ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1

Hannco (P.A.J.) Bakker (Naturalis)
Luc Willemse (Naturalis)
Emily van Egmond (Naturalis)
Ana Casino (CETAF)
Karsten Gödderz (CETAF)
Xavier Vermeersch (CETAF)

ICEDIG.EU

Summary

Anno 2018, the task of mobilizing data from natural history collections ahead of us is still enormous as 90% of these collections still needs to be mobilized. It is imperative for stakeholders, individual keepers of natural science collections, the community at large, and even for funding agencies, not only to tackle this backlog as quickly as possible, but to do it in the best possible order. To establish the mechanisms to prioritize the digitization of natural history collections, a demand driven framework is required based on, among others, criteria used to digitize collections. In this study, information available from previous surveys and studies was brought together into a comprehensive list of criteria used for prioritizing digitization of natural history collections. Criteria in the list were uniquified, where possible combined, categorized and finally included in a survey aimed at verifying and eliciting further comments. Results of the survey indicated that regardless of the role of respondents, size of the collection or kind of institute, scientific criteria are the most highly rated criteria used for prioritizing digitization of natural history collections. This is especially true for criteria linked to stimulating both fundamental research and research focused on understanding biodiversity processes and trends, and enhancing access to primary biodiversity data. Three methods are discussed that can be used to reach decisions regarding practical aspects of digitization (is it feasible?) and the order in which (parts of) the collection is digitized (where do we start?). For the former a decision tree based on a multivariate key is proposed, while for the latter a scoring method or panel review may be more suitable. If economic, scientific and social relevance are to be taken into account, a panel consisting of biodiversity scientists from various fields would offer the best guarantee for a well-informed decision-making process in relation to prioritization of digitization of natural history collections. It is important to set a clear international digital (research) agenda which can serve as a guideline within DiSSCo to determine what to prioritize in terms of digitization of collections specimens in more detail.

ICEDIG.EU

Contents

ICEDIG.EU

1. Background

Collections of natural history institutions worldwide reflect both the unique biodiversity of past and present life and the geological history of the Earth (Kemp 2015). Globally, it is estimated that natural history collections contain between 1.2 and 2.1 billion specimens (Ariňo 2010). European natural history collections are expected to contain up to one billion specimens (N. Raes, pers. comm.). However, only less than 20% of these collections has been digitized so far at the specimen level (internal DiSSCo Survey, unpublished data). The majority of natural history collections therefore is not accessible and cannot be used by the wider community to its fullest potential. Natural history collections are first and foremost essential for describing new species and studying biodiversity distribution patterns, for which digitization can make all the difference in terms of discoverability and accessibility (Kemp 2015). In addition, natural history collections are increasingly being used for other types of research, such as studying evolution (Holmes et al. 2016) and climate change or for a variety of educational, institutional, social and economic purposes (Baird 2010). By digitizing of biodiversity collections, this information can be made available on a large scale, thereby promoting accessibility for and usage by a wider community and a wider set of goals. As the backlog of specimens already present in collections which still need to be digitized is enormous and new specimens are added every year, digitization is an enormous task. Despite several efforts at mass digitization of an entire collection (e.g. Blagoderov et al. 2012, Heerlien et al. 2015), many projects thus far mainly had a local to national scope or focused on specific groups of specimens (e.g. FES[1], E-ReColNat[2]).

With the introduction of the new Research Infrastructure DiSSCo (Distributed System of Scientific Collections) into the environmental research infrastructure landscape, limitations will be addressed and the digitization process accelerated due to being a joint effort of the entire natural sciences community. DiSSCo aims to become a unique access point to collection specimens in their digital form, enabling science, opening up the collection data for the public by overcoming the current isolated access model, and by prioritizing specific groups of specimens in a new manner. DiSSCo means a completely new way to approach collections and their digitization, where all European collections will essentially be considered as one. Even more than before, this new vision will make it necessary to prioritize what needs to be digitized (e.g. Vollmar et al. 2010, Berents et al. 2010, Berendsohn and Seltmann 2010) as it will take many years before all specimens will have been processed. To this end, it is essential to set criteria for the prioritization of digitization.

In the recent past, criteria for the prioritization of digitization have been inventoried several times, for example by the Global Biodiversity Information Facility (GBIF; Frazier et al. 2008 and Krishtalka et al. 2016), Naturalis Biodiversity Center (Heerlien et al. 2015) and the Atlas of Living Australia (Kalms 2012). Based, among others, on these past inventories, criteria used to prioritize digitization were summarized and categorized. This summary of criteria was then presented to the wider community in the form of a survey for consolidation and gap analysis.

---

[1]FES (Fonds Economische Structuurversterking) project (EN):
https://www.naturalis.nl/media/library/2016/01/Self-evaluation_report_2009-2013.pdf
[2]E-ReColNat infrastructure: https://www.recolnat.org/en/

ICEDIG.EU

2. Objective

This deliverable D2.1 "Inventory of criteria for prioritization of digitization of natural history collections" will focus on creating a catalogue of the current criteria that can be used in order to prioritize digitization of natural history collections. Besides practical and funding criteria, we will specifically focus on criteria related to collection and economic, scientific and social relevance. The deliverable aims at contributing to an easy and well-informed decision-making process in relation to prioritization of digitization of natural history collections. To that end, it sets up a list of criteria to guide and lead to the implementation of the best strategy. Finally, the result of this task will be used in the design of a selection framework for DiSSCo to determine prioritization strategies. In preparation for this framework, three methods are described how to group and use the large set of criteria collected in this deliverable.

The deliverable builds on top of certain guiding principles and structural elements:
- DiSSCo utility
- Strategy oriented
- Bottom-up analysis
- Transnationality
- Diversity related, in terms of skills and roles
- Community usefulness
- Practical implementation

Anchored on those building blocks, the deliverable has been construed following a methodology and extracting conclusions that can be used for further setting a more detailed strategy under the DiSSCo framework.

ICEDIG.EU

3. Methodology

In light of the objective, a multi-stage process was adopted. It comprised of several steps that built subsequently on top of the preceding, as to produce a sound and thorough analysis that enabled us to draw solid conclusions regarding the prioritization of digitization of natural history collections.

The following stages were identified:
3.1) Literature and reports inventory
3.2) Targeted survey
3.3) Results revision
3.4) Identification of advisory methods

3.1. Collection of criteria from literature and reports

As a first step to get an overview of criteria used to prioritize digitization, an internet search was performed with combinations of the terms 'prioritization' 'digitization' and 'natural history collection'. We aimed to collect criteria from the entire range used to prioritize the digitization of natural history collections to date, therefore not distinguishing between international, national or institutional levels as an example. This resulted in a number of reports from which a first set of criteria was compiled (Appendix 3). Representatives from organizations such as E-ReColNat, iDigBio and ICEDIG partners were asked to review the set of criteria and to tackle all missing information and unclarities. The list was standardized, double entries merged and categorized (Appendix 4).

3.2. Collection of criteria by a targeted survey

3.2.1. Survey creation

An online survey was created using SurveyMonkey (https://www.surveymonkey.com/) (Appendix 2). It aimed at identifying which criteria were considered most important for the prioritization of digitization and whether any criteria were missing from the initial list. The first part of the survey consisted of 10 questions which addressed more general topics. In order to put the obtained results from the survey in an appropriate context, we first asked questions regarding the background of the person (e.g. role within institution) and the nature of their respective institution (e.g. name, type). Then, some questions tackled focus, diversity and estimated size of the collection of the institution of the respondent, usage of collections, and intention and progress of digitization efforts. The second and largest part of the survey dealt with the list of criteria composed in the first stage as a result of literature and reports inventory. Some questions contained a ranking option, asking respondents to assign a top 3 or a top 5 of criteria within one of four criteria categories (relevance categories: collection, economic, scientific and social relevance), besides practical and funding criteria (fundamental criteria categories). A separate question asked respondents to indicate which of the four relevance categories were most relevant when assessing digitization priorities. This ranking allowed us to discover which prioritization criteria received the highest value. Finally the third part of the survey included four questions asking for additional information (in free text) regarding the use of criteria and the digitization priorities of the respondents, and if in their opinion criteria were still missing. In total, the survey consisted of 20 questions with the final one being *'If you wish to be among the first to receive a summary of the results of this survey, please give your name and email address'*. This question was included to entice respondents to fill in the questionnaire and have them benefit from their contribution as well. However, the survey could always be completed anonymously if people did not wish to leave their contact details.

ICEDIG.EU

3.2.2. Contact list creation

Our intention was to reach out to a diverse audience in order to gauge whether the importance of certain criteria depended on variable conditions such as the role within an institution. We therefore attempted to receive responses from various types of institutions (e.g. herbarium, botanical garden and museum), from people with various roles within these institutions (e.g. collection manager, curator and manager) and from different countries across Europe. We established a list of over 900 contacts from 703 institutions (and departments within institutions) from 44 European countries (including Russia, Turkey, Israel and the Caucasus Republics). Where possible, we indicated the role of the contact within each institution and tried to address several persons within the same institution representing different roles. In this contact list, France had the highest number of contacts and from different institutions (145 contacts from 116 different institutions). Several sources were used to compile information, including the contact list of all participating institutions within DiSSCo and CETAF, GRBio (grbio.org, but currently offline (July 2018)), personal inquiries within the ICEDIG consortium and targeted internet searches. Only collection-holding institutions were included in this survey.

3.2.3. Survey distribution

The survey was distributed by sending an e-mail to each of the contacts and providing a direct link to the online survey. The call for participation was sent on the 9th of May 2018 and was open until the 15th of June 2018, approximately a period of 6.5 weeks (or 28 working days). The first reminder was sent on the 31st of May 2018 and the second reminder on the 12th of June 2018. Both times, the reminder resulted in a clear peak in the number of replies, thus being an adequate mean to increase the total number of responses.

3.2.4. Data analysis

After closing the survey, all individual responses were collected in a spreadsheet. Various responses containing information regarding institution names, roles and types of collections held by the institution, required a clean-up of the data. For example, institution names were written in different languages. Those responses needed an additional treatment to standardize the data obtained and to ensure that multiple responses from one institution were referred to with the same unique name. To indicate their position, respondents could select one role or the option 'other' for an open text answer. It soon became clear that various respondents did not have just one role within the institution but multiple ones. Therefore, frequently observed combinations of roles were added as an answer option. The same held true for the question regarding the type of institution, for which mixed nature was also allowed. Regarding the question on the types of collection held by the institution, the scope was enlarged as to include related collections such as lichens, wood, slime molds, and other plants or fungi. Since these types of collections were not initially among the answer options, they were included under 'Other plants and fungi (incl. lichens, mosses and algae)", as a possible answer. The answer option 'Fossils' included all responses in relation to geological collections. Within the answer option 'Minerals', we finally included all responses regarding petrology, meteorites and rock collections. This option was later renamed into 'Mineralogy, Petrology (incl. meteorites and rocks)'. In this survey, we considered a collection of up to 1,000,000 specimens as a small collection (incl. the answer options '0-100,000 specimens' and '100,000-1,000,000 specimens'), and above 1,000,000 specimens as a large collection (incl. the answer options '1,000,000-5,000,000 specimens' and 'more than 5,000,000 specimens'). Regarding the questions on criteria for prioritization, respondents were able to assign a top 3 or a top 5. Assigned ranks in these questions were converted into points. In top 5 rankings, rank 1 received 5 points, rank 2 received 4 points, etc. while in a top 3 ranking, rank 1 received 3 points, rank 2 received 2 points and rank 3 received 1 point. Thus, a high ranking (i.e. 1) resulted in a high value (i.e. 5).

The survey resulted in a total of 197 responses from 160 different institutions (and departments within institutions) from 34 European countries, though one response was excluded since it contained random texts and incomplete information. Of the remaining 196 responses, 127 resulted in a complete filled out questionnaire. The other 70 responses were incomplete and ended the survey before the seven questions about the prioritization of criteria were answered. Of the 127 there were only 68 respondents who correctly answered the ranking questions of the criteria. It happened often that respondents ranked not a top 5 or a top 3 as asked, but ranked within the same question various criteria with the same rank (e.g. 3 criteria within the same question were ranked '1'). Therefore the subsequent analysis of priorities was conducted on the 68 correctly answered survey only.

Chi-square statistical tests were performed to test the effect of the role of the respondent within his or her institution, the institution type and the size of the collection on the scoring of criteria within each of the four relevance categories. All statistical analyses were performed with the software RStudio.

3.3. Results revision and establishment of a list of criteria

The survey contained four questions where respondents were able to fill in a free-text response in case they felt criteria were missing or had additional suggestions. These four questions received hundreds of open-ended responses which were checked one by one towards the initial criteria list created under 3.1 above (Appendix 4). Suggested new criteria were validated by the authors and when accepted, added to the initial list. This exercise then resulted in the final set of criteria that could be used for prioritization of digitization (Appendix 5).

3.4. Identification of advisory methods

Due to the large number and varying types of criteria, we propose three methods that could be used to handle the large amount of criteria and help with developing a prioritizing process for digitization of bio- and geo-diversity collections (see section 5.4). These methods are proposed by the authors based upon numerous internal discussions on the topic. In regards to social relevant information we also obtained advice from LURIS, the knowledge Exchange Office from the University of Leiden, the Netherlands. LURIS creates and facilitates connections between science and social parties. We will further elaborate on these three methods in the section of Discussion below (see 6.4).

4. Results

4.1. Results of the literature study

The literature and reports inventory was based on a search for all information available in (scientific) literature regarding criteria being used for digitization of natural history collections. Combined, this resulted in a list of 177 criteria in total (Appendix 3). As a next step, we prepared a unique set of criteria by removing those criteria that were mentioned more than once and by merging similar criteria, resulting in an initial list of just over 100 criteria (Appendix 4). Due to this large number, we divided these criteria over six groups for a better overview: criteria relevant for collections, the economy, funding, practice, science and society. Within these six groups we created subgroups of no more than eight specified criteria each. Finally, all criteria were grouped and harmonized (and renamed where needed), resulting in a final list of the criteria more suitable for a survey and data analysis (Appendix 5).

4.2. Results of the survey

4.2.1. Survey respondents diversity

A large number of responses were received from a diverse realm of fields. We received in total 196 individual responses on the survey from 160 institutions (and departments within institutions) from 34 European countries. Most responses came from France, with 21 responses from 18 different institutions. We received nearly as many from Germany, with 20 responses from 17 different institutions. In relation to the institution itself, survey respondents were allowed to identify the type of their institution and could assign more than one type. This question Q2 (Fig. 1) was answered by 195 respondents, 112 of whom identified their institution as a museum, while only two (2) referred to it as an information/knowledge centre. Six (6) respondents identified themselves with an institution type that was not among the given options, which were: aquaria, non-governmental organizations (NGOs), libraries and zoological gardens. However, these six equally referred to their institutions to one of the given answer options, mostly as a museum. Respondents were also asked in Q3 to identify their role(s) within their institution (Fig. 2) which was completed by 195 respondents. The largest group of respondents were curators (79), while only nine (9) were technicians. Nine (9) respondents replied that they occupy a role different from those specified as answer options in the survey: citizen science expert, biodiversity informatics coordinator, data curator, data manager, development manager, digitization manager, GBIF node manager, liaison officer or teacher.

