ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Digitisation of private collections

DELIVERABLE D5.4

Luc Willemse (Naturalis)

Veljo Runnel (UTARTU)

Hannu Saarenmaa (LUOMUS)

Ana Casino (CETAF)

Karsten Gödderz (CETAF)

ICEDIG.EU

Summary

In this deliverable (D5.4), we report on solutions and procedures to incorporate private collections into the international data infrastructure. Central in this task were pilot projects carried out in three countries to test ideas on how to best motivate and equip citizen collectors for digitisation:

1. In Estonia, the approach was to outline the tools for registering, digitising and publishing private collection data, available in the biodiversity data management system PlutoF, developed and maintained by the University of Tartu. During the second stage, private collectors were contacted through the network of naturalist associations and also via social media. Private collectors who indicated interest in digitising their collections where assisted in the process. Six private entomological collections have been digitised and published to the GBIF portal on the metadata or specimen data level.

2. In Finland, the FinBIF portal offers a popular Notebook Service for citizens to store observations. This has been expanded to include collection specimens related to a field gathering event. Central to this new capacity is the ability to print labels with unique identifiers for each digitised specimen. The expanded system is now being tested by citizen collectors. Further functionality to be introduced includes a specific form geared towards digitising entire drawers containing specimens from several gathering events.

3. In the Netherlands, using dedicated software, private collection owners were approached directly and asked to start digitising their collection either by themselves or with the help of volunteers who were recruited specifically for this task. The first results indicate that volunteers clearly are more productive in terms of number of specimens digitised.

In addition to comparing management tools available for digitising private collections the pilot also looked into aspects like motivation, the person undertaking the work, scope, planning, specific knowledge or skills required and the platform for online publication. The pilot also revealed some underexposed aspects linked to private collections that may influence their digitisation like future ownership, the legality of specimens residing in private collections and the use of unique identifiers to attach to specimens. Besides looking for ways to streamline the overall process of digitising private collections and dealing with local, national or international challenges, developing a communication strategy is considered crucial in order to effectively distribute information and keep private collection owners aware of ongoing developments.

Besides collection owners other stakeholders are involved in and extremely important for digitising and publishing data from private collections. The stakeholders and their roles

are identified and for each of them a roadmap is outlined aimed at further streamlining the data from private collections into the international infrastructure.

In conclusion recommendations are presented based on challenges encountered during this task that are considered important to really make significant progress towards the overall accessibility of data stored in privately held collections.

ICEDIG.EU

Table of contents

1. Definition and scope

1.1 Digitisation

Digitisation is a wide concept that, depending on the person, context or situation, may have different meanings (see also p. 3 in Cobb et al., 2019[2]). Generally, digitisation is the process of creating a virtual representation of physical objects. When considering collection digitisation, however, this can relate to the electronic registration of the basic information of an entire collection, it may relate to registering information about the storage units (box, container, jar) or species present in a collection or it can mean electronically storing the label data of individual specimens. Using the term digitisation in some cases explicitly refers to making digital images whereas in others it is restricted solely to data. When using digitisation in combination with registering data, it may include all the data linked to a specimen or only part of the data[1]. By using the term 'digitisation' one also cannot assume that this also includes validation or georeferencing. In some cases this may apply but in many others it may not. For communication purposes it is crucial that both the sender of a message with the term 'digitisation' and the receiver of that message attach exactly the same meaning to the term 'digitisation'. If there is a mismatch between sender and receiver this may result in a misunderstanding which may lead to frustration, resistance and ultimately result in unwillingness to undertake digitisation tasks together from both sides.

In the current deliverable the term "digitisation" refers to the electronic registration of data and information from individual specimens or samples, possibly but not necessarily including imaging, validating of digitisation results or georeferencing, etc.

1.2 Private collections

In the original proposal, Task 5.3 was aimed at natural history collections of private collectors, amateur societies, and smaller museums and herbaria. This was in line with the precursor to this task within ICEDIG, Task 2.2, which was formulated as follows: "Inventory of content and incentives for digitisation of small and private collections." However, straight from the onset of the implementation of Task 2.2 it was decided to exclude "small" collections as a separate entity and restrict the target group of Task 2.2 to private collections only. The main argument for this adjustment is that a distinction based on size is rather artificial and "small" collections are therefore difficult to address as an entity. As privately owned collections outnumber collections of amateur societies and smaller museums and herbaria

---

[1] Cf., the discussion on Minimum Information on Digital Specimens (MIDS) in "Open access implementation guidelines for DiSSCo", ICEDIG Deliverable Report D6.5. https://doi.org/10.5281/zenodo.3465285

[2] Cobb NS, Gall LF, Zaspel JM, Dowdy NJ, McCabe LM, Kawahara AY. 2019. Assessment of North American arthropod collections: prospects and challenges for addressing biodiversity research. PeerJ 7:e8086 https://doi.org/10.7717/peerj.8086

many times and by their very nature form an easily distinguishable group, a distinction based on ownership therefore offers a clearer and more practical boundary than collection size. The distinction made for Task 2.2 has also been adopted for Task 5.3.

