Plant Sciences, 7(8), e11284.

41. Price, B. et.al. (2018) ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation. Preprint. Open Science Framework. doi:10.17605/OSF.IO/UVWRN

42. Rinaldo, C., Castronovo, D., DeVeer, J., & Rielinger, D. (2021, February). Supporting Natural History Collections by Connecting Collections. In Proceedings of the International Conference Collect and Connect. CEUR.

43. Sabaj, M. H. (2020). Codes for natural history collections in ichthyology and herpetology. Copeia, 108(3), 593-669.

44. Sadowski, E. M., Schmidt, A. R., Seyfullah, L. J., Solórzano-Kraemer, M. M., Neumann, C., Perrichot, V., ... & Nascimbene, P. C. (2021). Conservation, preparation and imaging of diverse ambers and their inclusions. Earth-Science Reviews, 220, 103653. https://doi.org/10.1016/j.earscirev.2021.103653

45. Saeedi, H. (2021). Marine biodiversity data digitisation and sharing. Beiträge der Jungen Akademie| Mainz 2021.

46. Phillips S., Dillen M., Groom, Q. et al. (2019) Report on New Methods for Data Quality Assurance. Verification and Enrichment Zenodo, vol. 10.

47. Schermer, M., & Duin, D. (2018). Leveraging the Benefits of Open Data Services for Natural History Collection Management. Biodiversity Information Science and Standards.

48. Schuettpelz, E., Frandsen, P. B., Dikow, R. B., Brown, A., Orli, S., Peters, M., ... & Dorr, L. J. (2017). Applications of deep convolutional neural networks to digitised natural history collections. Biodiversity data journal, (5).

49. Shelburne, E. (2018). Producing High-Quality 3-D Models for Specimen Research with the Creaform 3-D Scanner at the Sternberg Museum of Natural History. Biodiversity Information Science and Standards.

50. Stork, L., Weber, A., Miracle, E. G., Verbeek, F., Plaat, A., van den Herik, J., & Wolstencroft, K. (2019). Semantic annotation of natural history collections. Journal of Web Semantics, 59, 100462.

51. Ströbel, B., Schmelzle, S., Blüthgen, N., & Heethoff, M. (2018). An automated device for the digitisation and 3D modelling of insects, combining extended-depth-of-field and all-side multi-view imaging. ZooKeys, (759), 1.

52. Tegelberg, R., Kahanpää, J., Karppinen, J., Mononen, T., Wu, Z., & Saarenmaa, H. (2017, October). Mass digitisation of individual pinned insects using conveyor-driven imaging. In 2017 IEEE 13th International Conference on e-Science (e-Science) (pp. 523-527). IEEE.

53. Thorpe, E. D. (2019). Bringing microfossil specimens into the light: Using semi-automated digitisation techniques to improve collection accessibility. Making the Case for Natural History Collections: SPNHC 2019.

54. van Egmond E, Willemse L, Paul D, Woodburn M, Casino A, Vermeersch X, Gôdderz K, Bloothoofd J, Wijers A, Raes N (2019) Design of a Collection Digitisation Dashboard. Zenodo https://doi.Org/https://doi.org/10.5281/zenodo.2621055

55. Veiga, A. K., Saraiva, A. M., Chapman, A. D., Morris, P. J., Gendreau, C., Schigel, D., & Robertson, T. J. (2017). A conceptual framework for quality assessment and management of biodiversity data. PLoS ONE 12(6): e0178731. https://doi.org/10.1371/journal.pone.0178731

56. Willemse, L., & Caspers, M. (2018). Inter-Institutional Collections Storehouse. Biodiversity Information Science and Standards.

57. Wilson, R. J., de Siqueira, A. F., Brooks, S. J., Price, B. W., Simon, L. M., van der Walt, S. J., & Fenberg, P. B. (2022). Applying computer vision to digitised natural history collections for climate change research: Temperature-size responses in British butterflies. Methods in Ecology and Evolution.

58. Wu, Z., Kahanpää, J., Sihvonen, P., Koivunen, A., & Saarenmaa, H. (2019). Automated Methods in Digitisation of Pinned Insects. Biodiversity Information Science and Standards, 3, e38260.

59. Zizka, A. , Silvestro, D. , Andermann, T. , Azevedo, J. , Duarte Ritter, C. , Edler, D. , Farooq, H. , Herdean, A. , Ariza, M. , Scharn, R. , Svanteson, S. , Wengström, N. , Zizka, V. and Antonelli, A. (2019), CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. Methods Ecol Evol. Accepted Author Manuscript. doi:10.1111/2041-210X.13152

Appendix 4. Survey 1, M1.3 – DPP milestone report

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

Authors:

Louise Isager Ahl (NHMD) ORCHID: 0000-0002-4727-1882

Henrik Enghoff (NHMD) ORCHID: 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.

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.