ICEDIG.EU
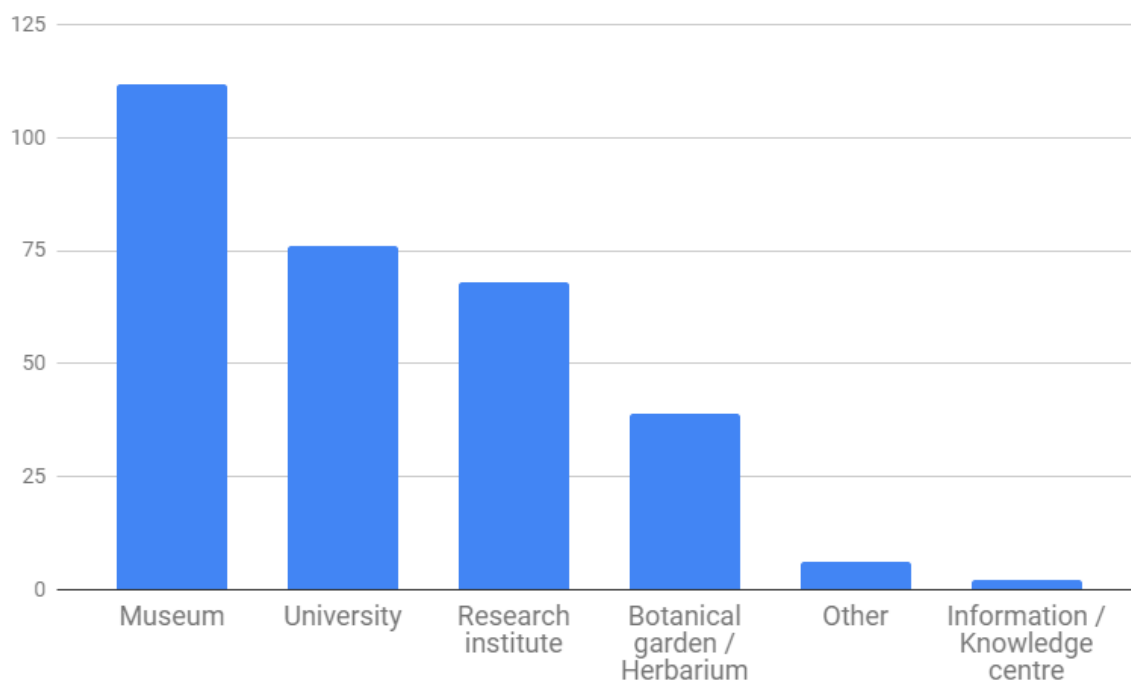
Fig. 1. Type of institution (Q2): What type of institution is it? From left to right: Museum (112); University (76); Research institute (68); Botanical garden / Herbarium (39); Other (6); Information/Knowledge center (2). *Numbers indicate the absolute number of responses.*
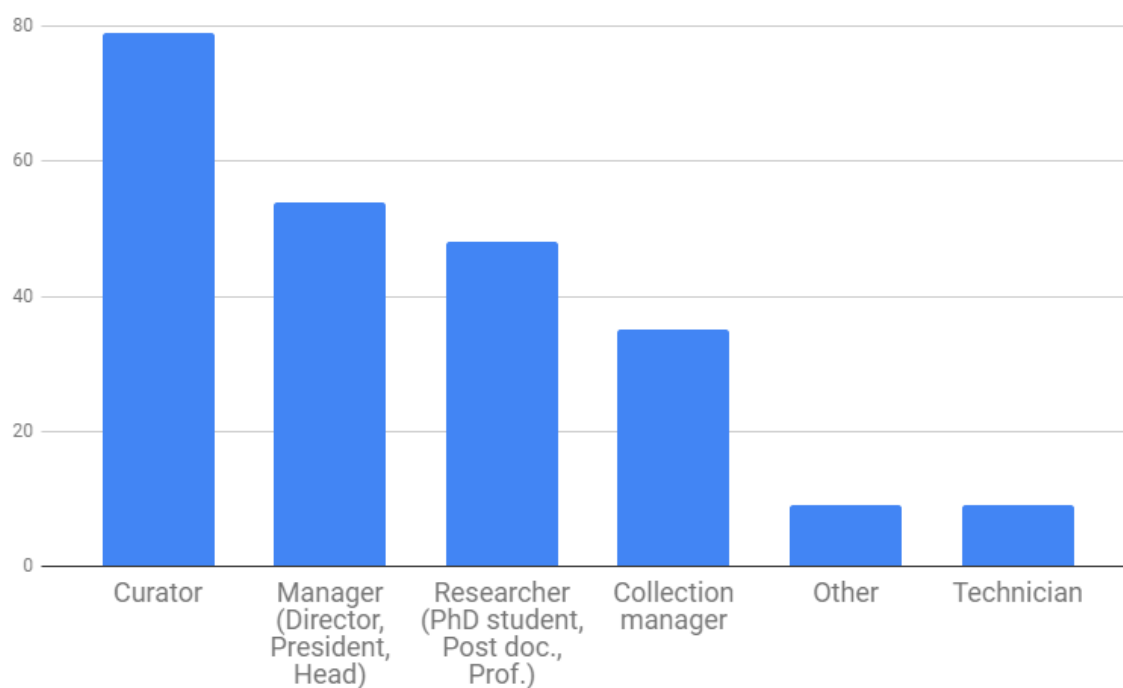
Fig. 2. Roles within the institution (Q3): What is your role within the institution? From left to right: Curator (79); Manager (Director, President, Head) (54); Researcher (PhD student, Postdoc, Professor) (48); Collection manager (35); Other (9); Technician (9). *Numbers indicate the absolute number of responses.*

ICEDIG.EU

4.2.2. Collection diversity and size

For the question regarding the taxonomic groups an institution has in its collection (Q4), we received 196 responses (Fig. 3). Most collections include vascular plants (151), while mineralogy and related collections were included least often (85). Seventeen (17) respondents included other non-taxonomic and non-biological collections types in their answer beside the given answer options including anthropology, archaeology, ethnology, library, scientific and technological collections. Some included other biological collections of living organisms, mainly for medical and pharmacological purposes.

We received 195 responses regarding the question on the approximate size of the collection within the institution (Q5; Fig. 4). Almost 60% of the respondents (116), representing 103 institutions from 31 countries identified their institution as a small collection, holding less than 1,000,000 specimens. On the other hand, almost 40% of the respondents (75), representing 56 institutions from 21 countries, indicated that their institution had a large collection (of more than 1,000,000 specimens). Four (4) respondents (2.1%) did not know the approximate size of their collection.
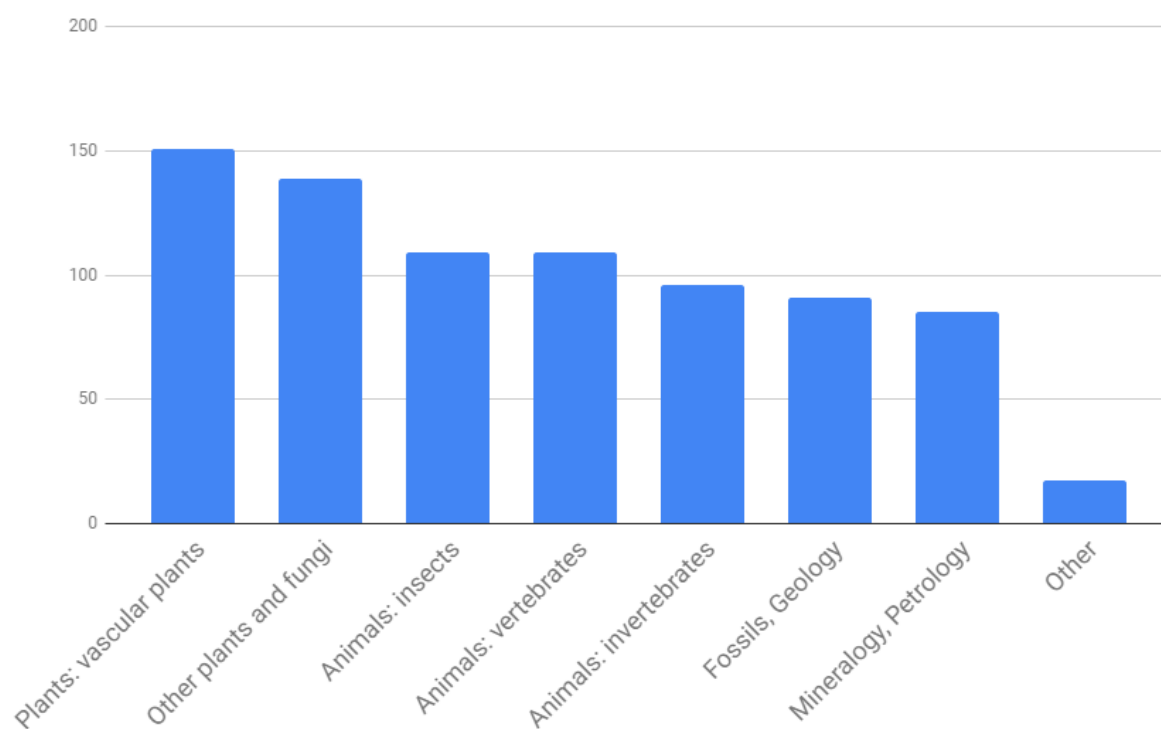
Fig. 3. (Taxonomic) groups in collections (Q4): Which (taxonomic) groups does your institution hold in its collection? From left to right: Plants: vascular plants (151); Other plants and fungi (incl. lichens, mosses and algae) (139); Animals: insects (109); Animals: vertebrates (birds, fish, mammals) (109); Animals: invertebrates (except insects) (96); Fossils, Geology (91); Mineralogy, Petrology (incl. meteorites and rocks) (85); Other (17). *Numbers indicate the absolute number of responses.*

ICEDIG.EU
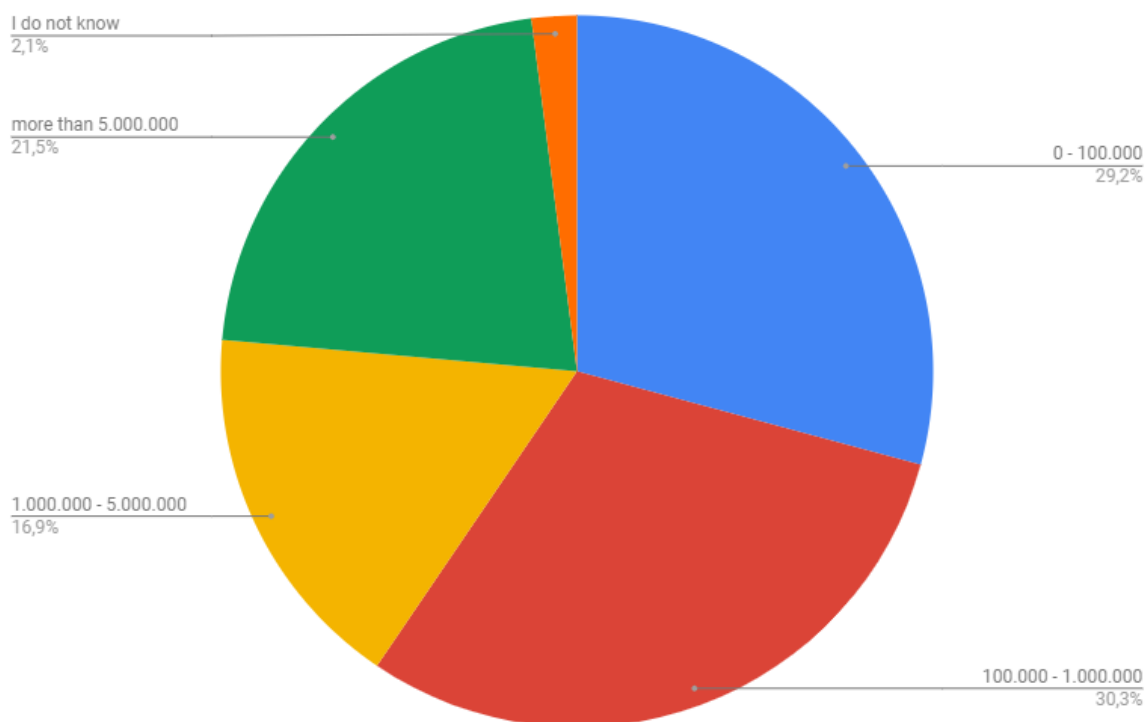
Fig. 4. Collection size (Q5): What is the approximate size of the collection, indicated by the number of specimens? *Data are given as a percentage.*

Smaller institutions often had their focus on a specific type of collection, either on plants (including 'Plants: vascular plants' and 'Other plants and fungi (incl. lichens, mosses and algae)'), animals (including 'Animals: invertebrates-except insects', 'Animals: insects' and 'Animals: vertebrates (birds, fish, mammals)') or fossils and minerals (including 'Fossils' and 'Mineralogy, Petrology (incl. meteorites and rocks)') (Fig. 5). Especially the smallest collections between 0-100,000 specimens often had a focus on a specific type of collection (60%), while the largest collections (more than 5,000,000 specimens) tended to have a wider scope (less than 25% focused on a specific type of collection). This is also likely connected to the results of the type of collections most institutions had, as plant collections were present among most institutions.

ICEDIG.EU
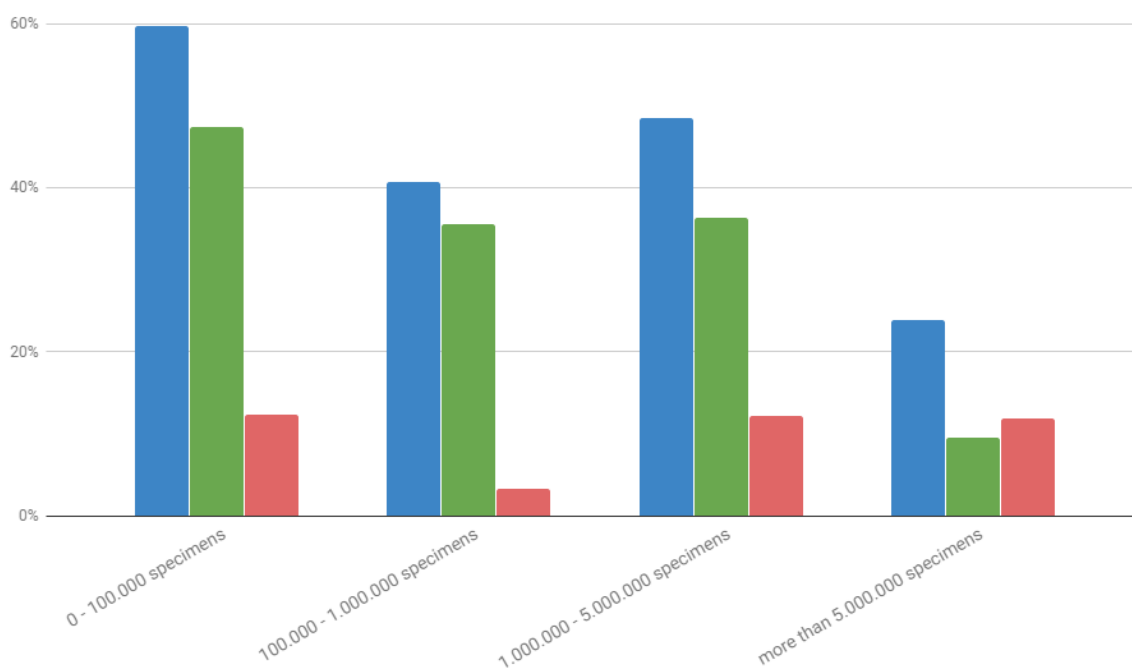
Fig. 5. Combination of institution size and type of collection (Q4 and Q5). *Blue bars* represent the percentage of institutions with a specific focus of just plants, animals or fossils and minerals. *Green bars* represent the percentage of institutions that focus on only plant collections. *Red bars* represent the percentage of institutions that focus on only animal collections. Blue bars: 0-100,000 specimens (59.74%), 100,000-1,000,000 specimens (40.7%), 1,000,000-5,000,000 specimens (48.5%), more than 5,000,000 specimens (23.8%). Green bars: 0-100,000 specimens (47.4%), 100,000-1,000,000 specimens (35.6%), 1,000,000-5,000,000 specimens (36.4%), more than 5,000,000 specimens (9.5%). Red bars: 0-100,000 specimens (12,3%), 100,000-1,000,000 specimens (3.4%), 1,000,000-5,000,000 specimens (12.1%), more than 5,000,000 specimens (11.9%).