The current task is therefore aimed at those collections that are privately owned, usually kept in a residential building and managed and cared for by a private person who does not receive any funding for this. Private collections on average are relatively small. Results from a survey carried out across Europe in 2018 and presented in ICEDIG deliverable report D2.2 [2] showed that most private collections hold less than 10,000 specimens, some 25% having more than 10,000 specimens. Private collections with more than 100,000 objects are exceptional but still form some 5% of the totalof private collections around. The 2018 European survey showed that the total number of specimens owned by private collectors, who responded(!) to the questionnaire lies between 9 and 31 million. This is only the tip of the iceberg, the overall total number of specimens in all European private collections is many times higher, as not all the collectors were reached, only a small percentage that were reached responded and there were big differences on how the various countries were covered. Based on the survey results it is not possible to make a reliable guesstimate of the total number but quite likely this surpasses 100 million or more. In comparison, institutional natural history collections in Europe hold 1.5 billion specimens.

65% of the respondents of the survey said that they manage their collection data fully or partly electronically. Over 90% are interested in sharing their data some way, preferably through a public website, and by listing metadata of their collection in some kind of public register. Most private collection owners (55%) indicated they need tools, such as a dataset template or a web based digitisation platform, followed by guidelines (36%) and physical equipment (27%). The report concludes that future efforts to help private collection owners to digitise their collections should focus on providing appropriate online tools and information on how to get started with digitisation. In particular, there is a large interest in a new European-wide website where private collection owners can register and subsequently share collection-level data.

---

[2] Egmond E & al. (2019) Prioritising scientific and societal needs for data of private collections. ICEDIG deliverable report D2.2. https://doi.org/10.5281/zenodo.2582994

2. Introduction

Specimens held in private natural history collections form an essential, but often neglected part of the specimens held in natural history collections worldwide. When engaging in regional, national or international initiatives aimed at increasing the accessibility of biodiversity data, it is paramount to include private collections as much and as often as possible. Compared to larger collections in national history institutions, private collections are numerous, anonymous, small and diverse in all aspects of collection management. This poses a unique set of challenges to take private collections into account.

Collection owners in general have neither biodiversity informatics knowledge nor the resources to digitise and share their collection data. In the current task (Task 5.3) pilot projects were carried out to look into aspects like strategies and the role of stakeholders as well as into management systems that can be used to digitise privately owned collections. The results, experiences and lessons learned from these pilots are described in this document.

Prior to the current deliverable, no less than five milestone reports were prepared within this task. The first milestone (MS29) centered on the role of citizen associations on digitisation of private collections, followed by two milestones on guidelines (MS30) and the status (MS31) of digitisation of private collections. The first preliminary results for the pilot projects were presented in MS32. The fifth milestone (MS33), reported on the progress within this task and the pilot projects up to and including June 2019 and included the work plan for the next six months.

Taking into account regional and cultural differences across Europe, the pilot projects not only served to gain insights into management systems that are available and practical obstacles that have to be tackled but also into factors beyond the direct influence of collection owners that interfere with digitising or sharing data.

Results laid out in this deliverable provide a basis for future actions to be taken aimed at improving digitisation of private collections and the inclusion of their data in the international infrastructure. First, in chapter 3, results of two approaches used in Estonia and the Netherlands to get private collection owners involved in digitising their collections are discussed. Partly based on the results obtained in Estonia and the Netherlands, chapter 4 essentially defines motivation, factors affecting it, steps to optimise motivation and the stakeholders involved to do so. Out of a multitude of factors affecting motivation, the (choice of a) management tool for digitising collection data is an important one. Chapter 5 combines management tools in four categories - three of which were used in pilots within the current task - and describes the pros and cons for each category. The next chapter introduces seven stakeholder groups involved in or having an interest in the digitisation of privately held collections. Chapter 7 provides roadmaps to be taken by stakeholders describing in practical terms what actions are recommended to facilitate digitisation of private collections. Before

the concluding remarks a separate chapter presents recommendations that require attention in order to start implementing measures to further facilitate the digitisation of private collection owners.

3. Involving private collection owners

Taking into account regional and cultural differences across Europe, two pilots were carried out in Estonia and the Netherlands to gain some insight into approaches toward collection owners to get them involved into digitising and/or sharing collection data. The two approaches used in the pilots and their results are described in this chapter. In Finland, the third pilot focussed purely on adjusting an online web portal to support digitisation by citizen collectors. A brief overview and discussion of data management systems including the systems used for the pilots in Finland, Estonia and the Netherlands is presented separately in Chapter 5.

3.1 Estonia

The main goal was to reach out to the private collection owners, present them the idea of collection digitisation along with the different strategies and means of digitisation. As a mediator, ideally, a collection owners association was to be approached. After the first contact with this association, testing would start where collection owners would digitise their collections through the PlutoF workbench and publish the data to GBIF. This process would be documented and analysed and finally presented as a best practice case for a workflow of private collection digitisation and publishing of collection data and metadata to GBIF.

UTARTU collaborated with the Estonian Lepidopterists' Society (ELS) to promote the concept of private collection digitisation and data sharing. The contract was signed between UTARTU and ELS for organising a public workshop and publishing an article in the ELS yearly magazine for promoting digitisation and giving guidelines to ELS members and the wider public. The Workshop was held and the article published in winter 2019 with twelve ELS members and other public participants.

Private collection owners who showed interest in the digitisation of their collections were assisted with their digitisation efforts. To this end, a specialist assistant was contracted by UTARTU. For some collection owners, access to the digital photography facilities in UTARTU were enabled to image the collection specimens.

3.2 The Netherlands

Prior to the ICEDIG project, in early 2017, a questionnaire about the composition of collections and their digitisation status was sent to members of the Dutch Entomological Society (NEV[3]). As part of the pilot in Task 5.3, a number of private collection owners who responded to the 2017 survey and had not yet started digitising their collections were approached and subsequently visited at their home address where the collection was kept.