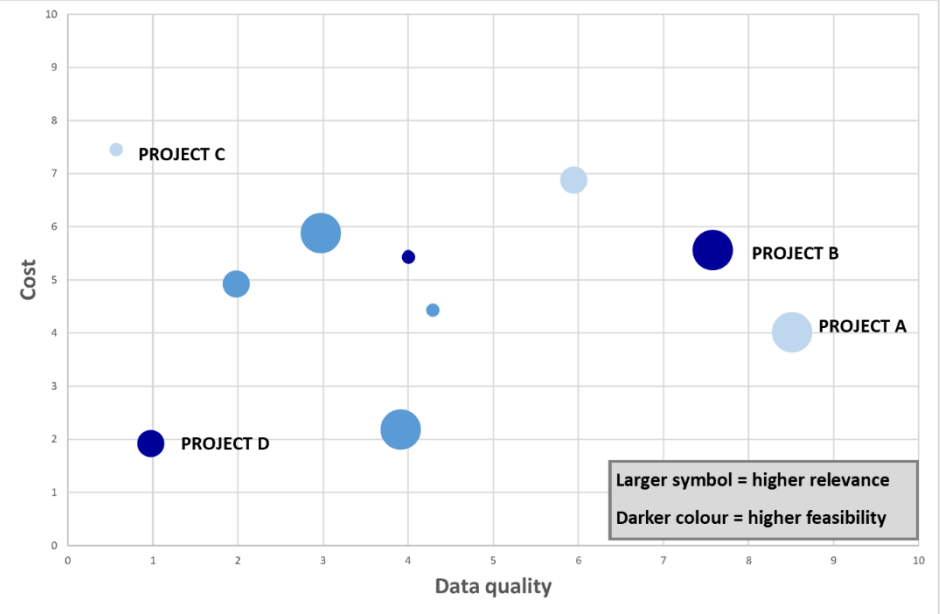


Fig. 1. Four main groups of criteria to be considered when prioritizing digitization projects. In the diagram, **data quality** and **cost** are represented on the horizontal and vertical axes (axis values are arbitrary), whereas **relevance** is represented by the size of the symbols, and **feasibility** 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 in M1.3)

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 in M1.3 – Compilation of Prioritisation of Digitisation answers)

Table of Contents

Introduction **2**

Compilation of information from DiSSCo partner institutions **3**

1. Digitisation strategy of the partner institutions (if available, please provide a copy or link). 4
2. Prioritisation criteria employed for digitisation which has been done or is in progress. 7

Highlighting relevant topics **10**

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.)? 10
2. Are you monitoring it? How? 12
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)? 13
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)? 15
5. Do you have a procedure for validating data (e.g., accuracy of identification and georeferenced)? 17
6. What are you planning to digitise next and what projects are planned for further down the line and why? 18
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.)? 20

Additional notes and comments **22**

RMCA, Belgium 22
Meise, Belgium 23
Estonian National Node, Estonia 25
LUOMUS, Finland 26
MNHN, France 27
MfN, Berlin, Germany 27
Florence, Italy 29
MnhnL, Luxemburg 30
NHMR, Netherlands 31
Naturalis, Netherlands 32
MACB, Spain 37
MAFH, Spain 38
MNCN, Spain 38
Herbarium GB, Sweden 39
NHM, London, England 40
RBGE, Scotland 43

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.

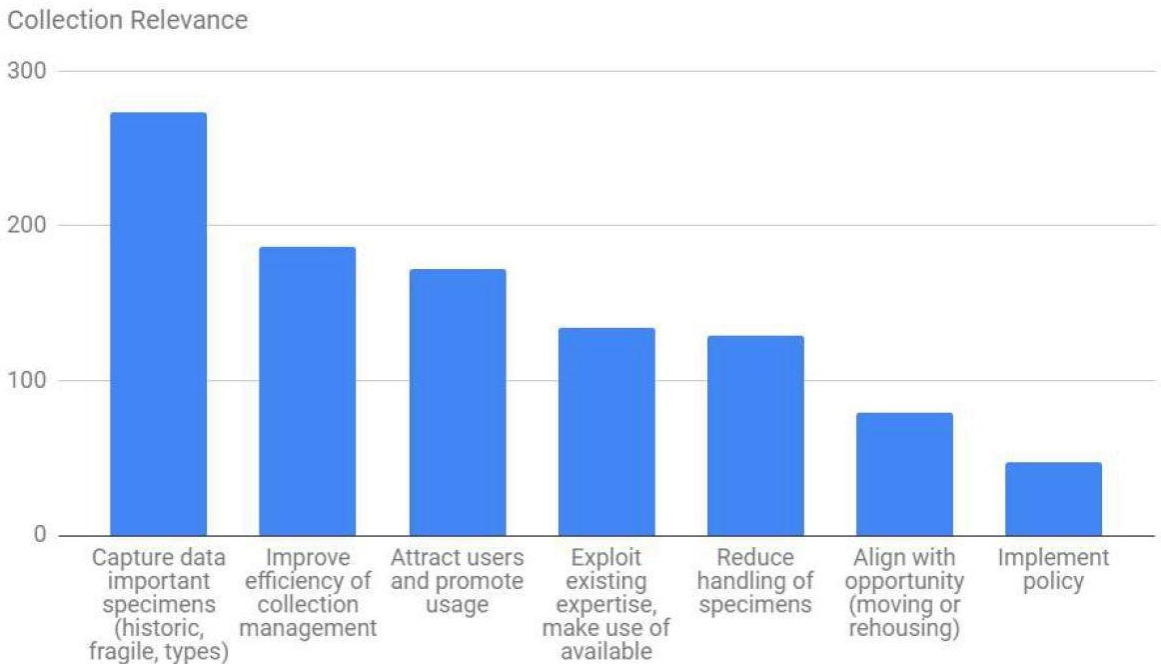
- Austria
- Belgium
    - Royal Belgian Institute of Natural Sciences (BE-RBINS)
    - Meise Botanic Garden*
    - Royal Museum of Central Africa (RMCA)
- Bulgaria
- Czech Republic
- Denmark
    - Natural History Museum of Denmark (NHMD)*
- Estonia
    - Estonian Museum of Natural History (EMNH)
    - 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.

<u>Estonia</u>

EMNH:

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

<u>France</u>

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.

Norway

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])

Spain

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.