4.2.3. Usage of collections and the intention to digitize collections

Three questions in the survey were aimed to gain insight in the use of collections for scientific research and the digitization activities. All three questions received 196 responses of which 98% indicated the collections were being used for scientific research, while only three (3) respondents answered this question with "No" and one (1) with "I do not know" (Q8). When it comes to the progress of digitization, we focused in the survey on two levels: databasing and imaging. With databasing we referred to the presence of a digital record of the specimen in a database, and with imaging, to the presence of a 2D or 3D image of the specimen in addition to the initial digital record. Almost half of all respondents indicated that their institution had databased more than 30% of the collection (Q10; Fig. 6). Over 90% of the institutions had databased some collection specimens. Most of the remaining respondents had the intention to digitize their collection. Only one 91) respondent answered the question with "No, and we are not planning to". This answer came from a small collection (0-100,000 specimens), which is nevertheless being used for research. When it comes to the creation of digital images of collection specimens, this had been less often performed than databasing collection specimens (Fig. 7). Only five (5) respondents answered that their institution had imaged all collection specimens. There were three (3) respondents who answered that they had both databased and imaged their collections completely ("Yes, 100%"), all three being representatives of institutions with collections below 100,000 specimens. Surprisingly, ten (10) respondents answered: "No, and we are not planning to".

ICEDIG.EU
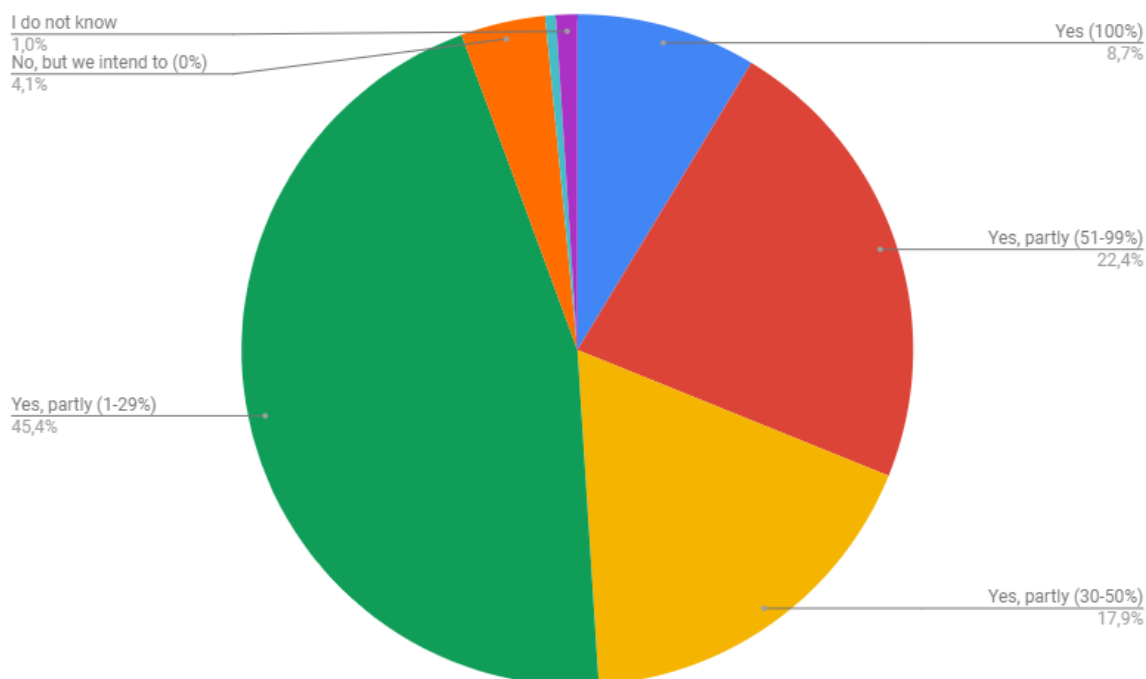
Fig. 6. Databased collections (Q9): Has your institution databased collection specimens? *Data is given as a percentage.*
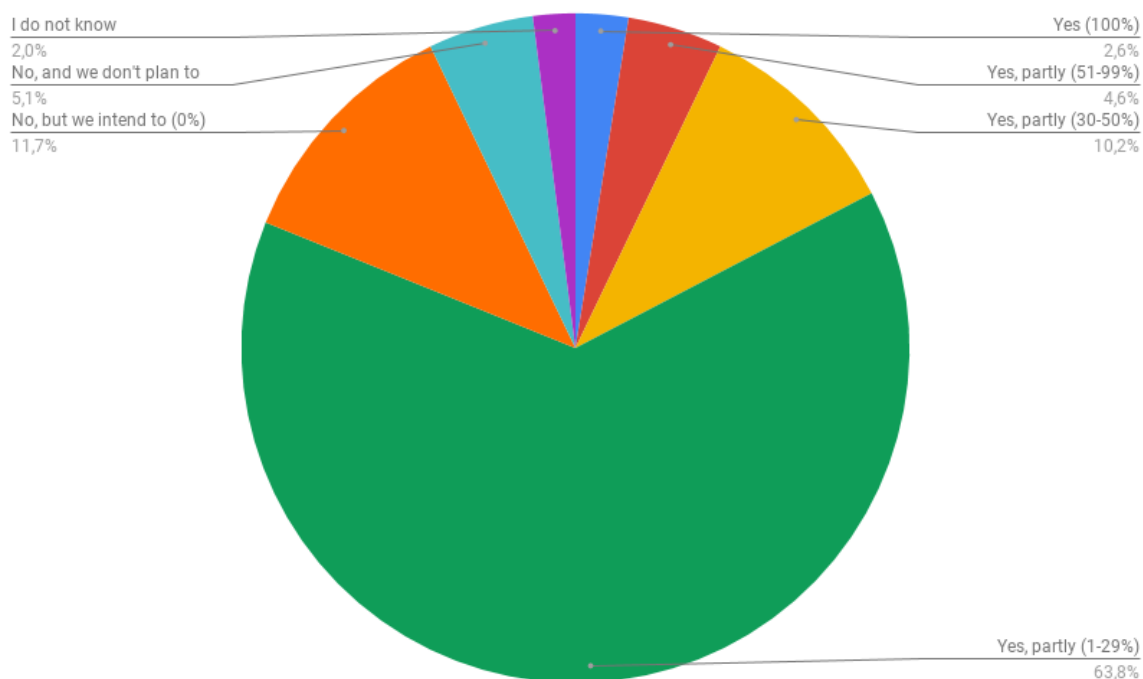
Fig. 7. Imaged collections (Q10): Has your institution imaged collection specimens? *Data is given as a percentage.*

ICEDIG.EU
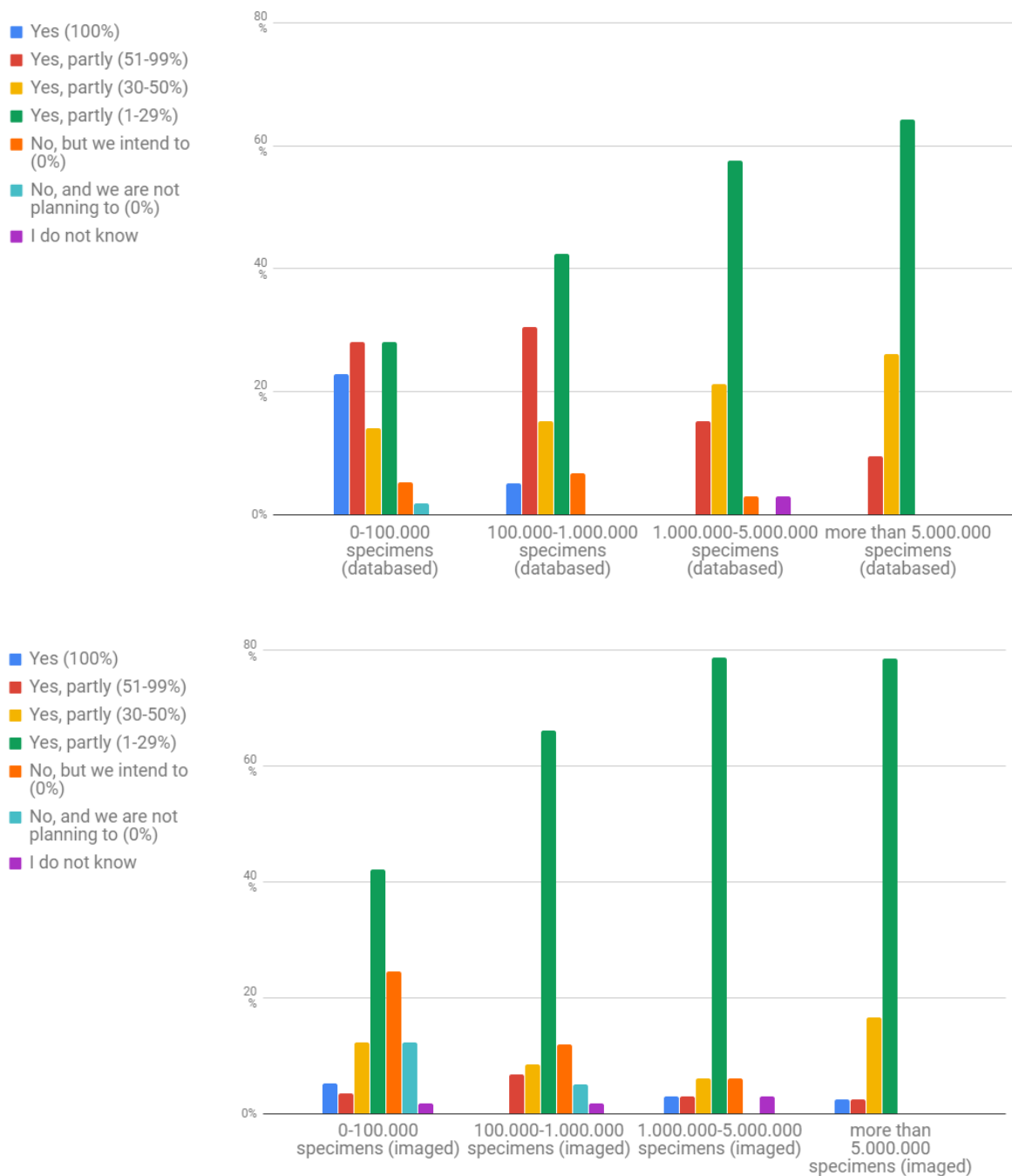
Fig. 8 and 9. Survey results for Q9 and Q10 in comparison with Q5 (collection size). Figure 8 and 9 represents the share of collections having databased (Fig. 8) or imaged (Fig. 9) their collections to a certain degree in relation to the collection size. Small collections (0-100,000 collection specimens) n=57, small collections (100,000-1,000,000 collection specimens) n=59, large collections (1,000,000-5,000,000 collection specimens) n=33, large collections (more than 5,000,000 collection specimens) n=42.

ICEDIG.EU

Table 1. Exact percentages accompanying figure 8.

Share	0 - 100,000 specimens	100,000 - 1,000,000 specimens	1,000,000 - 5,000,000 specimens	more than 5,000,000 specimens
Yes (100%)	22,8%	5,1%	0,0%	0,0%
Yes, partly (51-99%)	28,1%	30,5%	15,2%	9,5%
Yes, partly (30-50%)	14,0%	15,3%	21,2%	26,2%
Yes, partly (1-29%)	28,1%	42,4%	57,6%	64,3%
No, but we intend to (0%)	5,3%	6,8%	3,0%	0,0%
No, and we are not planning to (0%)	1,8%	0,0%	0,0%	0,0%
I do not know	0,0%	0,0%	3,0%	0,0%

Table 2. Exact percentages accompanying figure 9.

	0 - 100,000 specimens	100,000 - 1,000,000 specimens	1,000,000 - 5,000,000 specimens	more than 5,000,000 specimens
Yes (100%)	5,3%	0,0%	3,0%	2,0%
Yes, partly (51-99%)	3,5%	6,8%	3,0%	2,4%
Yes, partly (30-50%)	12,3%	8,5%	6,1%	16,7%
Yes, partly (1-29%)	42,1%	66,1%	79%	78,6%
No, but we intend to (0%)	24,6%	11,9%	6,1%	0,0%
No, and we are not planning to (0%)	12,3%	5,1%	0,0%	0,0%
I do not know	1,8%	1,7%	3,0%	0,0%

ICEDIG.EU

When it comes to the size of collections in comparison with the percentage of databased and imaged collection specimens, results showed that smaller collections had often databased a larger share of their collection than larger collections (Fig. 8 and Fig. 9, Table 1 and Table 2 for precise values). Larger collections had often at least started digitizing their collection, including both databasing and imaging of collection specimens. Of the larger institutions, 60 to 80% had databased and imaged up to a third of their collection specimens, while only about 3-6% of the respondents of institutions with large collections answered that they had not begun digitization yet but that they intended to do so. Among smaller institutions the share of those that had not started the digitization process but do intent to in the future, was 5-25%.

4.2.4. Prioritization of criteria

The prioritization questions received 68 completed responses, which were used in the following analysis. Results of the 127 responses – which include the 68 correct ones as well as the 59 incorrectly filled out answers to the ranking questions – are also available in Appendix 1.

Starting with the criteria for **collection relevance**, the criterion 'Capture data important specimens (historic, fragile, types)' scored the highest (total score of 273; Fig. 10). Second was the criterion 'Improve efficiency of collection management' (186), closely followed by the criterion 'Attract users and promote usage' (172). The criterion 'Implement policy' scored lowest (47). In regards to the **economic relevance**, the criterion 'Relevance for economic activities (indicator species, invasive species) and imperatives' was valued the most (total score of 171), followed by 'Create jobs' (127) and 'Generate revenues' (110; Fig. 11). Among the criteria for **scientific relevance** (Fig. 13), the criterion 'Relevant for fundamental research (taxonomy, ecology)' received the most points (total score of 235), followed by the criteria 'Relevant for access to primary biodiversity data' (187) and 'Relevant for research studying processes and/or trends (evolution, extinction, climate change)' (176). The criterion 'Part of national research initiative' scored lowest, with 'Part of in-house research programme' scoring only slightly higher. Finally, among the criteria related to **social relevance** (Fig. 14), the criterion 'Contribute to public awareness, education or outreach' was valued the most (total score of 236), followed by 'Contribute to conservation (policy)' (176) and 'Underpin importance of collections to stakeholders and public' (167). The criterion 'Comply with legal rules and regulations' was scored very low among this set of criteria (52).