---

[3] https://www.nev.nl/

During this meeting the collection was viewed, motivation probed and digitisation (specimen data registration only) of their collection discussed in terms of what was needed to get started.

Starting early 2019 nine collection owners joined the pilot. Out of the nine collection owners five indicated they preferred to do the digitisation themselves whereas four collection owners indicated they lacked the time but agreed to let volunteers do the work. For the latter group three collections would be digitised at an institute (instead of the home address) while in one instance the collection owner agreed to let a volunteer come and do the work at his home address. For collection owners doing the work themselves, all required items (manual, template file for data entry, registration codes) were prepared and handed out during an introductory session. Collection owners were then very much left alone, having been given contact details in case of questions.

Volunteers with no or hardly any experience in entomology were recruited using social media and were given a short instruction session. Each volunteer, as a rule, only worked for a single day per week. Altogether 7 volunteers were deployed, one working at the home address of a collection owner and 6 volunteers working at the institute. For the latter group using 1-2 workplaces, collections were brought to the institute, digitised and returned to the owner.

Table 1 summarises the results in numbers of specimens digitised per collection. Results clearly indicate that the output in terms of number of specimens digitised tend to be larger when volunteers are involved.

Main group	Subgroup	Estimated number of specimens	Volunteers deployed	Start date	Number of specimens digitised by 30/11/2019
Insecta	Diptera	1000	no	09/05/2019	0
Insecta	Lepidoptera	1600	no	18/01/2019	1
Insecta	Coleoptera	25000	yes	30/01/2019	3200
Insecta	Coleoptera	10000	yes*	14/01/2019	800
Insecta	Microlepidoptera	6000	no	18/04/2019	0
Insecta	Lepidoptera	10000	no	17/01/2019	215
Insecta	Heteroptera (spirit)	15.000	yes	22/02/2019	65
Insecta	Lepidoptera	17500	yes	15/02/2019	5000
Insecta	Lepidoptera	17000	no	17/01/2019	3240

* A volunteer digitised at the home of the collection owner.
Table 1: Progress in number of specimens digitised for each collection

In addition to the digitisation efforts, a one day meeting was organised for members of the Dutch Entomological Society. Among other reasons, this was organised to inform participants about the pilot that had taken place, the lessons learned, remaining questions, future plans and feedback. During the day a hands-on training on specimen digitisation was

included and several aspects (imaging, online portals, management tools) demonstrated. Ca. 25 participants attended the meeting.

3.3 Comparison and lessons learned

The diversity of private collections, their owners and the cultural aspects linked to private collections across Europe is enormous. In this respect the two approaches to get private collection owners involved in digitisation carried out in Estonia and the Netherlands only scraped the surface of a multitude of possible approaches. This diversity makes a comparison between approaches in terms of effectiveness difficult. An approach that works very well for amateur entomologists with private collections in Estonia or the Netherlands may not work at all in other countries. Advocating an approach that works very well for most entomological collections across Europe may turn out to be a complete failure when applied to malacological collections.

During the pilot in Estonia, a workflow was set up and tested for managing and publishing the digitised collection data, using PlutoF biodiversity data platform, GBIF IPT and GBIF API. The workflow proved to be suitable for intended use as a tool for private collection digitisation. However, the willingness of private collection owners to share the data was quite low and only a small number of them actually published the data. There are no clear explanations for this, but as indicated during meetings with the lepidopterologists society, collectors consider digitisation a time consuming activity which is not seen as something they are interested in in the first place.

The approach used in the Netherlands including nine participants showed that conviction of the importance of digitisation and having been given the tools to start digitising do not always provide enough impetus to get and keep collection owners going in the medium and long-term. Lack of time or rather making choices on how to spend your time hampers the start and/or continuation of digitisation related activities. Within the Dutch cultural setting the use of volunteers provided a workable solution to overcome time constraints. Along the way various remarks were made or questions raised about specific aspects of digitisation which were not taken into account at the start but have been incorporated in this deliverable. Such questions and comments very much helped in realising that making digitisation of private collections a success can only be brought about by a strong commitment and a joint effort from all stakeholders involved.

4. Factors affecting motivation

There are many factors that influence the motivation of a private collection owner toward digitising his or her collection. Motivation is defined here both in the positive sense as awareness and conviction about the importance of digitisation and publishing data and in the negative sense as factors that cause reservation or doubt toward various aspects of digitisation which may hamper or prevent collection owners from starting or continuing digitisation. Taking this into account, a communication strategy needs to be developed which allows each and every collection owner to be approached with a tailored message. Chapter 4 in the ICEDIG deliverable D2.2 which centers on a communication strategy starts as follows:

"In order to increase the degree of digitisation in private collections, collection owners need to be aware of the fact that digitising collection specimens and sharing data online is important for the global research community and contributes considerably to widening and increasing the use and general interest of collections. Being aware of this is a critical first step, but ideally collection owners should be entirely convinced of the necessity of digitisation. It will furthermore be equally important to clearly indicate how private collection owners can benefit from digitising and sharing their collection themselves, as private collection digitisation cannot solely rely on intrinsic, idealistic motivation of individuals (e.g. performing digitisation to help advance science). Collection owners who are convinced of the mutual benefits of digitisation are more likely to look for ways to actually begin with the digitisation process. The digitisation process itself could be started by the collection owners themselves. They may seek help from organisations (museums, institutes) or initiatives (DiSSCo) or they may become involved in national or international projects. A communication strategy to engage them should zoom in on both aspects: raise awareness and offer guidance for the process. This communication strategy is intended to serve as a guideline that can be used to encourage and motivate private collection owners to participate in Europe's overall digitisation efforts. In particular, this strategy aims to indicate how communication with private collection owners can be used to:

- Increase awareness of the need for digitisation of natural history collections among private collection owners, including private collections and sharing the digitised information.
- Encourage private collection owners to take action and join current and future digitisation activities at institutional, national and international levels.
- Make private collection owners aware about existing knowledge and experience related to digitising collections and the fact that advice and assistance related to the digitisation process can be provided in many forms, including solutions for a lack of time."