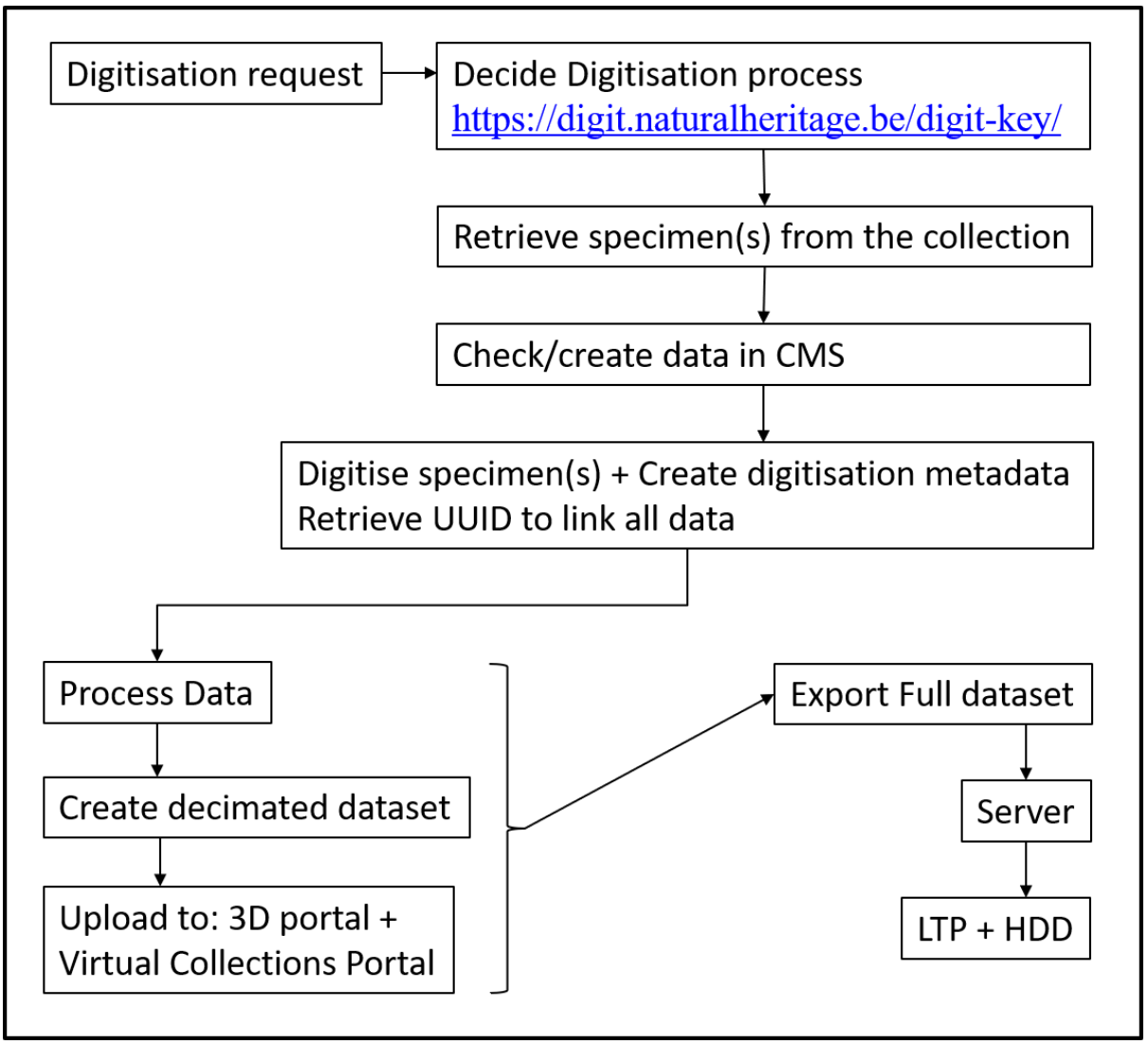


Figure 1: Digitisation Pipeline at the RMCA.

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.