ICEDIG.EU

Collection Relevance
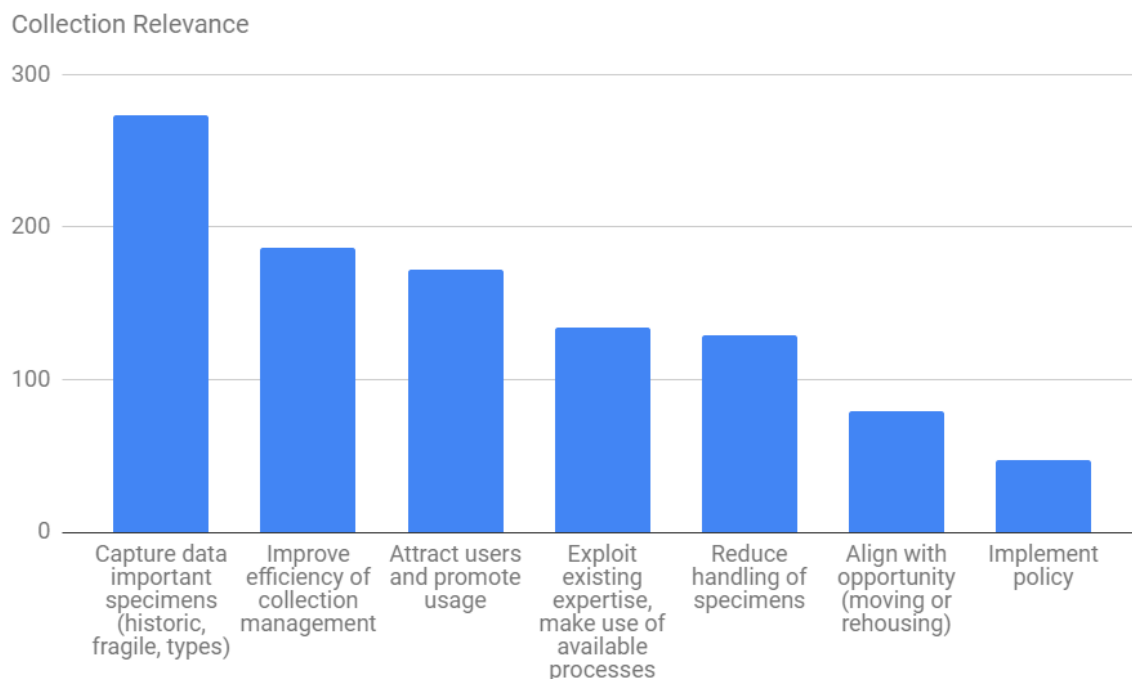
Fig. 10. Collection-relevant criteria (Q12): In the category 'collection relevance', which criteria do you think are the most important for the prioritization of digitization of natural history collections?. From left to right: Capture data important specimens (historic, fragile, types) (273); Improve efficiency of collection management (186); Attract users and promote usage (172); Exploit existing expertise, make use of available processes (134); Reduce handling of specimens (129); Align with opportunity (moving or rehousing) (79); Implement policy (47). *Numbers indicate the total points given by the respondents.*

Economic relevance
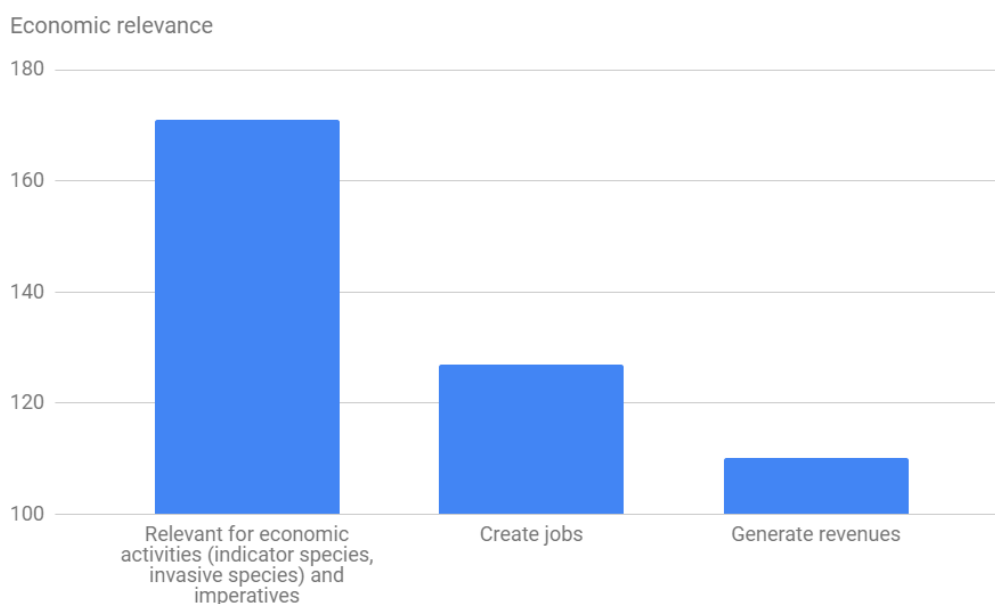
Fig. 11. Economic-relevant criteria (Q13): In the category 'economic relevance', which criteria do you think are the most important for the prioritization of digitization of natural history collections?. From left to right: Relevant for economic activities (indicator species, native species) and imperatives (171); Create jobs (127); Generate revenues (110). Numbers indicate the total points given by the respondents. *Numbers indicate the total points given by the respondents.*

ICEDIG.EU

Scientific relevance
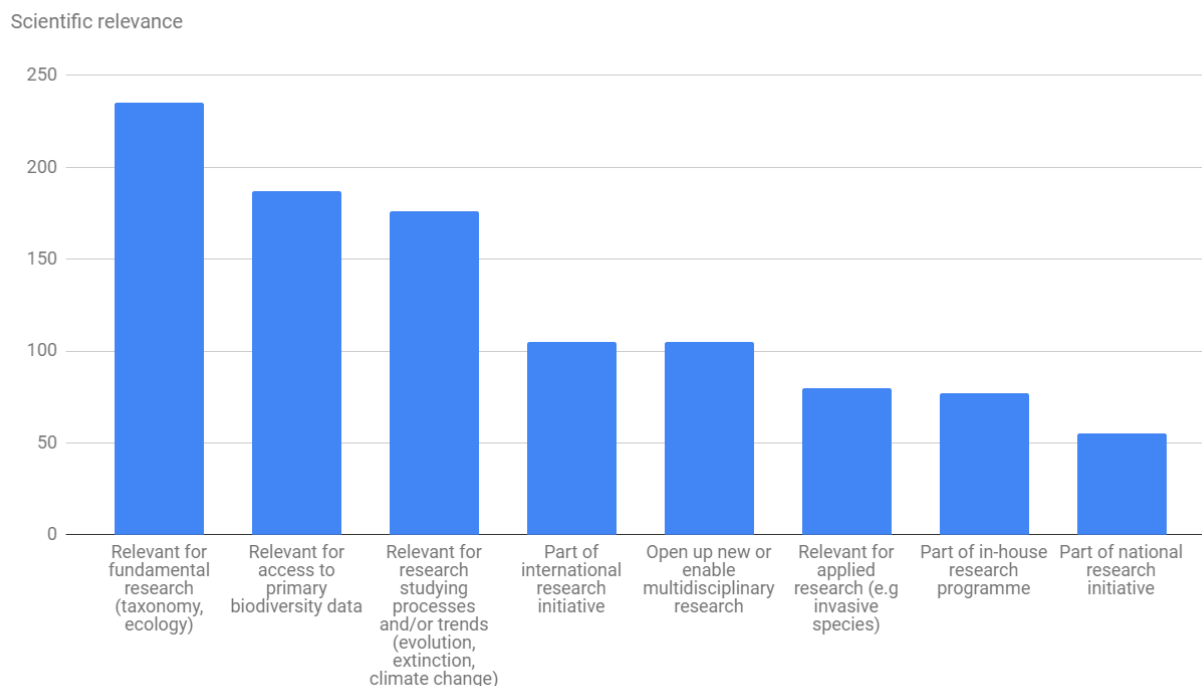
Fig. 13. Scientific-relevant criteria (Q14): In the category 'scientific relevance', which criteria do you think are the most important for the prioritization of digitization of natural history collections?. From left to right: Relevant for fundamental research (taxonomy, ecology) (235); Relevant for access to primary biodiversity data (187); Relevant for research studying processes and/or trends (evolution, extinction, climate change) (176); Part of international research initiative (105); Open up new or enable multidisciplinary research (105); Relevant for applied research (e.g. invasive species) (80); Part of in-house research programme (77); Part of national research initiative (55). *Numbers indicate the total points given by the respondents.*
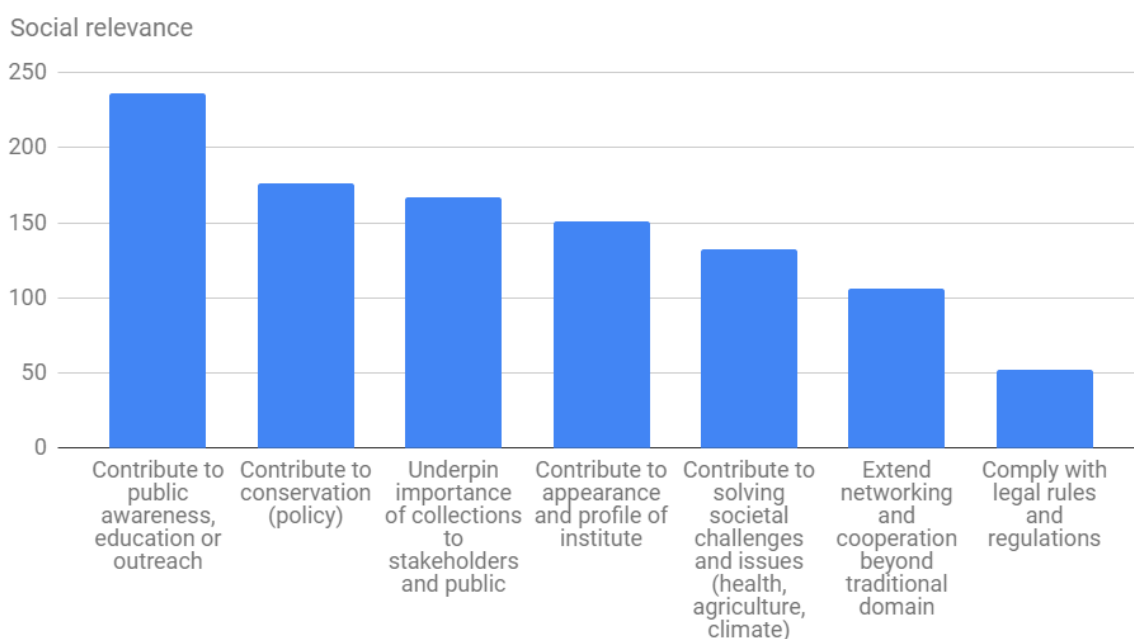
Social relevance

Fig. 14. Social-relevant criteria (Q15): In the category 'Social relevance', which criteria do you think are the most important for the prioritization of digitization of natural history collections?. From left to right: Contribute to public awareness, education or outreach (236); Contribute to conservation (policy) (176); Underpin importance of collections to stakeholders and public (167); Contribute to appearance and profile of institution (151); Contribute to solving societal challenges and issues (health, agriculture, climate) (132); Extend networking and cooperation beyond traditional domain (106); Comply with legal rules and regulations (52). *Numbers indicate the total points given by the respondents.*

ICEDIG.EU

In the questionnaire, respondents were able to select which of the criteria among the relevance categories they found the most relevant (Fig. 15) to prioritize future digitization efforts. This question received 127 responses. Scientific relevance was the most selected one (94) by far, followed by collection relevance (30). Social relevance was selected only three (3) times, while economic relevance was never selected. In addition to these results, some respondents made use of the open-ended answer questions to include comments or additions. Here, several respondents remarked that the collection is often the primary basis for science, therefore collection relevance and scientific relevance are intrinsically linked.
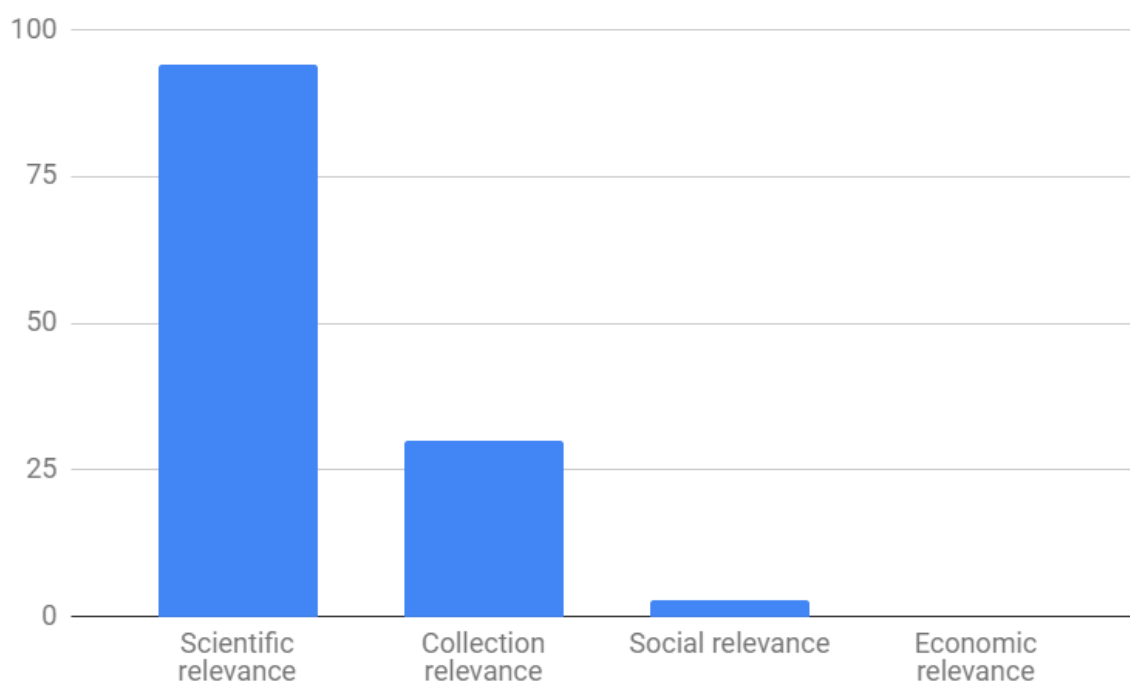
Fig. 15. Relevance categories (Q16): What is according to you the most relevant category for prioritization of collection digitization? From left to right: Scientific relevance (94), Collection relevance (30), Social relevance (3) and Economical relevance (0).

4.2.5. Relating role within the institution, institution type and collection size to scoring of criteria

The total scoring of the criteria within each category did not significantly depend on either the role of the respondent within the institution, the type of the institution or the size of the collection (Table 3). There was only a trend for the size of the collection to affect the scoring of the criteria related to collection relevance. When testing for the effect of collection size on the scoring of each individual criteria (Fig. 16) there was a significant effect for every criteria, except for the criterion 'Reduce handling of specimens' (Chi-square, df=3 $\chi2=7.7$, p>0.05). The largest difference between collections of different sizes were observed for the criterion 'Attract users and promote usage', with institutions with small collections giving a lower value to this criterion than institutions with large collections. Also, there was a large difference between collection sizes for the criterion 'Implementing policy', which was overall the criterion that scored lowest within this category. Institutions with very small collections valued this criterion higher than those with larger collections.

ICEDIG.EU

Table 3. Scoring of relevance categories: Results of chi-square tests to determine the dependence of the scoring of criteria within each of the relevance categories on the role within the institution, the type of the institution and the size of the collection. Significant values are indicated with an asterisk, near significant values (i.e. trends) are indicated with a circle. Tests are performed based on standardized data.