Regarding the first two bullet points, actions to be taken pertain to organising a range of messages and channels of communication toward private collection owners aimed at increasing awareness and encouragement. This, in essence, consists of adapting a similar kind of message in a multitude of ways to fit the audience it is aimed at and finding an optimal form or setting to spread it.

The third bullet point pertains to private collection owners who, rather than needing to be convinced about the importance of digitising their collections, are hesitant to start due to lack of knowledge, experience or time to do it or still have questions regarding specific aspects related to digitising them. The answers and activities required to deal with this involve various stakeholders and need to be handled at a local, national or international scale. Part of the lack of information and challenges interfering with collection digitisation were tackled within task 5.3 as part of the pilot, while others only became evident along the way.

Various aspects are briefly described below, indicating the approach required and stakeholders involved. The approaches described here (in random order) are partly repeated in Chapter 7 which presents separate roadmaps for each of the stakeholders.

1. Digitisation strategy
- Description: For collection owners who like to start but have not done so, choosing a strategy that best fits their collection and personal preferences is the first step to take. Elements included in a strategy are scope (what will be digitised, in which order?), who (who will do the actual work?), how (what do I need in terms of hard- and software, skills, knowledge, etc.?) and when (the time factor; how long will it take; how much time is invested?). For collection owners who have already started digitising their collection, a strategy can include contacting the museum the owner wishes to donate the collection to in the future and assess if and how digitisation should be adapted to match the protocol used by the museum.
- Approach: A general protocol offering guidance in the process of deciding on the best digitisation strategy could be an important means to assist collection owners in their decision.
- Stakeholders and their roles:
    - Private Collection Owner: user of the protocol
    - Research Community (e.g. CETAF): initiate protocol at an international level
    - Collection Holding Institutions: initiate protocol at a regional or national level
    - Non-Governmental Organisations: communication about protocol
    - Public Administrations: stress importance; stimulate development

2. Donation and digitisation
- Description: Most private collections sooner or later end up in a larger regional or national repository. Future ownership can already be taken into account when digitising collections, for instance for applying unique registration codes and digitisation standards.
- Approach: Each regional or national museum should have a protocol describing standards, formats and use of registration codes for datasets and collections that are going to be donated so that it can inform private collection owners. If the protocol is followed, regional or national repositories should be able to integrate an external dataset with the least possible effort into their own management system and

collection. Another aspect that requires attention is the provenance of collection specimens in an era when collecting is becoming increasingly bound to laws and regulations.

- Stakeholders and their roles:
  - Private Collection Owner: user of the protocol
  - Collection Holding Institutions: prepare protocol
  - Non-Governmental Organisations: communication about protocol (workshop; meetings)
  - Public Administrations: together with collection holding institutes formalise the acceptance of collections regardless of species and provenance, accepting the fact that private collection owners may not always have strictly adhered to laws and regulations when collecting.

3. Online publication
- Description: Digitisation is the first step necessary to publish data and/or images online and make them globally accessible. Like the process of digitisation, online publishing leads to a number of questions. Questions posed during the pilot relate to:
  - ownership of digital data/images
  - the portal
  - timing of online publication
  - access policy
  - privacy
  - sensitive data
  - synchronisation of digital (online) and physical information

Having adequate answers to all these questions will help collection owners to have a better understanding of the whole process, which will help to take away some of the reservations they may still have.
- Approach: Stakeholders should tackle these questions nationally and where appropriate internationally and summarise the answers in an easily readable document.
- Stakeholders and their roles:
  - Private Collection Owner: in need of clarity and answers
  - Non-Governmental Organisations: organise communication with other stakeholders
  - Collection Holding Institutions: practical issues like synchronisation and timing of publication
  - Research Community: look into international issues
  - Data aggregators: provide answers and guidelines on ownership, access policy, privacy and sensitive data

4. Data management tools

- Description: A very important choice collection owners have to make is the data management tool, module or software they wish to use for the registration of their collection data. There are many possibilities, offline and online, dedicated and non-dedicated (see also Chapter 5).

- Approach: information should be available online for collection owners, describing available solutions in general terms when it comes to data management tools, modules or software with the pros and cons.

- Stakeholders and their roles:
    - Private Collection Owner: in need of information
    - Non-Governmental Organisations: communication about websites with information
    - Collection Holding Institutions, Data Aggregators and/or publishers and Research Community should join forces to prepare an international overview of available solutions that is easy to read and interpret for laymen.

5. Alternatives for digitisation by the collection owner

- Description: Digitisation, even without imaging, is a time consuming activity. A collection owner may decide not to spend time digitising because of the sheer size of the collection or the fact that preference is given to spend time on other activities like research. These are perfectly valid arguments. However, there may be alternative solutions like the deployment of volunteers. In task 5.3 the deployment of volunteers was tested in the Netherlands.