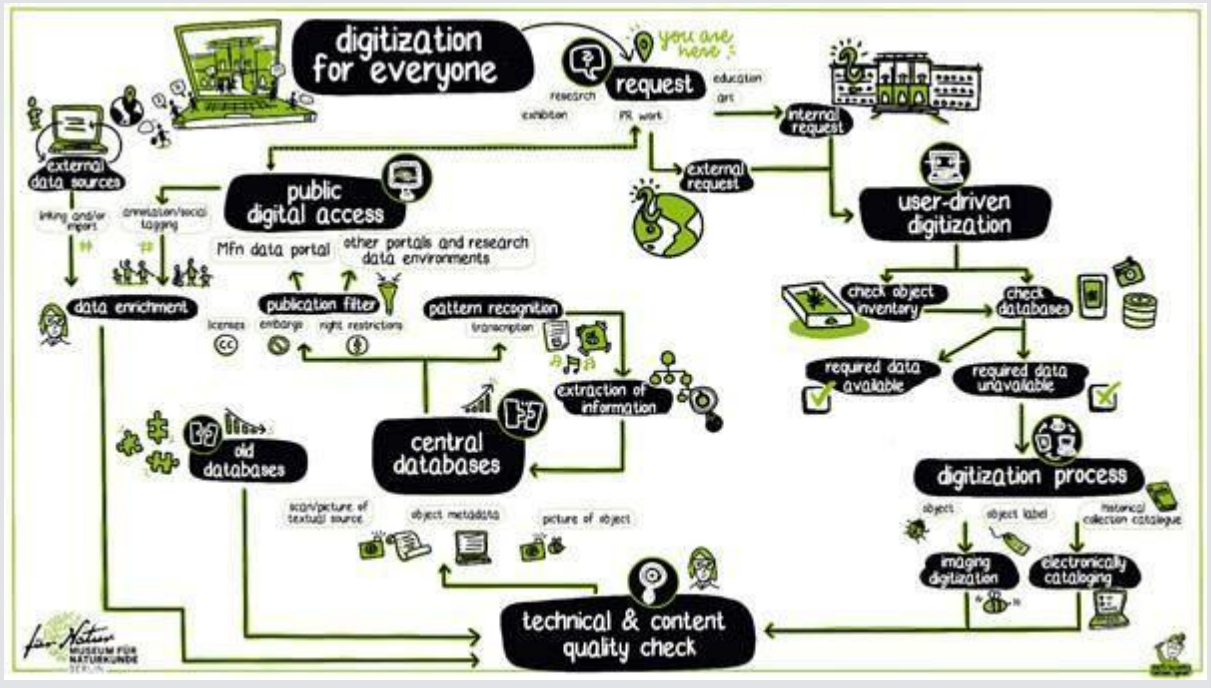
Germany

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, Scotland

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

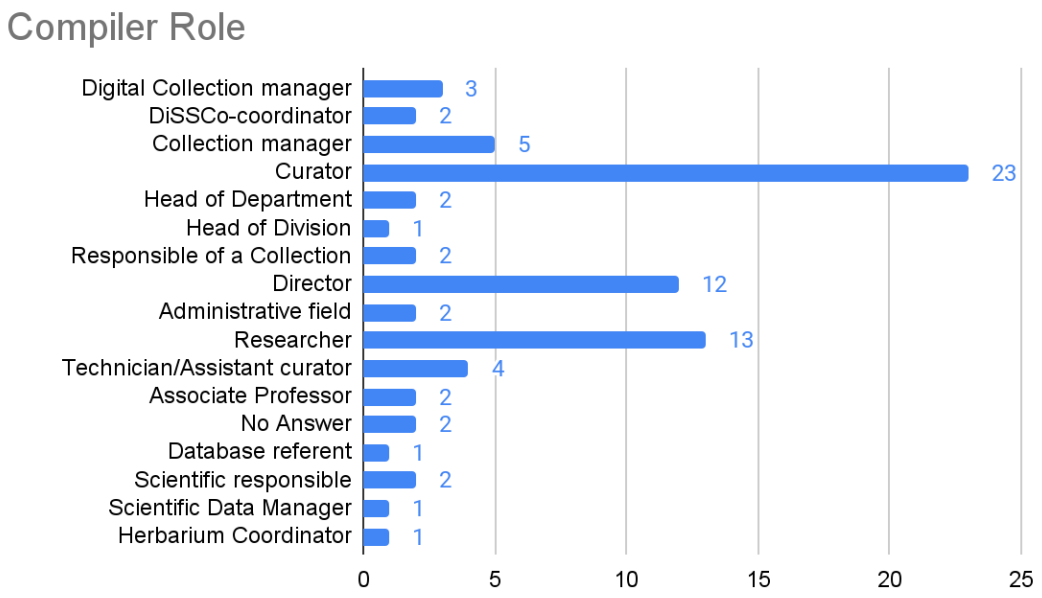
	Project management training Data management / data science training
Internal stakeholders	*Input required from other RBGE departments (agreed or not)* Horticultural staff input required: Maintain alignment between Herbarium and Living Collections Preparation of collection data for voucher specimens from 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

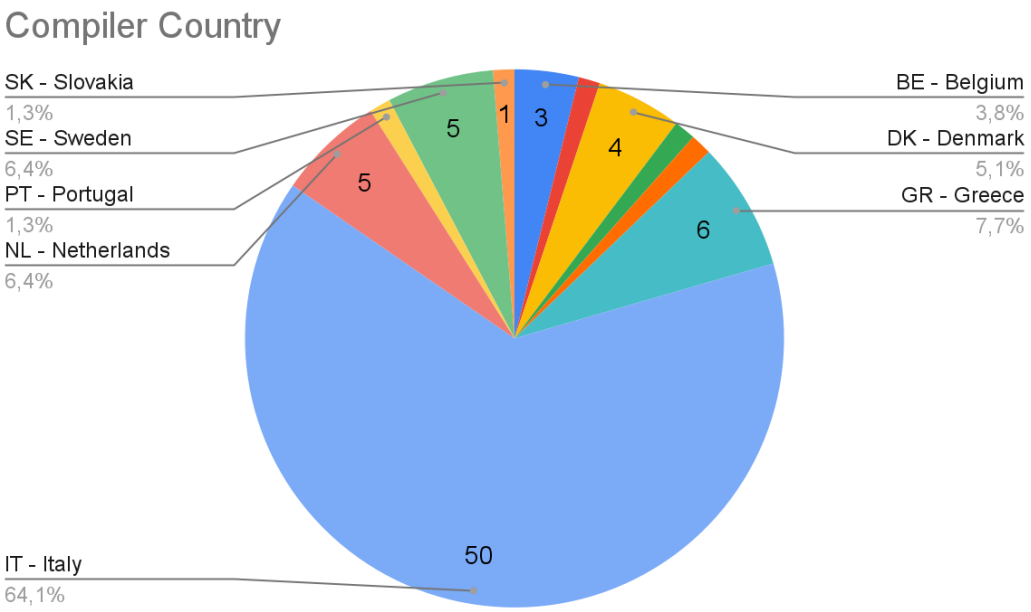
Appendix 5. Survey 2, Multiple-choice questionnaire

Task 1.3 Short questionnaire on digitisation activity

Question 1c) Role
77 answers. Most of all are curators, several are represented by researchers and directors of the collections. A small number has a specific role in digitisation of collections (Digital Collection Manager).