	Collection relevance			Scientific relevance			Economic relevance			Social relevance		
	df	χ^2	p	df	χ^2	p	df	χ^2	p	df	χ^2	p
Role within institution	24	22.0	0.58	28	25.4	0.61	8	4.5	0.81	24	20.7	0.65
Type of institution	24	15.5	0.91	28	27.6	0.49	8	1.9	0.98	24	21.2	0.63
Size of collection	18	26.3	0.09 ○	21	23.6	0.31	6	2.4	0.88	18	8.2	0.98
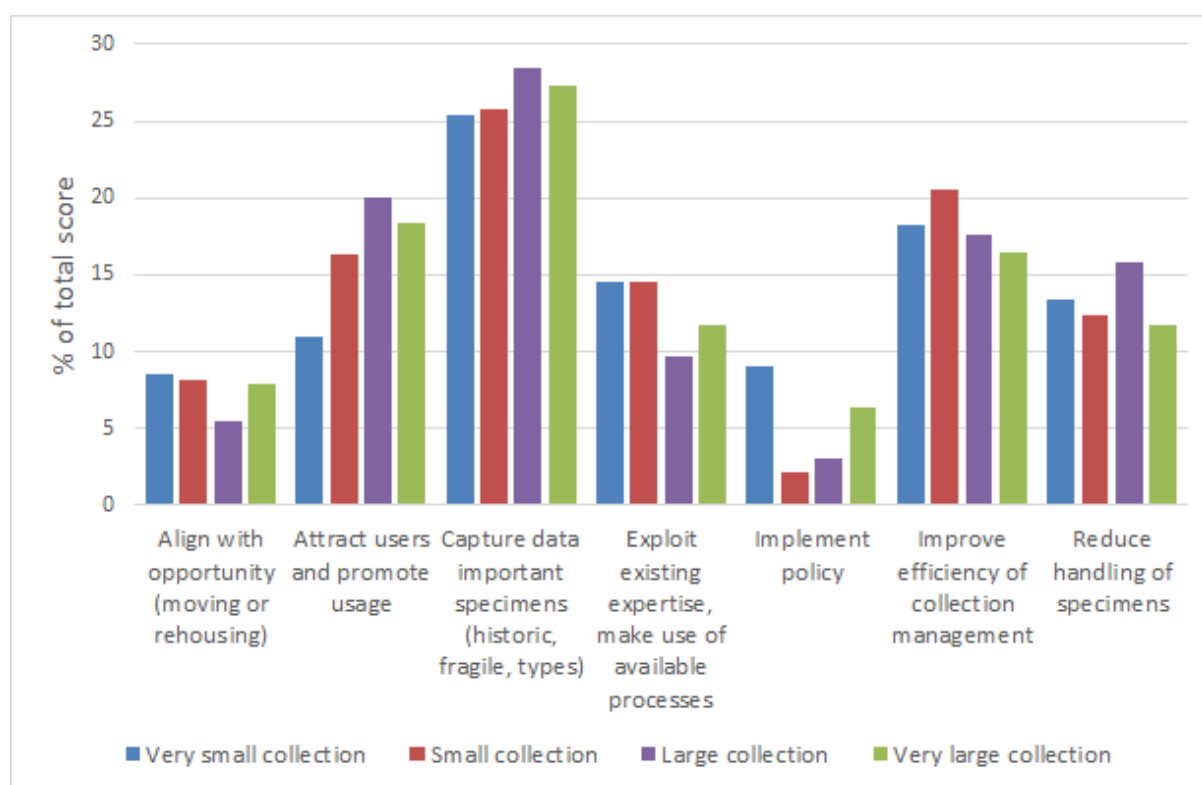
Figure 16. Percentage of the total score given to each of the criteria related to collection relevance by the respondents, shown for each of the four collection sizes.

4.2.6. Questions resulting in additional information regarding used criteria to prioritize digitization

A total of four questions were included in the survey to ask respondents if they were missing criteria or had suggestions for additional criteria. Among those four questions we received hundreds of suggestions from most of the respondents which were checked one by one to avoid duplications and errors. Though most responses contained suggestions for criteria which were already present in the initial list, it often resulted in new information that enabled us to make further adjustments and enhancements of individual criteria (Appendix 4 and 5).

Several responses also contained valuable information that established new criteria that were missing in the initial list provided. Most suggestions were related to criteria for collection relevance and practical criteria, however, several suggestions were given for funding criteria. All new suggested criteria were validated by task members and added to the criteria list. This resulted in a total of six new criteria, of which three related to collections

relevant, two to practical matters and one to funding. Two collection relevant criteria received an adjustment to include further detail to it.

One of the four questions was about current criteria used by the institution to prioritize digitization. This question was open-ended and 173 respondents answered to this question. Some of the most common answers included type material, regional collections, historical collections, new accessions, specimens of endemic and red-list/endangered species, as well as scientifically valuable collections.

ICEDIG.EU

5. Discussion

With this deliverable, we aimed to provide an index of criteria, related to scientific, economic, social and collection relevance or practical aspects, which are currently used for the prioritization of digitization of natural history collections. As a result of the consultancy process undertaken, we created a list of over 100 criteria in total that can be used for the prioritization of digitization, indicating there may be many reasons to choose what (not) to digitize first in a certain natural history collection. The analysis of the results obtained from the literature and report inventory allowed us to make a distinction between the criteria specifically related to the feasibility of digitization in the first place (mainly focused on resource availability) and the criteria leading to determine what to digitize first, once digitization takes place.

The process of gathering results from the literature and report search together with the data obtained from the conducted survey also allowed an internal debate on the primary topics to consider. In particular, several subjects were outlined and were placed at the heart of the discussions and lead the analysis of the data compiled, namely, 1) the different progress stages of digitization; 2) the criteria deemed to be important; 3) the relationship between criteria selected and the nature of the collection/personal/institution involved; and 4) the diverse methods for establishing prioritization of digitization.

5.1. Progress of databasing and imaging of collection specimens

A majority of the participants in our study indicated that their institution had databased at least part of the specimens, while imaging had been performed to a much smaller degree. Databasing thus seems to be prioritized over imaging, which is probably related to the fact that imaging specimens requires more time, expertise, specialized software, equipment and therefore funding than databasing. Especially institutions holding smaller collections indicated they were not planning to image their collection specimens, and to a (much) smaller degree not to database their collection specimens. While we did not investigate their reasons for not databasing and/or imaging collection specimens, it may be that institutions holding small collections found it too much work, did not have the resources or expertise or deemed the task of digitization less important.

Only five (5) respondents answered that their institution had imaged all collection specimens. Surprisingly, this answer appeared to be in contrast with the previous question regarding databasing collection specimens. That is because respondents that answered the question about imaging with "Yes, 100%", did not always answer the previous question about databasing with "Yes, 100%". While databasing is often considered to be the first step in digitization, followed by imaging (Krishtalka et al. 2016), when it comes to herbarium specimens, it sometimes is the other way around. Here, the first step in digitizing can be imaging of the sheets (specimens) and then databasing them based on that image which could explain this discrepancy. The respondents who did answer that they have imaged more specimens than databased always had at least herbarium collections. There may have been a different understanding of terminology among respondents, as the definition of digitization is not uniform and may for example differ between types of collection or simply between different persons.

ICEDIG.EU

5.2. Criteria that are deemed most important are related to scientific relevance

Scientific relevance was most often indicated as the most important criteria category, which is in accordance to Krishtalka et al. (2016). Despite the acknowledged priority of scientific needs, in reality this process has been often driven by other factors (e.g. funding). Within DiSSCo, it will therefore be crucial to set a clear, harmonized and unified digital (scientific) agenda to be able to determine which specimens need to be digitized first. Since many respondents indicated that they see scientific and collection relevance to be closely intertwined, the prioritization will have to be based on criteria related to both, scientific and collection relevance. While for collection relevance, the criterion 'Capture data important specimens (historic, fragile, types)' received the highest score for scientific relevance it was the criterion 'Relevant for fundamental research (taxonomy, ecology). Thus, it will be important to first **determine both the relevance for fundamental research and whether data-important specimens will be captured** when making a prioritization of digitization of natural history collections.

5.3. Collection size tended to affect the scoring of collection relevance criteria

When testing whether belonging to a certain group influences the ranking, we only found a trend for an effect of collection size on the scoring of criteria related to collection relevance. This suggests that criteria related to scientific, social and economic relevance are scored equally among respondents occupying different roles, working at different institution types and being associated with different collection sizes. It means that even though people had different backgrounds, they had largely similar ideas to what is important when prioritizing the digitization of natural history collections. Therefore, this finding is expected to support future discussions between DiSSCo partners, highlighting that although there are many cultural and other differences between partners, we find **consensus on the prioritization criteria important for the digitization** of natural history collections.

For collection relevance, however, the largest differences between collection sizes appeared to be for the criterion 'Attract users and promote usage'. Institutions with small collections deemed this criterion to be less important than institutions with large collections. A possible explanation is that institutions with large collections are more aware of the potential future uses of their collection and may generally be more focused on outreach. On the other hand, it could also be related to funding. Although government funding is not directly linked to actually working with a collection, there is the idea that more usage supports negotiations when it comes to funding. Nevertheless, it is surprising that there is quite a large difference in assigning value to this specific criterion, as the increase of collection use, especially for research, is often mentioned on a higher governmental level as the main reason to be digitizing collection specimens in the first place.

ICEDIG.EU

5.4. Three methods to use the criteria for prioritization

The final topic to be addressed in this deliverable is how to use the results from this task in practice. Based on the large set of criteria resulting from the analysis of questionnaire and literature, we propose a few optional methods of using these criteria for determining the strategy for a digitization project.

1. Decision tree (based on practical and funding criteria);
2. Scoring method (for an extensive evaluation);
3. Panel review (for a less extensive and (nationally) focused evaluation).

The first method considers two fundamental groups of criteria (namely funding and resources), which determine if all necessary practical aspects have been taken into account and whether the goal set out is achievable. For these two sets of criteria a **decision tree** can be established as an online tool. A first attempt as an illustration has been included as Appendix 6. This example is not a finalized decision tree by any means, but it is used to visualize this potential method. A decision tree makes it possible to tackle every single question as a multivariate key, as often used in taxonomy. Using a decision tree will make the user think about every single practical and funding criteria and finally gives an advice regarding the amount of funding (still necessary) or which practical resources and knowledge are still needed before starting a digitization process.

The criteria of the relevance categories can be used with either one of the other two possible methods, i.e. **scoring method and panel review**, or by a combination of both. However, it is important to recognize that these criteria may be differently valued at institutional, national and international level.

In our survey we requested to select a top 3 or top 5 for criteria, which can be used as a prioritization score at European level. A scoring method at a different level or at a different point in time could be based on the prioritization resulting from the survey and the same method, giving the most valued criteria the highest score, and the lowest valued criteria the lowest score. A first attempt on creating this scoring method would be time-consuming for setting the right parameters, but once established and robust, ultimately time-saving. Economic, scientific and social relevance can change over time and these criteria seem to be more difficult to prioritize.

This last consideration was received as information and advice from LURIS, but also by other ICEDIG partners. Therefore the option of a panel was strongly suggested, either seperate or in combination with the scoring method. The panel is expected to have validators with expertise in all different criteria groups, so they can value the benefit of digitization efforts and prioritize digitization based on the criteria delivered here. A panel could be formed at international level, e.g. organized by related organizations such as CETAF that could bring in its expertise and advice. A panel could also be formed at national or institutional level to validate and prioritize criteria at these respective levels. The formed panels would then set out their priorities of criteria for future digitization efforts which would then apply to the level the panel is convened on and could define a time period after which the priorities should be revisited.

5.5. Limitations of the study

The survey was focused to receive responses from a wide audience related to collection-holding institutions. The total number of respondents can be considered as high and covered a wide range of roles, institutions and countries. Although this diversity in respondents is highly valued in terms of drawing broad conclusions, it did not allow for an in-depth analysis

ICEDIG.EU

at institutional or national level as we received only a few responses per country and/or per institution. Therefore, the focus of this study is at European level.

Furthermore, answering the questions on ranking priorities was not as easy to respond to as expected. From the total of 198 respondents only 68 completed the survey with correct answers including their rankings. Two factors are supposedly at work here. First it appears that the design of this type of specific questions was difficult to implement correctly in the online tool SurveyMonkey, which may have hampered responding correctly to these questions. This will be important to take into account when conducting other surveys within ICEDIG. Secondly, respondents rushing through the questionnaire and not reading the instructions beyond the first few words also contributed to this phenomenon. This is backed up by the rather high number of respondents who contacted us when facing difficulties with an explanation hinting at this problem. However, this was only possible to become a problem because of the first issue of imperfect implementation of the questionnaire.

Time limitation was another element hindering the optimization of results. Within the confines of this study, it was not possible to keep track of changes within scientific and social relevance due to the lack of long-term monitoring opportunities. This study took place over a period of six months and for achieving well-grounded conclusions in terms of importance of economic, scientific and social aspects, information over several years should be taken into account. Therefore, this study reflects the current state of valuing criteria for the prioritization of digitization, which may only be to some extent indicative for future evolvement.

These limitations, focus on the European level, imperfect results, limited time, are why it will be important to use the proposed methods to set the criteria for prioritization of digitization at any other moment in time and at different levels.

ICEDIG.EU

6. Conclusions and recommendations

In order to determine which (parts of) natural history collections in Europe need to be digitized first, it is crucial to have clear criteria in order to set these priorities. This study provides a thoroughly review and comprehensive list of such criteria. Also, it is useful to know which criteria for the prioritization of digitization are considered to be most important by collection-holding institutions. We find that scientific criteria to prioritize digitization are most valued by personnel associated with collection-holding institutions, especially when digitization promotes biodiversity research focused on taxonomy, ecology, evolution, extinction and/or climate change. In particular, the prioritization of digitization should initially focus on stimulating fundamental research (e.g. taxonomy), closely followed by encouraging research focused on studying processes and trends (e.g. evolution) and enhancing access to primary biodiversity data. To exactly determine which specimens should be digitized first in a scientific context, it will be essential to further prioritize digitization using most likely a panel consisting of biodiversity scientists from different fields.

Most importantly, our study highlights that regardless of background, people at collection-holding institutions find scientific criteria most important to determine what to digitize first. Nevertheless, first and foremost it needs to be decided what the digitization project will entail, what practical aspects need to be covered and what is needed in terms of resources and funding. We recognize that institutions will have to make a final decision regarding digitization while taking into account **competing interests at an international, national and institutional scale**. Within DiSSCo, it will be important to set a clear **digital (research) agenda** which can serve as a guideline to determine what to prioritize in terms of digitization of collection specimens in more detail.

ICEDIG.EU

7. Acknowledgements

We would like to thank LURIS, the knowledge Exchange Office from the University of Leiden (Leiden, the Netherlands), for their advice regarding social relevant criteria. We would like to express our gratitude to members of E-ReColNat and Deborah Paul (iDigBio) for their help with obtaining information about criteria being used for digitization prioritization. Finally, we would like to thank many members of the ICEDIG consortium during various stages of the project, particularly for reviewing the criteria.

ICEDIG.EU

8. Cited literature

Ariňo, A.H. 2010. Approaches to estimate the universe of natural history collections data. *Biodiversity Informatics*, 7: 81-92.

Baird, R. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics*, 7: 130-136.

Berendsohn, W.G. and Seltmann, P. 2010. Using geographical and taxonomic metadata to set priorities in specimen digitization. *Biodiversity Informatics*, 7: 120-129.

Berents, P., Hamer, M. and Chavan, V. 2010. Towards demand-driven publishing: Approaches to the prioritization of digitization of natural history collections data. *Biodiversity Informatics*, 7: 113-119.

Blagoderov, V., Kitching, I.J., Livermore, L., Simonsen, T.J. and Smith, V.S. 2012. No specimen left behind: Industrial scale digitization of natural history collections. *Zookeys*, 209: 133-146.

Heerlien, M., van Leusen, J., Schnörr, S., de Jong-Kole, S., Raes, N. and van Hulsen, K. 2015.The natural history production line: An industrial approach to the digitization of scientific collections. *ACM Journal on Computing and Cultural Heritage*, 8: article 3, 11 pages.