- Approach: Due to cultural differences, deployment of volunteers to digitise private collections may differ per country. Therefore a national approach seems most appropriate. National stakeholders should examine whether collection owners are willing to allow volunteers to digitise their collection, as well as whether finding and organising volunteers (and workspace) is a viable option.

- Stakeholders and their roles:
    - Private Collection Owner: prepared to allow volunteers to digitise the collection
    - Collection Holding Institution: prepared to allocate staff time and workspace to organise volunteers to digitise private collections
    - Non-Governmental Organisations: in consultation with collection holding institutes, communicate about the possibility to let volunteers digitise collections
    - Research Community: examine the possibility to internationally tackle this issue
    - Citizen Science Associations: in consultation with collection holding institutes and non-governmental organisations investigate whether or not they can play a role providing manpower for digitisation

6. Manuals and protocols
- Description: Data registration, imaging, data/image quality control or georeferencing are all activities which initially are not the core interest of private collection owners. Providing manuals and protocols to assist in these activities will not only help private collection owners but also form a very strong tool when it comes to standardising ongoing work.
- Approach: Prepare manuals and protocols aiming for private collection owners as the target audience, based on a national or international inventory of available manuals and protocols linked to all aspects included in the concept of digitisation,.
- Stakeholders and their roles:
  - Private Collection Owner: beneficiary of manuals and protocols
  - Collection Holding Institutes, Research Community and Data Aggregators and/or Publishers together look into the best approach to tackle this in order to standardise this as much as possible at an international level taking into account the multilingual challenge.
  - Non-Governmental Organisations: keep members informed about developments and results

7. Stakeholders and their roles
- Description: Chapter 6 describes the stakeholder communities linked to the digitisation of private collections. In the process of collection digitisation, the owners (and other stakeholders for that matter) may benefit from a document summarising the stakeholders and the role they play around the digitisation of their collection.
- Approach: A document needs to be drawn up giving a clear overview of stakeholders, their responsibilities and how and where they play a role in the process of digitising and online publication of data originating from private collections.
- Stakeholders:
  - Research Community (e.g. CETAF) should take the lead

5. Data management tools

An important choice for private collection owners to make when starting to digitise natural history collections is which data management tool to use. Considerations for making a choice are diverse and depend of course partially on personal preferences. Important aspects to be included, apart from obvious matters like costs, support or user friendliness, are:

- generic or dedicated
- flexibility
- export formats
- use of thesauri
- built in tools (label printing; maps)
- learning curve
- linking media

The amount of background information on data management tools is huge. An overview is available at, amongst other sources, iDigBio, TDWG and GBIF where more information can be found about data management tools. None of the currently available solutions for data management tools has proven to be outstanding in its own right, is used globally and has a very satisfied user community. Consequently, the number of data management tools available is still enormous and providing an integral overview is beyond the scope of this document. For the sake of clarity though, data management tools in this document have been organised into four categories which are discussed in some detail below. With the exclusion of dedicated offline software, three of the four categories have been used during the pilot in task 5.3 tools belonging to.

5.1 Online portals for observation data
5.1.1 Description and examples
In quite a number of countries, websites have been built that allow capturing observation data with (or without) the possibility to attach images and usually using input of specialists to validate identifications. Examples of such websites are: iNaturalist (USA), Observation.org (Netherlands), Biodiversidad Virtual (Spain), Artportalen (Sweden and Norway), and iSpot (UK and Ireland). Although they are meant for observation data, they could be expanded to include collected material as well. In the current pilot this was investigated for the FinBIF (Finland) portal.

5.1.2 Results of the pilot
The FinBIF project offers a Notebook Service for citizen scientists to report their biodiversity observations. The service is tri-lingual and quite popular. For instance, for the

year 2018 a total of 212,490 records of Lepidoptera have been reported so far. The service now contains in total over 4 million Lepidoptera records created by citizen scientists worldwide. Most of these records are observations without a preserved specimen, but there are also 264,000 records of preserved specimens.

It was decided to enhance the system so that preserved specimens could be better marked and counted and to have a unique ID which can be printed onto the specimen labels. Marking the specimens which have been digitised with unique identifiers is critical for preventing them to be digitised again in future, creating duplicate records. Data entry in the Notebook Service happens through customised forms or spreadsheet file uploads. Until now, the only generic form to enter data has been the "Field Trip Report" form which is organised around an observation event (gathering) (Figure 1 and 2). Under one event, observations and specimens of many species (taxa) can be entered.

While this form can be used and easily enhanced to enter data of preserved specimens and print labels for them, it works well only when the observations and specimens belong to one gathering event. In the current design there is no direct relation (one-to-many) between an occurrence record and the specimens gathered as evidence of that occurrence. Instead, there will be several records of the same species without any clear link between them. They will only be grouped together under the gathering event. For instance, if 100 individuals of an insect taxon are observed and only 3 of them collected as preserved specimens, four records need to be created with individual counts 97, 1, 1, and 1, respectively. This is not very elegant, but may work in the case of whole individuals. However, if the specimens represent parts of the same individual (for instance branches of a tree), a concrete link between the records would be necessary.