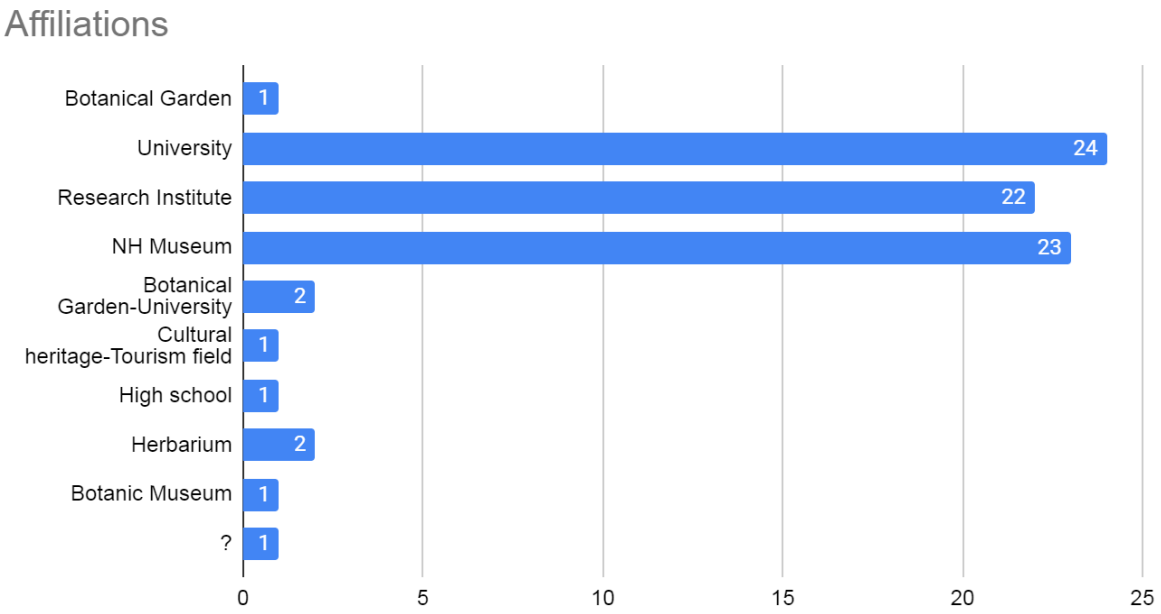


Question 2) select your country
79 Answers. Only 10 of the 23 countries partners of DiSSCo, most of all from Italy.



Question 3) Institution

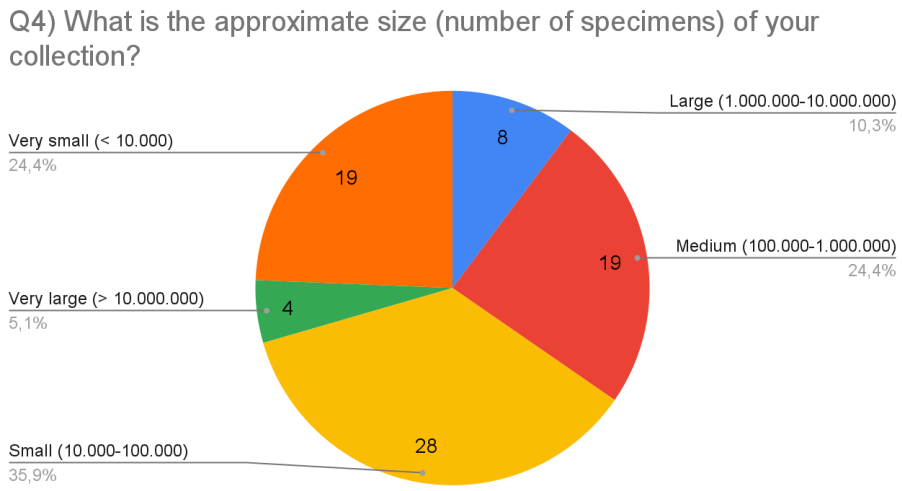
79 answers (one blank). Most of all coming from NH Museums or University Museums and Research Institutes.

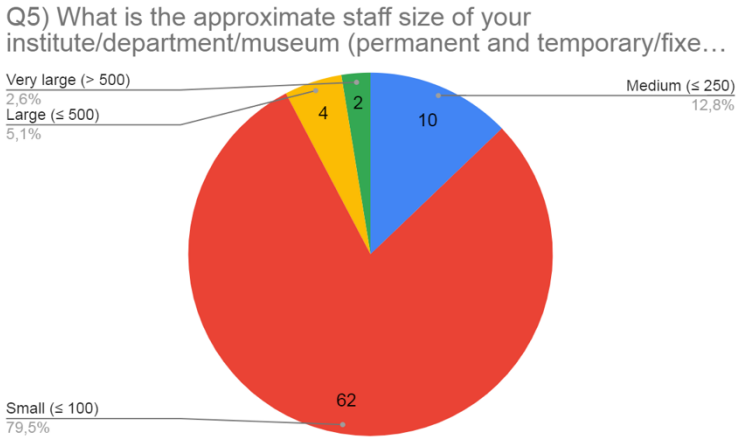


Q4-5: Size of Collections and size of the staff

79 answers. More than 50% is represented by Very small-Small size collections followed by Medium-size with 23%, and only 14% of large-very large size collections. In general a small number of persons is employed for Very small, small and medium size collections; with some

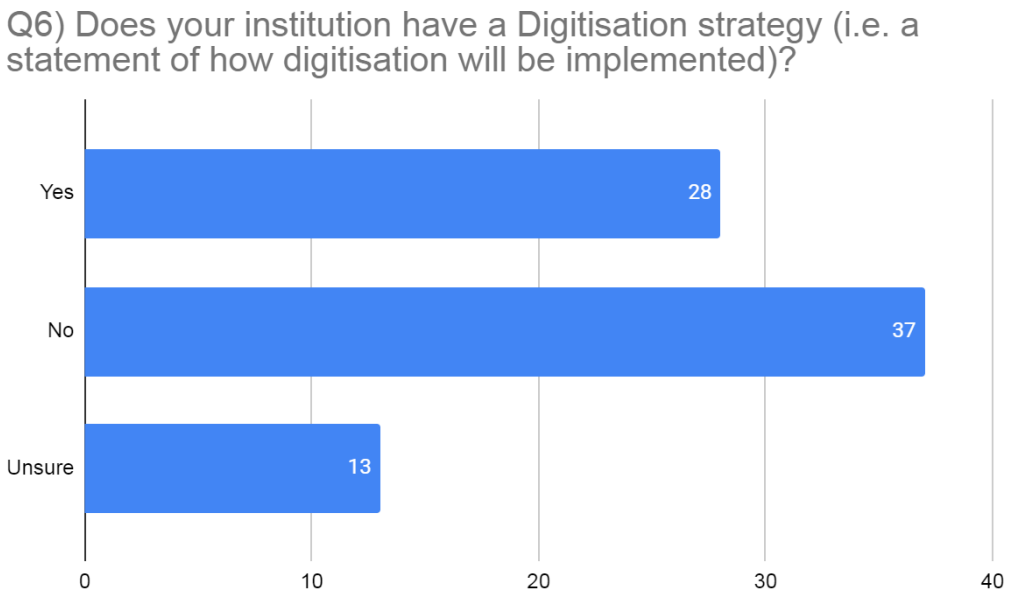
exception: 5 of them, with Large-very Large collections have small size staff, 6 with very small collections have Medium size staff, only one has a very small collection but with Large size staff.