Holmes, M.W., Hammond, T.T., Wogan, G.O.U., Walsh, R.E., LaBerbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M. and Nachman, M.W. 2016. Natural history collections as windows on evolutionary processes. *Molecular Ecology*, 25: 864-881.

Frazier, C.K., Wall, J. and Grant, S. 2008. Initiating a Natural History Collection Digitisation Project, Copenhagen: GBIF Secretariat. Available online at: https://www.gbif.org/document/80574/initiating-a-collection-digitisation-project.

Kalms, B. 2012. Digitisation: A strategic approach for natural history collections. *Atlas of Living Australia, CSIRO Ecosystem Sciences*, Australia. 95 pp.

Kemp, C. 2015. The endangered dead. *Nature*, 518: 292-294.

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

Vollmar, A., Macklin, J. A., and Ford, L.S. 2010. Natural history specimen digitization: challenges and concerns. *Biodiversity Informatics*, 7: 93-112.

ICEDIG.EU

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1 – Appendix 1

ICEDIG.EU

Appendix 1. Criteria rankings

The following tables (2-9) contain all exact scores for the different criteria resulting from the survey. These scores were established for both, biased and unbiased data, where biased data also contained information of incomplete and incorrectly scored top 3 and top 5 ranking questions from the survey.

Table 1: Unbiased data for Collection relevance scores.

	Align with opportunity (moving or rehousing)	Attract users and promote usage	Capture data important specimens (historic, fragile, types)	Exploit existing expertise, make use of available processes	Implement policy	Improve efficiency of collection management	Reduce handling of specimens	Total score
Score by Overall = 68	77	171	270	134	47	181	125	1005
Score by Collection Managers (n=16)	17	57	61	25	9	40	31	240
Score by Curators (n=25)	31	43	105	59	11	74	52	375
Score by Manager (n=21)	24	46	79	39	19	51	42	300
Score by Researchers (n=11)	11	30	46	18	9	36	15	165
Score by Technicians and others (n=5)	5	16	15	9	8	18	4	75
Score by Botanical Gardens (n=12)	16	28	47	24	18	22	25	180
Score by Museums (n=46)	52	119	178	86	32	126	82	675
Score by Research Institutes (n=31)	35	75	121	56	17	87	59	450
Score by Universities (n=14)	14	43	59	27	9	31	27	210
Score by Information / Knowledge centers and others (n=5)	3	9	21	13	2	14	13	75
Score by small institutes (0 - 100.000 specimens) (n=11)	14	18	42	24	15	30	22	165
Score by small institutes (100.000 - 1.000.000 specimens) (n=22)	27	54	85	48	7	68	41	330
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=11)	9	33	47	16	5	29	26	165
Score by large institutes (5.000.000 and more specimens) (n=22)	25	58	86	37	20	52	37	315

ICEDIG.EU

Table 2: Biased data for Collection relevance scores.

	Align with opportunity (moving or rehousing)	Attract users and promote usage	Capture data important specimens (historic, fragile, types)	Exploit existing expertise, make use of available processes	Implement policy	Improve efficiency of collection management	Reduce handling of specimens	Total
Score by Overall (n=127)	182	339	510	286	102	366	288	2073
Score by Collection Managers (n=28)	42	93	110	69	12	89	83	498
Score by Curators (n=49)	81	106	208	111	39	148	112	805
Score by Manager (n=42)	63	110	163	93	44	123	104	700
Score by Researchers (n=27)	30	72	109	63	27	84	53	438
Score by Technicians and others (n=6)	5	19	20	13	8	22	8	95
Score by Botanical Gardens (n=26)	35	59	95	46	25	45	57	362
Score by Museums (n=82)	131	225	328	165	74	245	175	1343
Score by Research Institutes (n=51)	76	132	203	115	33	158	114	831
Score by Universities (n=42)	49	133	169	100	32	122	115	720
Score by Information / Knowledge centers and others (n=9)	8	25	38	29	5	22	22	149
Score by small institutes (0 - 100.000 specimens) (n=29)	45	58	105	71	29	80	59	447
Score by small institutes (100.000 - 1.000.000 specimens) (n=34)	41	93	134	81	20	112	85	566
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=27)	39	85	113	50	26	76	68	457
Score by large institutes (5.000.000 and more specimens) (n=35)	55	95	148	75	27	96	77	573

ICEDIG.EU

Table 3: Unbiased data for Economic relevance scores.

	Create jobs	Generate revenues	Relevant for economic activities (indicator species, invasive species) and imperatives	Total score
Score by Overall = 68	126	107	169	402
Score by Collection Managers (n=16)	24	30	42	96
Score by Curators (n=25)	52	38	60	150
Score by Manager (n=21)	37	33	50	120
Score by Researchers (n=11)	23	15	28	66
Score by Technicians and others (n=5)	9	7	14	30
Score by Botanical Gardens (n=12)	23	21	28	72
Score by Museums (n=46)	83	73	114	270
Score by Research Institutes (n=31)	53	48	79	180
Score by Universities (n=14)	26	23	35	84
Score by Information / Knowledge centers and others (n=5)	11	8	11	30
Score by small institutes (0 - 100.000 specimens) (n=11)	23	17	26	66
Score by small institutes (100.000 - 1.000.000 specimens) (n=22)	41	37	54	132
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=11)	17	19	30	66
Score by large institutes (5.000.000 and more specimens) (n=22)	42	31	53	126

ICEDIG.EU

Table 4: Biased data for Economic relevance scores.

	Create jobs	Generate revenues	Relevant for economic activities (indicator species, invasive species) and imperatives	Total score
Score by Overall = 127	229	210	306	745
Score by Collection Managers (n=16)	43	49	72	164
Score by Curators (n=25)	96	78	115	289
Score by Manager (n=21)	72	67	101	240
Score by Researchers (n=11)	54	46	62	162
Score by Technicians and others (n=5)	11	8	17	36
Score by Botanical Gardens (n=12)	47	40	54	141
Score by Museums (n=46)	146	133	200	479
Score by Research Institutes (n=31)	88	85	125	298
Score by Universities (n=14)	76	75	101	252
Score by Information / Knowledge centers and others (n=5)	18	17	20	55
Score by small institutes (0 - 100.000 specimens) (n=11)	56	52	66	174
Score by small institutes (100.000 - 1.000.000 specimens) (n=22)	59	57	82	198
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=11)	47	47	65	159
Score by large institutes (5.000.000 and more specimens) (n=22)	64	51	87	202

ICEDIG.EU

Table 5: Unbiased data for Scientific relevance scores.

	Open up new or enable multidisciplinary research	Part of in-house research programme	Part of international research initiative	Part of national research initiative	Relevant for access to primary biodiversity data	Relevant for applied research (e.g invasive species)	Relevant for fundamental research (taxonomy, ecology)	Relevant for research studying processes and/or trends (evolution, extinction, climate change)	Total score
Score by Overall = 68	105	77	105	55	187	80	235	176	1020
Score by Collection Managers (n=16)	21	25	23	10	47	20	56	38	240
Score by Curators (n=25)	34	33	34	29	76	21	82	66	375
Score by Manager (n=21)	43	18	27	12	48	27	73	52	300
Score by Researchers (n=11)	19	7	23	7	20	16	40	33	165
Score by Technicians and others (n=5)	11	8	12	2	17	7	10	8	75
Score by Botanical Gardens (n=12)	14	13	19	17	23	20	44	30	180
Score by Museums (n=46)	82	54	66	33	123	46	149	122	675
Score by Research Institutes (n=31)	40	45	54	27	74	30	112	68	450
Score by Universities (n=14)	11	13	24	10	42	21	48	41	210
Score by Information / Knowledge centers and others (n=5)	14	4	7	1	14	2	18	15	75
Score by small institutes (0 - 100.000 specimens) (n=11)	20	10	18	17	31	9	35	25	165
Score by small institutes (100.000 - 1.000.000 specimens) (n=22)	37	16	24	12	65	31	76	69	330
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=11)	16	8	19	6	39	3	43	31	165
Score by large institutes (5.000.000 and more specimens) (n=22)	26	38	37	16	44	32	74	48	315

ICEDIG.EU

Table 6: Biased data for Scientific relevance scores.

	Open up new or enable multidisciplinary research	Part of in-house research programme	Part of international research initiative	Part of national research initiative	Relevant for access to primary biodiversity data	Relevant for applied research (e.g invasive species)	Relevant for fundamental research (taxonomy, ecology)	Relevant for research studying processes and/or trends (evolution, extinction, climate change)	Total score
Score by Overall (n=127)	214	157	249	145	382	207	456	318	2128
Score by Collection Managers (n=28)	54	28	59	29	91	57	107	81	506
Score by Curators (n=49)	79	79	79	66	159	73	166	129	830
Score by Manager (n=42)	93	39	90	47	119	66	158	103	715
Score by Researchers (n=27)	38	29	64	34	73	49	99	67	453
Score by Technicians and others (n=6)	11	8	16	2	22	12	15	13	99
Score by Botanical Gardens (n=26)	24	39	51	31	55	40	84	58	382
Score by Museums (n=82)	160	97	149	74	246	125	291	220	1362
Score by Research Institutes (n=51)	90	83	105	53	145	71	188	119	854
Score by Universities (n=42)	52	43	92	61	132	89	157	126	752
Score by Information / Knowledge centers and others (n=9)	25	9	20	6	25	12	28	26	151
Score by small institutes (0 - 100.000 specimens) (n=29)	56	28	64	42	80	44	99	63	476
Score by small institutes (100.000 - 1.000.000 specimens) (n=34)	62	30	45	36	109	59	126	97	564
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=27)	43	43	67	41	98	26	104	64	486
Score by large institutes (5.000.000 and more specimens) (n=35)	51	51	69	23	87	73	125	93	572

ICEDIG.EU

Table 7: Unbiased data for Social relevance scores.

	Contribute to public awareness, education or outreach	Contribute to appearance and profile of institute	Comply with legal rules and regulations	Contribute to conservation (policy)	Contribute to solving societal challenges and issues (health, agriculture, climate)	Underpin importance of collections to stakeholders and public	Extend networking and cooperation beyond traditional domain	Total score
Score by Overall (n=68)	236	151	52	176	132	167	106	1020
Score by Collection Managers (n=16)	61	26	18	41	36	43	15	240
Score by Curators (n=25)	95	65	21	62	27	53	52	375
Score by Manager (n=21)	62	50	11	52	44	51	30	300
Score by Researchers (n=11)	30	25	5	27	26	27	25	165
Score by Technicians and others (n=5)	20	12	3	17	8	11	4	75
Score by Botanical Gardens (n=12)	38	25	9	37	25	31	15	180
Score by Museums (n=46)	168	101	40	110	81	118	57	675
Score by Research Institutes (n=31)	94	73	24	81	63	73	42	450
Score by Universities (n=14)	48	30	7	37	28	32	28	210
Score by Information / Knowledge centers and others (n=5)	19	19	1	13	3	16	4	75
Score by small institutes (0 - 100.000 specimens) (n=11)	37	35	6	27	15	26	19	165
Score by small institutes (100.000 - 1.000.000 specimens) (n=22)	88	40	13	62	42	53	32	330
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=11)	39	25	10	32	18	26	15	165
Score by large institutes (5.000.000 and more specimens) (n=22)	61	46	21	50	49	53	35	315

ICEDIG.EU

Table 8: Biased data for Social relevance scores.

	Contribute to public awareness, education or outreach	Contribute to appearance and profile of institute	Comply with legal rules and regulations	Contribute to conservation (policy)	Contribute to solving societal challenges and issues (health, agriculture, climate)	Underpin importance of collections to stakeholders and public	Extend networking and cooperation beyond traditional domain	Total score
Score by Overall (n=127)	405	322	126	338	258	339	251	2039
Score by Collection Managers (n=28)	99	66	28	86	70	85	45	479
Score by Curators (n=49)	162	134	46	124	71	129	108	774
Score by Manager (n=42)	130	118	36	110	96	118	99	707
Score by Researchers (n=27)	69	72	32	71	55	69	64	432
Score by Technicians and others (n=6)	20	16	8	22	13	14	4	97
Score by Botanical Gardens (n=26)	67	68	29	65	51	60	34	374
Score by Museums (n=82)	270	202	84	212	150	243	147	1308
Score by Research Institutes (n=51)	142	125	47	147	110	132	108	811
Score by Universities (n=42)	132	118	42	113	76	125	96	702
Score by Information / Knowledge centers and others (n=9)	25	29	5	24	18	29	19	149
Score by small institutes (0 - 100.000 specimens) (n=29)	87	96	39	64	54	57	61	458
Score by small institutes (100.000 - 1.000.000 specimens) (n=34)	135	70	28	93	60	92	65	543
Score by small institutes (1.000.000 - 5.000.000 specimens) (n=27)	89	71	27	78	57	80	61	463
Score by large institutes (5.000.000 and more specimens) (n=35)	88	84	31	98	79	103	62	545

ICEDIG.EU

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1 – Appendix 2

Appendix 2. Survey questions

Table 1: Survey questions and their various ways of options for answering.

No.	Question	Answers	Remarks
1	What is the name of your institute?	Open-Ended Response	
2	What type of institute is it?	Botanical garden	Choose one
		Information / Knowledge center	
		Museum	
		Museum and Research institute	
		Museum and University	
		Research institute	
		University	
		Other (please specify)	
3	What is your role within the institute?	Collection manager	Choose one
		Curator	
		Manager (Director, President, Head)	
		Researcher (PhD student, Post doc., Prof.)	
		Technician	
		Other (please specify)	
4	Which (taxonomic) groups does your institute hold in its collection? Multiple answers possible	Animals: invertebrates (except insects)	Option to choose multiple
		Animals: insects	
		Animals: vertebrates (birds, fish, mammals)	
		Fossils	
		Minerals	

		Plants: mosses, algae, fungi	
		Plants: vascular plants	
		Other (please specify)	
5	What is the approximate size of the collection?	0 00.000 specimens	Choose one
		100.000 .000.000 specimens	
		1.000.000 - 5.000.000 specimens	
		more than 5.000.000 specimens	
		I do not know	
6	How many collection managers/curators are working in the collection of the institute? Give an average number	Open-Ended Response	
7	How many paid employees are involved in collection digitisation at the institute? Give an average number	Open-Ended Response	
8	Is the collection being used for (scientific) research?	Yes	Choose one
		No	
		I do not know	
9	Has your institute databased collection specimens?	Yes (100%)	Choose one
		Yes, partly (51-99%)	
		Yes, partly (30-50%)	
		Yes, partly (1-29%)	
		No, but we intend to (0%)	
		No, and we are not planning to (0%)	
		I do not know	
10	Has your institute imaged collection	Yes (100%)	Choose one

		specimens?	Yes, partly (51-99%)	
			Yes, partly (30-50%)	
			Yes, partly (1-29%)	
			No, but we intend to (0%)	
			No, and we are not planning to (0%)	
			I do not know	
11	If yes (even partly), what were the main criteria for the prioritisation of digitisation (databasing and imaging combined) of the collection?		Open-Ended Response	
12	In the category 'collection relevance', choose the top 5 of criteria you think are the most important for the prioritisation of digitisation of natural history collections and rank them (1 = most important, 5 = least important).		Align with opportunity (moving or rehousing)	Select a top 5
			Attract users and promote usage	
			Capture data important specimens (historic, fragile, types)	
			Exploit existing expertise, make use of available processes	
			Implement policy	
			Improve efficiency of collection management	
			Reduce handling of specimens	
13	In the category 'economic relevance', choose the top 3 of criteria you think are the most important for the prioritisation of digitisation of natural history collections and rank them (1 = most important, 3 = least important).		Create jobs	Select a top 3
			Generate revenues	