A specification of work was written in November 2018. Design of the new functionalities is underway by the FinBIF development team. It has been concluded that there are at least two different use cases for the digitisation of specimens: 1) saving a few individuals for evidence during an observation event, and 2) digitising a collection by drawers. Each requires a different approach. Currently, only the first case is supported by the FinBIF Notebook Service. Some enhancements have already been made such as label printing, assigning unique IDs for specimens, and allowing field numbers (collector-given ID) to be saved.
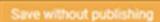
Figure 1. Upper part of the "Field Trip Report" form on the Notebook Service of the FinBIF portal (www.laji.fi), where details of the gathering event can be entered.

Figure 2. Lower part of the "Field Trip Report" form on www.laji.fi, where observations and their details can be entered.

5.1.3 Strengths and challenges

The FinBIF Notebook Service is the result of several generations of developments of Software as a Service (SaaS) for citizen scientists. As early as the 1990's, bulletin boards on the web were used to share records of observations, followed by several (sometimes competing) databases with web interface, etc. The large number of data in the current system is the result of migrating all data from several earlier systems into the Notebook Service. It has then been integrated into the FinBIF portal with data from digitised scientific collections of museums. (If we count these as well, the number of Lepidoptera records is 5.6 million out of which 617,000 are specimens, cf. above figures). This is a valuable pool of data which attracts users to also contribute their own observation and specimen records, further accelerating the accumulation of data.

Another strength is that the system is being hosted and technically developed by the Finnish Museum of Natural History (LUOMUS). Most citizen associations (with the exception

of Birdlife) have endorsed the system, which relieves the associations from the chore of maintaining a complex web database system with their limited resources.

The challenge the system faces, considering its use for private collection digitisation, is its focus on field observations. There are forms for field gathering events only, not for digitising a box of insects organised by taxonomy. Citizen scientists who are serious about their own collection need something that resembles a Collection Management System (CMS) and has views and functions which are focussed on specimens and taxonomy. This is also relevant from the perspective of the museum which will one day inherit the private collection. Such a CMS for private collectors from SaaS is not yet available.

5.2 Online data management platforms
5.2.1 Description and examples

Online data management platforms have been developed with a much wider range of uses in mind, collection management being only one of them. A good example of an online data management platform is PlutoF which during the pilot has been used in Estonia. Other systems with a similar aim include Symbiota, an open source content management system for curating specimen- and observation-based biodiversity data. It has been used to build dozens of portals for various taxonomically and regionally oriented initiatives. More data management platforms and software can be found, for instance on iDigBio.

5.2.2 Results of the pilot

In Estonia, there is little information on private collections which are publicly available. On a very basic level (taxonomic scope, collectors name), some Estonian private collection owners have registered their collections at Global Registry of Scientific Collections (GRSciColl). GRSciColl used to be the only international open registry which also allowed private collection owners to register their collections. Currently the registry is managed by GBIF. The registry is in transitional phase and will be opened for the validated users for submissions and edits.

The University of Tartu Natural History Museum and the Botanical Garden (UTNHM) have developed a biodiversity data management platform called PlutoF which is used nationally by several research institutions but also globally. It also allows individuals to register as users and manage their private observation and collection data. The PlutoF platform is also an official publisher to GBIF and has its own IPT instance installed. Recent developments of the PlutoF platform even allow direct publishing of datasets to GBIF from the system, without IPT. This enables users to share data more easily and with hardly any technical knowledge.

Communication

Private collection owners in Estonia are to some degree organised in non-governmental organisations like the Estonian Lepidopterologists' Society and Estonian Naturalists Society. But those organisations usually consolidate expertise and knowledge on biodiversity, not collection management and digitisation. Nevertheless, they provide a network for people

who share interests. Societies usually have annual meetings for face-to-face communication as well as webpages. For the individuals who are not organised in NGOs but still want to share observations or other biodiversity information, there are many social media groups available, usually on a specific topic like butterflies, bumblebees, mosses, etc.

The aims of the pilot

The pilot project aimed at:

A. reaching out to private collectors (who own natural history collections),
B. presenting the concept of digitisation and data sharing to collection owners,
C. establishing the collaboration between research institution (museum) and private collectors for digitisation attempts,
D. utilising biodiversity data management tools for private collection registration and digitisation, and
E. sharing private collection data through GBIF portal.

The course and results of the pilot

A. In order to maintain focused action, we targeted mainly entomology collection owners but initially reached out to a wider community of naturalists. An official invitation to participate in the pilot project was sent to specific social media groups such as Estonian Butterflies or to e-mail lists of naturalist associations or clubs. We contacted and subcontracted the Estonian Lepidopterists' Society (ELS) to collaborate including their members in the pilot. During this stage, we collected contact information of private collectors who indicated interest to collection digitisation and data sharing.

B. On the UT Natural History Museum website a subpage was created which included a detailed description of the pilot project, its aims and also a tutorial for digitisation. As a part of the subcontracting, ELS facilitated the digitisation campaign at their membership events. Also, a featured article about collection digitisation was included in their journal "Lepinfo".

C. A curator of zoological collections of UTNHM was taking the responsibility of introducing the private collectors who were to digitise their specimens to the photographic facilities of the museum. However, there was only one collector who approached the museum for imaging his collections.

D. Before introducing private collectors to the PlutoF platform, the workflow for digitised collection data management was tested, and some modifications and adjustments were made by the UTNHM IT team. As the data publishing was also an important component of the pilot, the GBIF publishing functionality of PlutoF was tested as well. A specialist was subcontracted with the assignment to instruct the collection owners who were involved in the pilot project. Registration of collections and publishing the data was intended on two levels: metadata-only and full specimen data. For metadata-only publishing two alternatives were used: publishing through IPT and

publishing through the PlutoF-integrated module. In both cases the subcontracted specialist assisted private collectors and was given system access to GBIF IPT for data management.