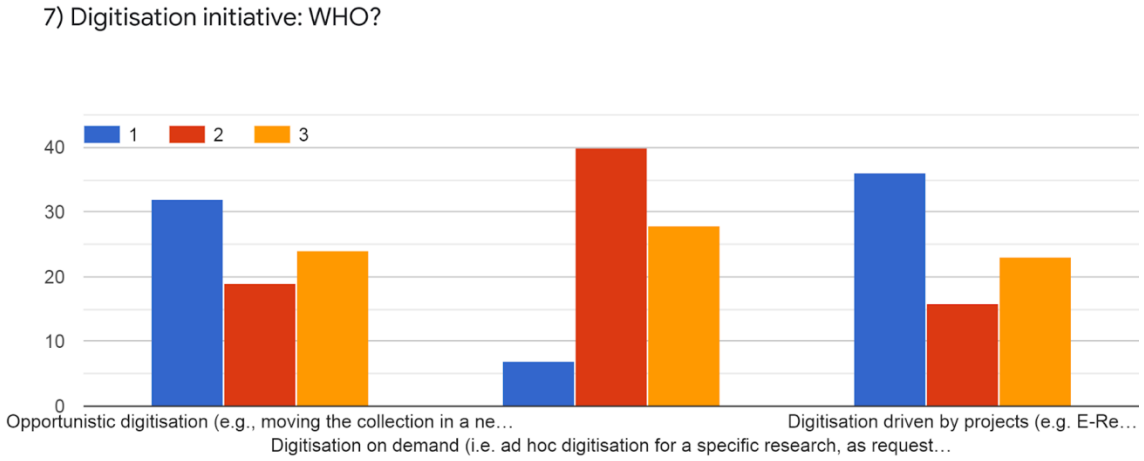
Q6) Digitisation strategy

79 answers. Around 37% of the institutions do not have any digitisation strategy.

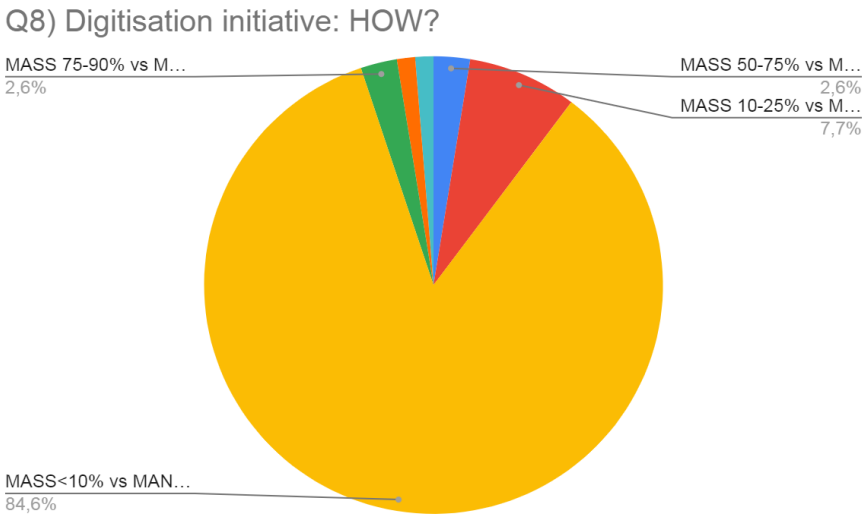


Q7 and 8) Who and How Digitisation

The digitisation seems to be driven by Projects and Opportunistic digitisation (e.g., moving the collection in a new site, out-going loans, new specimens entering the collection, exhibition and other contingent events) that showed similar values. Digitisation on demand (i.e., *ad hoc* digitisation for a specific research, as requested by external researchers; e.g., VA Synthesys+) is the second choice. In any case, the DIGITISATION MAINLY BY MANUAL data entry is the most executed.



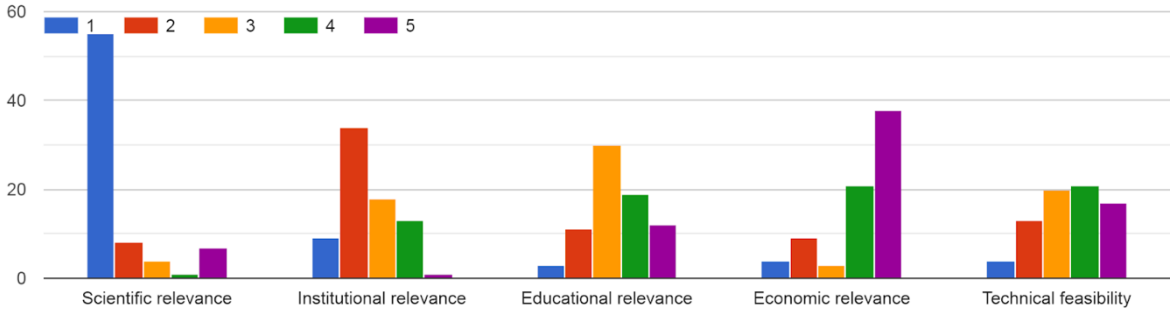
Q8 HOW



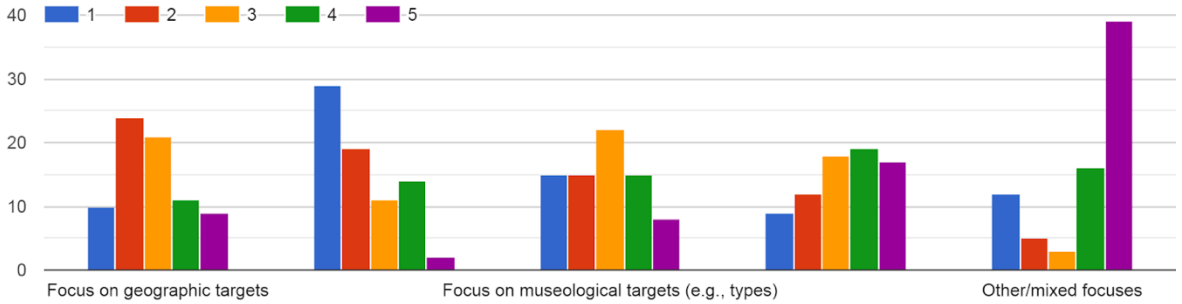
Q9a-b-c-d-e) Digitisation Priorities:

1. **Scientific relevance** is the most important criteria (within this, in order of priority: the most important is "focusing on taxonomic targets", followed by " geographic targets", "museological targets" and "global challenges activities") **followed by**
2. **Institutional relevance** (within this there is a strong balance between both choices "Importance for the museum itself" and "Strategic for national and/or regional programs/projects/guidelines") **and then**
3. **Educational relevance** (within this the most important is "Education and training young people", followed by "Citizen science initiatives" and "Other public engagement")
4. **Technical feasibility** has a balance of preference between third and fourth place (within this, in order of priority: **"**Ease in specimens handling", "Remote digitisation (e.g., from paper catalogues)", "Availability of dedicated technologies (e.g., conveyor belt for herbaria and pinned insects)")
5. At last place there **Economic Relevance** (in order of priority:"Overall performance in respect to human resources and tools", "Overall performance in respect to financial resources", "Faster digitisation improving cost/volume rate")

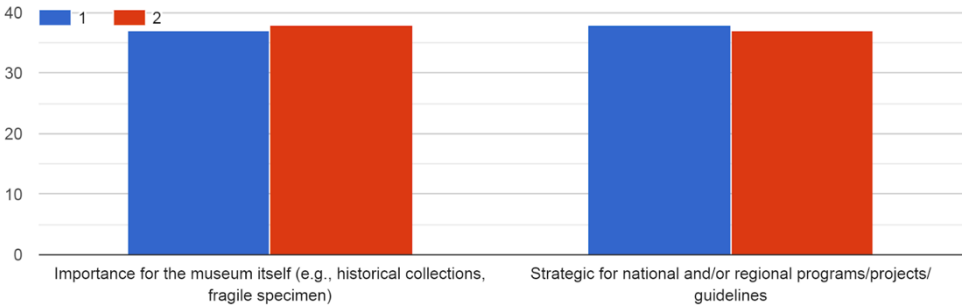
9) Digitisation initiative: WHY?



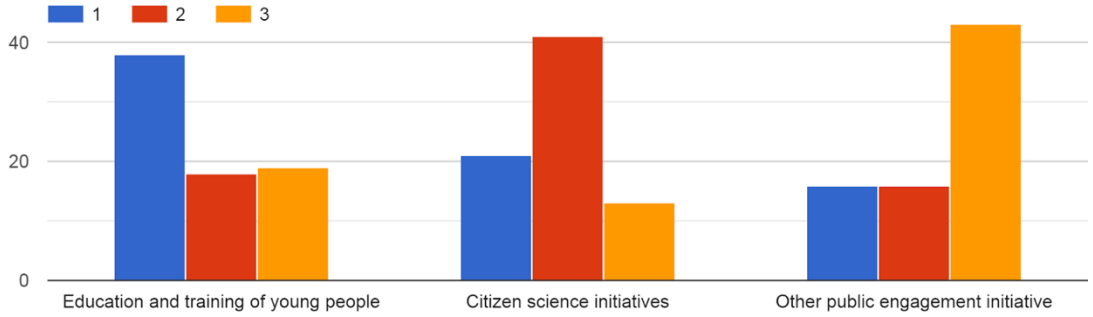
9a) Which are your SCIENTIFIC RELEVANCE priorities when approaching digitisation?



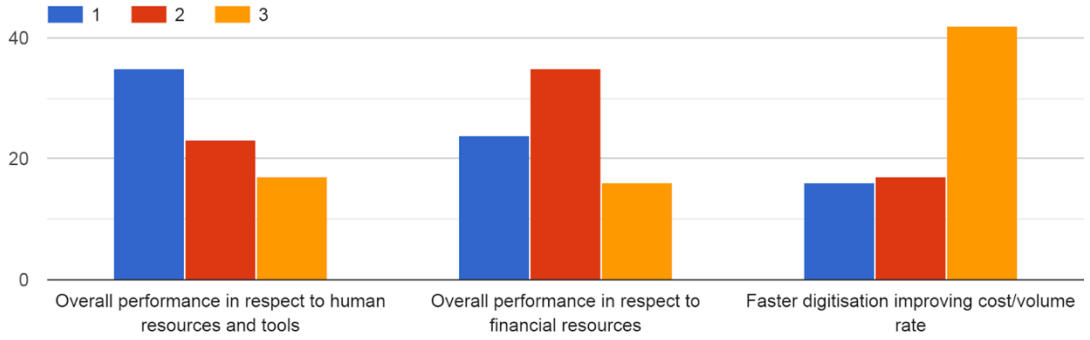
9b) Which are your INSTITUTIONAL RELEVANCE priorities when approaching digitisation?



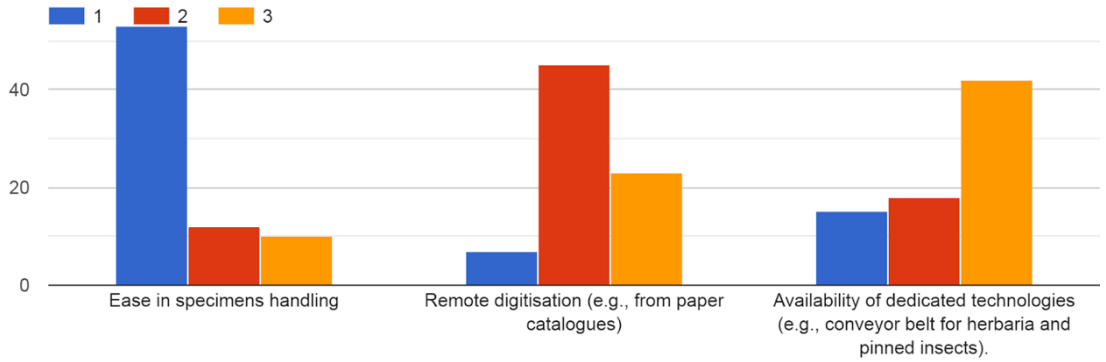
9c) Which are your EDUCATIONAL RELEVANCE priorities when approaching digitisation?