		Relevant for economic activities (indicator species, invasive species) and imperatives	
14	In the category 'scientific relevance', choose the top 5 of criteria you think are the most important for the prioritisation of digitisation of natural history collections and rank them (1 = most important, 5 = least important).	Open up new or enable multidisciplinary research	Select a top 5
		Part of in-house research programme	
		Part of international research initiative	
		Part of national research initiative	
		Relevant for access to primary biodiversity data	
		Relevant for applied research (e.g. invasive species)	
		Relevant for fundamental research (taxonomy, ecology)	
		Relevant for research studying processes and/or trends (evolution, extinction, climate change)	
15	In the category 'social relevance', choose the top 5 of criteria you think are the most important for the prioritisation of digitisation of natural history collections and rank them (1 = most important, 5 = least important).	Contribute to public awareness, education or outreach	Select a top 5
		Contribute to appearance and profile of institute	
		Comply with legal rules and regulations	
		Contribute to conservation (policy)	
		Contribute to solving societal challenges and issues (health, agriculture, climate)	
		Underpin importance of collections to stakeholders and public	

		Extend networking and cooperation beyond traditional domain	
16	What is according to you the most relevant category for prioritisation of collection digitisation?	Collection relevance	Select one
		Economic relevance	
		Scientific relevance	
		Social relevance	
17	Are there any criteria that according to you are missing in the complete list?	Open-Ended Response	
18	Are there any criteria that according to you are missing in the following list of funding criteria? List:	Cover costs of requirements	Indicate if anything is missing in the open-ended response field
		Cover exploitation costs	
		Feasible within the expected costs	
		Increase funding opportunities	
		Indicate expected costs	
		Attract external funding in addition to in-house funding	
		Receive additional funding in different ways for this collection	
		If yes, please indicate which criterion you miss:	
19	Are there any criteria that according to you are missing in the following list of practical criteria? List:	Accessibility of collection specimens	Indicate if anything is missing in the open-ended response field
		Availability of equipment and protocols (including safety)	
		Availability of staff	
		Feasibility within the given time	

		Indicate current condition of specimens	
		Indicate project goals/digitisation results	
		Indicate size and type of collection and its specimens	
		Indicate level of digitisation (databasing, imaging)	
		If yes, please indicate which criterion you miss:	
20	If you wish to be among the first to receive a summary of the results of this survey, please give your name and e-mail address.	Name	Optionally; open-ended response field
		Company	
		Address	
		Address 2	
		City/Town	
		State/Province	
		ZIP/Postal Code	
		Country	
		Email Address	
		Phone Number	

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1 – Appendix 3

Appendix 3. First set of criteria

Table 1: A first set of criteria, including the sources.

Criteria	Source
Address a recognized need or gap for learning, teaching or research provision	Kalms 2012
Are at risk of being lost to the community through sale, deterioration and disaggregation	Kalms 2012
Condition of items in the collection	Kalms 2012
Contributing to create critical mass within a given area	Kalms 2012
Digitisation of objects in poor condition	Kalms 2012
Discoverability of specimens within a collection	Kalms 2012
Enable access to and use of difficult or impossible to access collections	Kalms 2012
Help to create a theme across previously unassociated materials	Kalms 2012
Improved accountability for objects	Kalms 2012
Increased public accessibility of specimens	Kalms 2012
Inspire new avenues of research, or new approaches within learning and teaching	Kalms 2012

Likely users of the digital assets	Kalms 2012
Map to a particular area of the curriculum or research interest	Kalms 2012
Nature of the collection (e.g. wet or dry specimens)	Kalms 2012
Objects at high risk of losing their integrity (e.g. objects in need of repair)	Kalms 2012
Objects of international significance (e.g. type specimens)	Kalms 2012
Objects of local significance	Kalms 2012
Objects of national significance (e.g. explorer notebooks)	Kalms 2012
Objects with high user demands	Kalms 2012
Purpose of digitisation (e.g. accounting of all items)	Kalms 2012
Range of items in the collection	Kalms 2012
Reduced handling of high value specimens	Kalms 2012
Would not otherwise be funded, or be able to attract significant funding from other sources	Kalms 2012
Enhances biodiversity collection infrastructure for sustained curation	BCoN - Questionnaire
Enhances capacity for using collections data in research	BCoN - Questionnaire

Enhances use of biodiversity collection data in conservation policy	BCoN - Questionnaire
Enhances use of biodiversity collection data in formal education	BCoN - Questionnaire
Enhances use of biodiversity collection data in informal education	BCoN - Questionnaire
Improves curation and storage of collections	BCoN - Questionnaire
Increases human diversity in collections management and research personnel using collections	BCoN - Questionnaire
Involvement of more collections in the digitization process	BCoN - Questionnaire
Priority for national or international collections	BCoN - Questionnaire
Digitisation of collections associated with museum projects (e.g. exhibitions)	Berents 2010
Digitisation of collections associated with research projects	Berents 2010
Digitisation of collections that result from research projects in ecosystems that provide services	Berents 2010
Ecosystem relevance	Berents 2010

Historical significance	Berents 2010
Prioritization of type specimens (specifically holotypes)	Berents 2010
Species of special concern: diseases or disease vectors	Berents 2010
Species of special concern: harvested species (e.g crops)	Berents 2010
Species of special concern: invasive alien species	Berents 2010
Species of special concern: pests	Berents 2010
Species of special concern: threatened, endangered, endemic species	Berents 2010
Taxonomic priorities (ideally linked to global or national initiatives)	Berents 2010
Scientific panel to determine the scientific relevance	E-Recolnat
Small or large collection (size)	E-Recolnat
Aid research by reducing future transcription time of specimen labels	Frazier 2008
Choose how much data is captured per specimen (number of data fields)	Frazier 2008
Collection data needs to be available due to legislation (access to information for publicly funded institutions)	Frazier 2008
Data capture or data interpretation	Frazier 2008
Data quality requirements	Frazier 2008

Digitisation project must fit within the goals/requirements of the institution	Frazier 2008
Enable your collection data to be studied in different ways	Frazier 2008
Enhance curational activities (e.g. keep track of specimen state)	Frazier 2008
Enhances the ability of the institution to contribute in areas beyond its traditional domain	Frazier 2008
Funding requirements	Frazier 2008
Identification of the end users of the digitisation effort	Frazier 2008
Location where the digitisation will take place	Frazier 2008
Monitor the size, growth and usage of the collection	Frazier 2008
Number of records to digitize	Frazier 2008
Project deadlines already set	Frazier 2008
Protect your specimens (reduced specimen handling)	Frazier 2008
Raising the profile of the institution and/or collection	Frazier 2008
Scale of digitisation (institution-wide or one-person project)	Frazier 2008
Sensitivity of revealing collection data (e.g. geographical location of red list species)	Frazier 2008
Sensitivity of revealing collection data due to legislation	Frazier 2008

Support from the institution/knowledgeable people on digitisation	Frazier 2008
Target a specific family or species	Frazier 2008
Target important specimens	Frazier 2008
Time scale for digitisation/data entry	Frazier 2008
What personal is available	Frazier 2008
Wider dissemination of collection data	Frazier 2008
Availability of expertise / required expertise	Krishtalka 2016
Availability of funding / required funding	Krishtalka 2016
Availability of time / required time	Krishtalka 2016
Does effort exceeds perceived benefit	Krishtalka 2016
Does the data contain possible errors	Krishtalka 2016
Is the task feasible (not overwhelming)	Krishtalka 2016
Presence of partnership in a larger community effort	Krishtalka 2016
Priority for the individual in charge of the collection	Krishtalka 2016
Relevance for climate change impacts	Krishtalka 2016

Relevance for discovery and access to primary biodiversity data	Krishtalka 2016
Relevance for economic well-being	Krishtalka 2016
Relevance for environmental well-being	Krishtalka 2016
Relevance for extinction, threatened species and habitat loss	Krishtalka 2016
Relevance for food security	Krishtalka 2016
Relevance for geographic focus	Krishtalka 2016
Relevance for human well-being	Krishtalka 2016
Relevance for indicators of minerals (e.g. plants as indicators of minerals)	Krishtalka 2016
Relevance for institutional priority / policy / mission	Krishtalka 2016
Relevance for invasive species	Krishtalka 2016
Relevance for public health (e.g. zoonotic diseases and environmental contaminants)	Krishtalka 2016
Relevant for health and human services	Krishtalka 2016
Sufficient information on digitization process available	Krishtalka 2016
Target the most urgent biodiversity science imperatives of our time	Krishtalka 2016
Target the most urgent economic imperatives of our time	Krishtalka 2016

Target the most urgent environmental imperatives of our time	Krishtalka 2016
Target the most urgent social imperatives of our time	Krishtalka 2016
Top 3 criteria: (1) (increases?) research	Krishtalka 2016
Top 3 criteria: (2) (increases?) funding/grant opportunities	Krishtalka 2016
Top 3 criteria: (3) taxonomic focus	Krishtalka 2016
World's key challenges: food security	Krishtalka 2016
World's key challenges: impacts of climate change	Krishtalka 2016
World's key challenges: invasive species	Krishtalka 2016
World's key challenges: zoonotic disease outbreaks	Krishtalka 2016
Are exploitation costs results covered	Vermeulen 2013
Are more than one party involved	Vermeulen 2013
Are there additional finances	Vermeulen 2013
Captures irreplaceable knowledge of employees	Vermeulen 2013
Collection is historical value	Vermeulen 2013
Collection is of scientific value	Vermeulen 2013

Content of the collection	Vermeulen 2013
Contributes in the disclosure of collection	Vermeulen 2013
Contributes to the conservation of collection	Vermeulen 2013
Direct or indirect necessary for research	Vermeulen 2013
Does it contribute to the permanent digital infrastructure	Vermeulen 2013
Duration of digitization	Vermeulen 2013
Enables multidisciplinary analyzes	Vermeulen 2013
Enables new publications or improves publications	Vermeulen 2013
Feasibility in costs	Vermeulen 2013
Feasibility in time	Vermeulen 2013
Fits well with already finished digitalized collection	Vermeulen 2013
Follows the research question the policy of the institute	Vermeulen 2013
If research is part of the research programme from within the institute	Vermeulen 2013
Implementation requirements (procedures; materials, etc.)	Vermeulen 2013
Increases recognition of the institute	Vermeulen 2013

Is of interest of public of the museum	Vermeulen 2013
Is the minimum set of required data digitized? (e.g. minimum set of dataquality)	Vermeulen 2013
Is the used material of the institute/museum	Vermeulen 2013
Necessary in order to increase feasibility and completion of other projects	Vermeulen 2013
Open new fields within scientific research	Vermeulen 2013
Outreach: of interest of audience	Vermeulen 2013
Personal (collection manager) opinion to be necessary to be done	Vermeulen 2013
Physical size of the collection	Vermeulen 2013
Quality of digitization	Vermeulen 2013
Research in progress or planned within the institute	Vermeulen 2013
Scientific interest relevance (based on survey outcome)	Vermeulen 2013
Serves a social interest	Vermeulen 2013
What are the costs	Vermeulen 2013
Assets selected for web presentation for use to publicize the type of resources available for use	Johnson 2012
Collection materials selected for exhibition digitized to facilitate exhibition activities	Johnson 2012

Collection materials selected for publication being made more broadly available	Johnson 2012
Creation of 3-D objects of all sizes and complexity	Johnson 2012
Creation of digital descriptive records	Johnson 2012
Creation of mixed media	Johnson 2012
Digitization is completed in conjunction with improving efficiency of administering digital assets	Johnson 2012
Digitization is completed in conjunction with moving or rehousing collections	Johnson 2012
Digitization is done in support of a special event	Johnson 2012
Digitization of collection and associated metadata to improve access to materials that are hidden or hard to find	Johnson 2012
Digitization of live collections	Johnson 2012
Digitization of other materials that include books, journals, maps, etc., in and of themselves.	Johnson 2012
Digitization to make it possible to share some collection materials in multiple venues as a means to generate revenue	Johnson 2012
Make collection materials available for wider educational purposes	Johnson 2012
Make collection materials more readily available for collaboration or distant research	Johnson 2012
Preserve and protect collection materials that are at risk and cannot be used	Johnson 2012

Are sufficient staff and volunteers available	Meise
Are sufficient staff available in-house	Meise
Are sufficient staff available outsourced	Meise
Are the objects accessible and findable in the collection	Meise
Are the objects for mass digitisation or digitisation on demand	Meise
What is the average time of the complete digitisation process of one object	Meise
Collections personnel will prioritize their specimen holdings for digitisation	IDigBio
Advice from stakeholders	IDigBio
Advice of collections professionals	IDigBio
Advice of oversight committees	IDigBio
Are images included	IDigBio
Are methods in place to handle the different types of collection? (e.g. glas preparates, pinned, alcohol, dry, flat, different sizes, etc.)	IDigBio
Defined metrics for measuring progress against explicit goals and reporting progress	IDigBio
Does it enable new science	IDigBio

Does it provide more effective monitoring and regulatory activities of biodiversity	IDigBio
Open access availability	IDigBio
Amount of data available per specimen	Vollmar 2010
Available equipment (i.e. software and storage space)	Vollmar 2010
Available staff (number of persons and number of knowledgeable persons)	Vollmar 2010
Spatial structure of the collection (including physical)	Vollmar 2010
Time needed to image/database one specimen	Vollmar 2010

Sources:

Berents, P., Hamer, M. and Chavan, V. 2010. Towards demand-driven publishing: Approaches to the prioritization of digitization of natural history collections data. *Biodiversity Informatics*, 7: 113-119.

Frazier, C.K., Wall, J. and Grant, S. 2008. Initiating a Natural History Collection Digitisation Project, Copenhagen: GBIF Secretariat. Available online at: https://www.gbif.org/document/80574/initiating-a-collection-digitisation-project.

IDigBio. 2010. A Strategic Plan for Establishing a Network Integrated Biocollections Alliance. IDigBio. https://digbiocol.files.wordpress.com/2010/08/niba_brochure.pdf

Johnson, K. 2012. *Unit Digitisation Plan*. Smithsonian NMNH. 49 pp.

Kalms, B. 2012. Digitisation: A strategic approach for natural history collections. *Atlas of Living Australia, CSIRO Ecosystem Sciences*, Australia. 95 pp.

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

Vermeulen, J.J. 2013. *Aanpak prioriteitstelling binnen de digitalisering van de collectie, aan de hand van de prioritering in de digistraat Entomologie.* Naturalis Biodiversity Center, Leiden, The Netherlands. 34 pp.

Vollmar, A., Macklin, J. A., and Ford, L.S. 2010. Natural history specimen digitization: challenges and concerns. *Biodiversity Informatics*, 7: 93-112.

BIodiversity Collections Network - Questionnaire: https://bcon.aibs.org/2018/02/07/survey-harnessing-biodiversity-collections-data-for-addressing-national-challenges/

E-ReColNat: *Personal communication with E-ReColNat (Paris, France)*

Meise: *Personal communication with Botanic Garden Meise (Meise, Belgium)*

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1 – Appendix 4

Appendix 4. Unique set of criteria

The following list contains a unique set of criteria, in which criteria are standardized, double entries merged and categorized. The column for 'topic' describes the six different types of criteria that we distinguish: collection relevance, economical relevance, funding, practical criteria, scientific relevance and social relevance. The column for 'subtopic' describes the criteria that we used in the survey, the descriptions are based on the criteria in the last column. It was more feasible to represent 3-8 criteria per question instead of up to 20 criteria to choose from for survey respondents. When the subtopic contains '(new)' the criteria are the result of new information from the survey.