Collectors were asked to submit their collection metadata via Google Form with following data fields:

- email address
- name of collection owner
- the name of the collection (e.g. "Private herbarium of John Smith")
- the type of collection (different types of collections should be registered separately)
- number of specimens in the collection (estimate)
- country of location
- owner's address
- owner's email (if the contact email differs)
- owner's phone number
- linked webpages (LinkedIn or similar)
- collecting region (Europe, Asia, etc.) or countries (Estonia, Russia, etc.) for most of the specimens
- which groups of organisms are present in the collection? (As detailed as possible or desired: "Lepidoptera", "Cerambycidae", etc. If needed, enter several families, genera, etc.)
- indication of species list in separate file (to be sent by email)
- period of collection of specimens
- in case of paleontological collection, geological period
- number of storage units in collection (boxes, jars, etc.)
- web address of collection (in case the data is already shared online)
- storage type of specimens (pinned, herbaria, alcohol etc)
- indication of intent to digitise at specimen level in the future, so that additional information can be sent
- agreement to publish data on GBIF portal

For full specimen data registration, the collection owner would use the PlutoF taxon occurrence management module (Figure 3).

Figure 3. The PlutoF taxon occurrence management module.

Using the PlutoF platform in the collection digitisation process does not limit it to data publishing only. PlutoF can be used also for collection data management, label publishing, biodiversity data visualisation and analysis.

E. After the metadata submission, the collections were registered on the metadata level either via GBIF IPT or the PlutoF GBIF publishing module. In both cases the subcontracted specialist entered the data and processed publishing to GBIF (Fig. 4). All the required IT solutions were arranged by the museum's IT team.

Figure 4. The PlutoF module for GBIF publishing.

Five collection owners (1, 2, 3, 4, 5) submitted the metadata-only of their collections which was processed by GBIF PlutoF IPT instance or the PlutoF integrated publishing

module. Only one private collector submitted digitised collection <u>data on specimen level</u> (with images included). <u>Guidelines</u> for digitisation of private collections were translated into Estonian and made available on the UT Natural History Museum's webpage.

5.2.3 Strengths and challenges

Sharing private collection data for research in the biodiversity domain is quite rare and technical facilities for this are lacking or not adequate. The PlutoF platform opens up a solution for digitising and publishing digitised data to GBIF. A data management platform can help to keep track of collections. When managing collection data through a platform on the fly as the new specimens come in, the digitisation process is done simultaneously with data management. Only imaging of specimens needs a separate workflow. Data publishing in a meaningful way can increase the value of a private collection. Simple data sharing gives an opportunity to open collections for research. In Estonia, reaching out to private collection owners remains a challenge.

5.3 Generalistic software

5.3.1 Description and examples

Generalistic software is software that can be used to store biodiversity data in a structured way but has <u>not</u> been developed specifically to serve as a data management system for natural history collections. Examples of such software tools are Microsoft Excel, Microsoft Access or FileMaker. Most of such software is used offline but online variants also exist like a Google spreadsheet which for instance has been used in the pilot in the Netherlands.

5.3.2 Results of the pilot

In order to keep the threshold as low and the learning curve as flat as possible - i.e. the data management tool not being reason to not to start digitising - a Google Spreadsheet in a predefined format was used in the task 5.3 pilot in the Netherlands. Altogether, 5 collection owners volunteered to start digitising their collection themselves, including the use of QR codes for registration numbers, and using a Google spreadsheet with an accompanying manual. The format of spreadsheets was developed in such a way that, once delivered to the museum to which the collection would eventually be donated, the museum could easily upload the file to their central data management system. Unfortunately, over the course of six months, only one out of five collection owners actually registered a substantial number of specimens (> 3000). A second one registered a few hundred records, and the three others not more than a few. For the latter group, lack of time rather than the use of a Google spreadsheet was stated as the underlying reason for having registered so few records.

5.3.3 Strengths and challenges

The strengths of generalistic software is that these are readily available, as a rule do not have a steep learning curve, and are flexible when it comes to structuring data. Weaknesses of generalistic software is that carrying out specific tasks like printing a label or producing a map is not possible or requires a lot of detailed knowledge. Uploading data is another issue. Because generic software normally are used without a thesaurus providing standardised lookup tables, data quality may as well be an issue. Furthermore, there is no built-in support for generating unique identifiers to attach to the digitised specimens.

5.4 Dedicated software

5.4.1 Description and examples

Dedicated software refers to management systems specifically built to support ongoing activities in natural history collections. Examples are for instance widely used programs that are installed on a PC or local network like Specify or BRAHMS. Of course there are also a multitude of dedicated systems developed for specific categories of plants or animals for small user communities like for instance Klasse developed specifically for entomologists in the Netherlands for storing observations and collection data. Except for widely used software, a lot of collection holding institutes have over the years developed their own inhouse management system which usually is not used anywhere else.

5.4.2 Results of the pilot

During the pilot in task 5.3 this type of software was not specifically tested.

5.4.3 Strengths and challenges

Besides storing data in a structured way and possibly linking media (images; sound recordings) to species or specimens, dedicated software usually has built in tools aimed to support collection related activities, e.g. to print labels. Although they are better adjusted to user requirements, they are as a rule less flexible and the learning curve may be steep. Options to upload data to the internet may be absent or may be included.