9d) Which are your ECONOMIC RELEVANCE priorities when approaching digitisation?



9e) Which are your TECHNICAL FEASIBILITY RELEVANCE priorities when approaching digitisation?

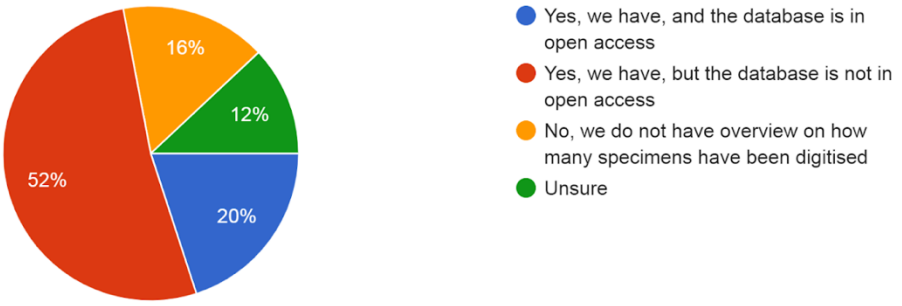


Q.10-11-12) Digitisation overview and CMS: 70% of compilers declared that there is a clear overview of the digitisation status (how many specimens are in the database, how many imaged, open access database, etc) but for most of all the database it is not in open access.

Only 18% has an automated way to monitor the digitisation status and 50% do not have a CMS, they use traditional databases.

10) Do you have a clear overview of the digitisation status of your institution (how many specimens databased, how many imaged, open access database, etc.)?

75 risposte

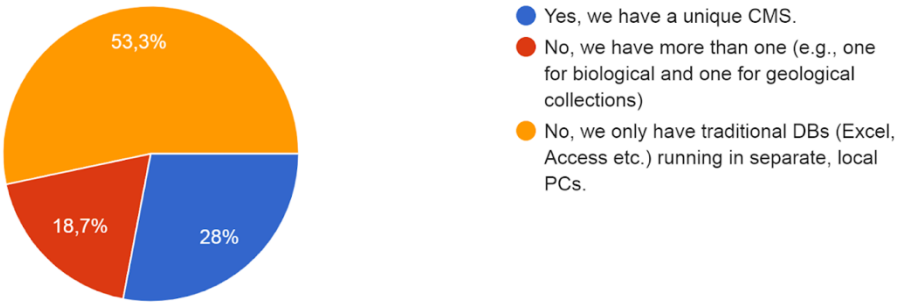


11) How are you monitoring the digitisation activity?

75 risposte

12) Do you have a unique Collection Management System or more than one?

75 risposte

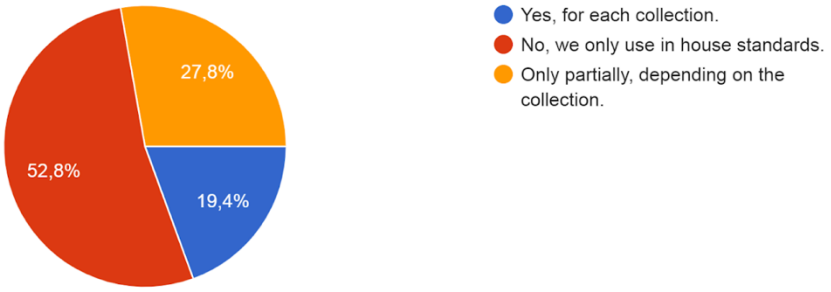


Q13-14: Data Validation and Databasing standards adopted:

For 70% of compilers, data are validated totally (21.6%) or only partially (50%) by the curator and/or by other specialists. It is interesting that **23%** declared they do not have a validation procedure in place. In house standards are the most applied.

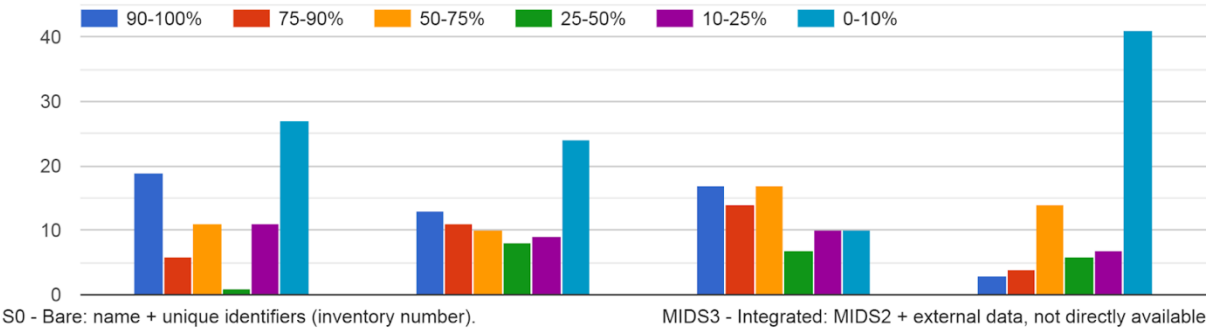
13) Do you have a procedure for validating data (e.g., accuracy of identification and georeferencing)?
74 risposte

14) Do you adopt international databasing standards (such as DarwinCore and ABCD) for specimens?
72 risposte



Q.15-16) Digitisation level:probably higher percentage in MIDS0 - Bare: name + unique identifiers (inventory number), reaching a good percentage in MIDS2 - Complete: MIDS1 + label information (collection locality, collector, date). In general, there is a low percentage of imaged items and 3D models.

15) How many digitised items are databased (textual data) reaching one of the following MIDS (Minimum Information about Digital Specimen)?



16) How many digitised items are imaged or 3D-modelised?