Table 1: A unique set of criteria. Based on the first set of criteria, but standardized, double entries merged and categorized.

Topic	Subtopic	Criteria
Collection relevance	(new)	Does it contribute to ensure continuity for digitisation processes?
Collection relevance	(new)	Is collection being used or worked on?
Collection relevance	Align with opportunity (moving or rehousing)	Is digitization completed in conjunction with moving or rehousing collections
Collection relevance	Attract users and promote usage	Does it contribute to the usage of collection
Collection relevance	Attract users and promote usage	Does it help to identify the end users and usages of the collection and its digital assets
Collection relevance	Attract users and promote usage	Is it relevant for the collection users of the collection

Collection relevance	Capture data important specimens (historic, fragile, types)	Does it involve objects at high risk of losing their integrity
Collection relevance	Capture data important specimens (historic, fragile, types)	Does it involve objects at risk of being lost to the community through sale
Collection relevance	Capture data important specimens (historic, fragile, types)	Does it involve objects of international significance
Collection relevance	Capture data important specimens (historic, fragile, types)	Does it involve objects of national or cultural significance
Collection relevance	Capture data important specimens (historic, fragile, types)	Is the collection to be digitised of historical value
Collection relevance	Exploit existing expertise, make use of available processes	Can other collections be involved in the same digitisation process
Collection relevance	Exploit existing expertise, make use of available processes	Can other collections use the same digitisation process
Collection relevance	Exploit existing expertise, make use of available processes	Does it capture irreplaceable knowledge of employees
Collection relevance	Exploit existing expertise, make use of available processes	Is it a follow up of already finished digitised collection
Collection relevance	Exploit existing expertise, make use of available processes	Is more than one party involved in the digitisation process

Collection relevance	Exploit existing expertise, make use of available processes	Is the required collection expertise available
Collection relevance	Implement policy	Is it in line with the institution priority / policy / mission
Collection relevance	Improve efficiency of collection management	Does it contribute to the accessibility, conservation and storage of collection
Collection relevance	Improve efficiency of collection management	Does it contribute to the digitisation infrastructure
Collection relevance	Improve efficiency of collection management	Does it contribute to the disclosure and dissemination of previously hidden collection data
Collection relevance	Improve efficiency of collection management	Does it contribute to the possibility of (re)discovering high value specimens
Collection relevance	Improve efficiency of collection management	Does it facilitate the creation of digitally available collection data
Collection relevance	Improve efficiency of collection management	Is digitization occuring following standardized methods set by the wider community
Collection relevance	Improve efficiency of collection management	Is it relevant for the collection manager(s) in charge of the collection
Collection relevance	Improve efficiency of collection management	Start digitizing objects that are related to the natural history collections

Collection relevance	Reduce handling of specimens	Does it reduce future transcription time of specimen labels
Collection relevance	Reduce handling of specimens	Does it reduce the physical handling of (high value) specimens
Economical relevance	Create jobs	Does it contribute to the creation of jobs
Economical relevance	Generate revenues	Does digitization make it possible to share collection materials in multiple venues to generate revenue
Economical relevance	Relevant for economic activities (indicator species, invasive species)	Is it relevant for understanding invasive alien species
Economical relevance	Relevant for economic activities (indicator species, invasive species) and imperatives	Does it target the most urgent economical imperatives of our time
Economical relevance	Relevant for economic activities (indicator species, invasive species) and imperatives	Is it relevant for economic well-being
Economical relevance	Relevant for economic activities (indicator species, invasive species) and imperatives	Is it relevant for indicators of minerals
Funding	(new)	Are data management, data processing and data storage (incl. long-term storage) costs covered?
Funding	Costs	Are costs of requirements covered

Funding	Costs	Are exploitation costs results covered
Funding	Costs	What are the expected costs
Funding	Feasibility	Is it feasible within the expected costs
Funding	Funding opportunities	Are there additional fundings involved
Funding	Funding opportunities	Does it increase funding/grant opportunities
Funding	Funding opportunities	Is it possible to receive funding in different ways for this collection
Practical criteria	(new)	Are the necessary information technology and data platforms available?
Practical criteria	(new)	Are the required facilities in place / available?
Practical criteria	(new)	Is (rich) metadata available?
Practical criteria	Accessibility	Are all objects of the same museum / institute
Practical criteria	Accessibility	Are objects accessible and findable in the collection
Practical criteria	Accessibility	Is there a location with the necessary location conditions available
Practical criteria	Detail	Are the objects for mass digitisation or digitisation on demand

Practical criteria	Detail	Is the collection already at some level digitised
Practical criteria	Detail	What are the project goals
Practical criteria	Detail	What is the current condition of the specimens
Practical criteria	Detail	What level of data quality will be captured
Practical criteria	Procedure	Are the requirements in place / available
Practical criteria	Procedure	Is imaging process included
Practical criteria	Procedure	What level of digitisation is required?
Practical criteria	Procedure	Are metrics or procedures involved in order to measure progress and check data quality
Practical criteria	Results	Does effort exceeds perceived benefit
Practical criteria	Results	Is data open access available
Practical criteria	Results	Is there a plan for communication on the results of digitization
Practical criteria	Safety	Does objects for digitisation require special treatment and safety regulations
Practical criteria	Size	What is the physical size and structure of the collection
Practical criteria	Size	What is the range of sizes and forms of objects involved

Practical criteria	Size	What is the taxonomical range
Practical criteria	Staff	Is specific training of staff needed
Practical criteria	Staff	Is the necessary staff in-house available
Practical criteria	Staff	Will voluntary workers or staff outsourced be used
Practical criteria	Time	Can the digitization project be split up and spread out over time if necessary
Practical criteria	Time	How much time is needed to database/image one specimen
Practical criteria	Time	Is it feasible in the expected duration
Scientific relevance	Open up new or enable multidisciplinary research	Does it enable the performance of multidisciplinary research or analyses
Scientific relevance	Open up new or enable multidisciplinary research	Does it open up new research avenues or fields
Scientific relevance	Open up new or enable multidisciplinary research	Does it stimulate new or improve scientific publications
Scientific relevance	Other	Is the collection directly or indirectly necessary for research
Scientific relevance	Other	Is the collection to be digitised of scientific value

Scientific relevance	Other	Will data acquired from digitisation comply with FAIR principles (Findable, Accessible, Interoperable and Reusable)
Scientific relevance	Part of in house research programme	Is it part of a research programme within the institute
Scientific relevance	Part of international research initiative	Does it contribute to a partnership within a larger scientific community effort
Scientific relevance	Part of national research initiative	Does it contribute to a research infrastructure
Scientific relevance	Relevant for access to primary biodiversity data	Is it relevant to the discovery and access to primary biodiversity data
Scientific relevance	Relevant for applied research (e.g invasive species)	Is it relevant for understanding invasive alien species
Scientific relevance	Relevant for fundamental research (taxonomy, ecology)	Is it relevant for ecological and geographical imperatives
Scientific relevance	Relevant for fundamental research (taxonomy, ecology)	Is it relevant for systematic and taxonomical imperatives
Scientific relevance	Relevant for research studying processes/trends (evolution, extinction, climate change)	Does it target the most urgent biodiversity and environmental imperatives of our time
Scientific	Relevant for research studying processes/trends (evolution,	Is it relevant for understanding extinction, threatened or

relevance	extinction, climate change)	endemic species and habitat loss
Social relevance	Data access	Are there sensitivity issues in revealing collection data
Social relevance	Data access	Does collection data needs to be publicly available due to legislation
Social relevance	Data access	Does it increase public access to specimens
Social relevance	Other scientifc domains	Is it relevant for climate change impacts
Social relevance	Other scientifc domains	Is it relevant for ecosystem services
Social relevance	Other scientifc domains	Is it relevant for environmental well-being
Social relevance	Other scientifc domains	Is it relevant for food security and agricultural issues
Social relevance	Other scientifc domains	Is it relevant for human well-being
Social relevance	Other scientifc domains	Is it relevant for monitoring and conservation of biodiversity
Social relevance	Other scientifc domains	Is it relevant for public health issues
Social relevance	Other scientifc domains	Is it relevant for understanding impact of invasive alien species
Social relevance	Other scientific domains	Does it help to create a theme across previously unassociated objects

Social relevance	Other stakeholders	Are external stakeholders involved in the definition and implementation of digitization
Social relevance	Other stakeholders	Are the collections to be digitized associated to museum projects, ongoing or planned
Social relevance	Other stakeholders	Does it address a recognized need or gap within (in)formal education
Social relevance	Other stakeholders	Does it contribute to the recognition of the institute and raise its profile
Social relevance	Other stakeholders	Does it contribute to the use of biodiversity collection data for the development of (conservation) policy
Social relevance	Other stakeholders	Does it increase collaboration with different stakeholders
Social relevance	Other stakeholders	Does it increase the potential of the institute to contribute beyond its traditional domain
Social relevance	Other stakeholders	Does it inspire new approaches within (in)formal education
Social relevance	Other stakeholders	Does it target the most urgent social imperatives of our time
Social relevance	Other stakeholders	Is it of interest to a wider social audience
Social relevance	Other stakeholders	Is it of interest to the museums public

Innovation and consolidation for large scale digitisation of natural heritage

Inventory of current criteria for prioritization of digitization

D2.1 – Appendix 5

Appendix 5. Rewritten set of criteria

Table 1: A rewritten set of criteria. Based on the unique set of criteria and rewritten in a standardized form, leaving out question marks and open-ends.

Topic	Criteria
Collection relevance	Aligns with the institution priority / policy / mission
Collection relevance	Availability of required collection expertise
Collection relevance	Capture irreplaceable (historical) knowledge of employees
Collection relevance	Collection is being used or worked on
Collection relevance	Contribute to the (re)discovery of high value specimens
Collection relevance	Contribute to the accessibility, conservation and storage of collection
Collection relevance	Contribute to the continuity of digitisation processes
Collection relevance	Contribute to the digitisation infrastructure
Collection relevance	Contribute to the disclosure of previously hidden collection data
Collection relevance	Contribute to the usage of collection
Collection relevance	Digitise collection of historical value

Collection relevance	Digitise following standardized methods set by the wider community
Collection relevance	Digitise in conjunction with moving or rehousing collections
Collection relevance	Digitise objects related to natural history collections (e.g. journals, maps)
Collection relevance	Facilitate the creation of digitally available collection data
Collection relevance	Follow up on already finished digitised collection
Collection relevance	Help to identify the end users and usages of the collection and its digital assets
Collection relevance	Involve more than one party in the digitisation process
Collection relevance	Involve objects at high risk of losing their integrity
Collection relevance	Involve objects at risk of being lost to the community through sale
Collection relevance	Involve objects of international significance (e.g type specimens)
Collection relevance	Involve objects of national significance (e.g. national natural history)
Collection relevance	Other collections can be involved in the same digitisation process
Collection relevance	Other collections can use the same digitisation process
Collection relevance	Reduce future transcription time of specimen labels

Collection relevance	Reduce the physical handling of (high value) specimens
Collection relevance	Relevance for the collection manager(s) in charge of the collection
Collection relevance	Relevance for the collection users of the collection
Economical relevance	Contribute to the creation of jobs
Economical relevance	Facilitate sharing collection materials between multiple venues to generate revenue
Economical relevance	Involve external stakeholders in the definition and implementation of digitisation
Economical relevance	Relevance for economic well-being
Economical relevance	Relevance for identifying (plant) indicators of valuable minerals
Economical relevance	Target the most urgent economical imperatives of our time
Funding	Cover costs of data management, data processing and data storage (incl. long-term storage)
Funding	Cover costs of requirements
Funding	Cover exploitation costs

Funding	Feasible within the expected costs
Funding	Increase of funding and/or grant opportunities
Funding	Indicate the expected costs
Funding	Involve additional fundings in addition to in-house funding
Funding	Receive additional funding in different ways for this collection
Practical criteria	Availability of a location with the necessary conditions (e.g. light)
Practical criteria	Availability of digitisation requirements (e.g. procedures, materials)
Practical criteria	Availability of (rich) metadata
Practical criteria	Collection already at some level digitised
Practical criteria	Effort exceeds perceived benefit
Practical criteria	Feasibility within the expected duration
Practical criteria	In-house availability of necessary staff
Practical criteria	Include imaging process within digitisation project
Practical criteria	Indicate needed time to digitise one specimen
Practical criteria	Indicate the current condition of the specimens

Practical criteria	Indicate the digitisation project goals (e.g. number of specimens)
Practical criteria	Indicate the level of data quality to be captured
Practical criteria	Indicate the physical size and structure of the collection (e.g. specimens, drawers)
Practical criteria	Indicate the range of sizes and forms of objects involved
Practical criteria	Indicate the taxonomical range
Practical criteria	Involve metrics or procedures in order to measure progress
Practical criteria	Need for specific training of staff
Practical criteria	Objects are accessible and findable in the collection
Practical criteria	Objects are present at the same institute
Practical criteria	Objects are suitable for mass digitisation or digitisation on demand
Practical criteria	Objects require special treatment and safety regulations
Practical criteria	Open access availability of data
Practical criteria	Possibility to split up and spread out a digitization project over time
Practical criteria	Presence of a communication plan on digitisation results

Practical criteria	Presence of required facilities
Practical criteria	Presence of required information technology and data platforms
Practical criteria	Use of voluntary workers or outsourced staff
Scientific relevance	Compliance of acquired data with FAIR principles (Findable, Accessible, Interoperable and Reusable)
Scientific relevance	Contribute to a partnership within a larger scientific community effort
Scientific relevance	Contribute to a research infrastructure
Scientific relevance	Enable the performance of multidisciplinary research or analyses
Scientific relevance	Is the collection directly or indirectly necessary for research
Scientific relevance	Open up new research avenues or fields
Scientific relevance	Part of a research programme within the institute
Scientific relevance	Relevance for ecological and geographical imperatives
Scientific relevance	Relevance for systematic and taxonomical imperatives
Scientific relevance	Relevance for understanding extinction, threatened or endemic species and habitat loss
Scientific relevance	Relevance for understanding invasive alien species

Scientific relevance	Relevant to the discovery and access to primary biodiversity data
Scientific relevance	Scientific value of collection to be digitised
Scientific relevance	Stimulate new or improve scientific publications
Scientific relevance	Target the most urgent biodiversity and environmental imperatives of our time
Social relevance	Address a recognized need or gap within (in)formal education
Social relevance	Associate with ongoing or planned museum projects (e.g. exhibitions)
Social relevance	Contribute to the recognition of the institute and raise its profile
Social relevance	Contribute to the use of data for the development of (conservation) policy
Social relevance	Help to create a theme across previously unassociated objects
Social relevance	Increase collaboration with different stakeholders (e.g. distant research)
Social relevance	Increase public access to specimens
Social relevance	Increase the potential of the institute to contribute beyond its traditional domain
Social relevance	Inspire new approaches within (in)formal education
Social relevance	Need of public availability of data due to legislation

Social relevance	Presence of sensitivity issues in revealing collection data (e.g. location of red list species)
Social relevance	Relevance for climate change impacts
Social relevance	Relevance for ecosystem services (e.g. water filtering)
Social relevance	Relevance for environmental well-being
Social relevance	Relevance for food security and agricultural issues (e.g. crops and pests)
Social relevance	Relevance for human well-being
Social relevance	Relevance for monitoring and conservation of biodiversity
Social relevance	Relevance for public health issues (e.g zoonotic diseases and their vectors)
Social relevance	Relevance for understanding impact of invasive alien species
Social relevance	Relevance to the museums public
Social relevance	Relevant to a wider social audience
Social relevance	Target the most urgent social imperatives of our time