6. Stakeholders

To a certain degree it is fair to state that in case of private natural history collections the stakeholders are the same as those for smaller or larger collection holding institutes in general. The data from private collections are comparable to data of institutes and museums and influence research and policy making in exactly the same way, just on a different scale. However, there are also certain aspects that are unique for private collections like competition between collectors for obtaining rare species or the economic value of specimens. When considering digitisation of privately owned biodiversity collections and looking for ways to improve it, the following stakeholders were identified:

1. private collection owners [see ICEDIG Milestone 29: p7-9]
2. citizen science associations [see ICEDIG Milestone 29: p. 10]
3. non-governmental organisations
4. collection holding institutions [see ICEDIG Milestone 29: p. 11]
5. research community
6. public administrations
7. data aggregators and/or publishers
8. industry

Each of these stakeholder groups can contribute significantly towards improving the overall success rate of digitisation of private collections. Mutual recognition as involved parties as well as awareness of the respective interests are key factors, both to optimise strategies and to streamline activities and workflows. Here we adopted the 4Rs method to describe stakeholders. For this approach, digitisation of private collections is seen both as a process and as a result affecting different stakeholders. In this chapter the stakeholders are defined and then described in more detail according to the following aspects:
Rights - How can stakeholders access and use the data, originating from digitised private collections.
Responsibilities - What is expected from stakeholders to accomplish digitisation.
Revenues (benefits) - Direct and indirect benefits for stakeholders both from the results of digitisation and from the process itself.
Relationships - How are the stakeholders related to the process and to other stakeholders in domains of trust, respect, and legitimacy.

6.1 Private Collection Owners

Private collection owners are at the heart of their collection digitisation. They are at the same time customer and supplier.

Rights: Private collection owners are the primary data owners who decide about digitisation in the first place and whether - and if so in which way - or not are shared.

Responsibilities: Digitisation of private collections is impossible without the initial consent or action of private collection owners. However, their actions are based on voluntarism and cannot be forced.

Revenues (benefits): By having their collection digitised, private collection owners can much more easily manage their collections or share information with other collection holders for instance for specimen exchange. Digitisation is also required in order to publish data online and opening the collection up to the global research community. This makes collections a lot more meaningful not only for the community at large but also for the owners.

Relationships: Depending on the national context, Private collection owners are most closely linked to non-governmental organisations or citizen science associations. Private collection owners can be members of those associations and share information about their collections through the membership network. Depending on country and cultural aspects, private collection owners may be in contact with collection holding institutes as regular visitors or even as external specialists.

6.2 Citizen Science Associations

The citizen science associations are networks of institutions and/or individuals who utilise citizen science methods, crowd-sourcing or are representing citizen scientists.

Rights: Citizen science associations have no specific rights regarding data originating from digitised private collections.

Responsibilities: Citizen science associations do not carry an active role in collection data digitisation itself. However they could play an important role in communication, not only to encourage their members to list their collections in public catalogues but also to inform them about ongoing developments and activities.

Revenues (benefits): The main aim of citizen science associations is to promote knowledge of and stimulate interest in the subject of the association. Digitising and publishing data from private collections may contribute significantly to that.

Relationships: Citizen science associations carry a networking value for collectors. They can play an important role as "middleman" between more formal structures (governmental agencies, institutes, universities) and individual citizen scientists.

6.3. Non-Governmental Organisations

Non-governmental organisations represent research communities linked to - or whose daily activities center on - specific topics or tasks (e.g. Estonian Naturalist's Society, Estonian Ornithological Society or the Vlinderstichting, EIS and Floron in the Netherlands) including citizen scientists who for instance submit observational data.

Rights: Non-governmental organisations do not usually have the rights to access digitised data from private collections. In some countries, however, natural history associations can act as a non-governmental organisation by representing private collectors or observers can obtain rights to the data they are submitting.

Responsibilities: Non-governmental organisations can support private collectors in their digitisation efforts, collection management or taxonomic expertise.

Revenues (benefits): If private collectors share data with Non-governmental organisations, this can be leveraged to form partnerships with other institutions, government agencies, etc. In some cases the non-governmental organisations have exclusive rights to the data being submitted and raise revenue by providing consultation services for the public administrations based on these data. This sometimes results in data not being shared openly.

Relationships: These specialised interests groups can bring in expertise to governmental projects or serve public interests.

6.4 Collection-Holding Institutions

Collection-holding institutions comprise all the smaller and larger regional and national museums, institutes, universities, etc. holding a natural history collection.

Rights: When there is an agreement on the future donation of a private collection to a biodiversity repository, the data use rights as a rule are transferred as well but sometimes the timing can coincide with the physical transition.

Responsibilities: As a rule of thumb, digitisation of donated collections is the responsibility of the recipient collection-holding institute. Prime goal of the current task within the ICEDIG project is to ensure that in the future more and more collections that are being donated to a collection-holding institute have already been digitised according to the standards and protocol used by that same institute and to ensure that the effort to migrate data into the local management system can be kept to a minimum. Moreover, digitisation by collection owners allows for a more accurate and complete dataset as they consciously or unconsciously include their knowledge and information in the process. Chapter 4 describes

a number of tasks that can be undertaken by collection holding institutes in order to facilitate the digitisation of collections prior to the donation.

Revenues (benefits): There are two important benefits for collection-holding institutes to be gained from digitising privately owned collections. First, it saves time to digitise the collection once it has been donated. Secondly, curators or research staff working at the collection-holding institute may directly use the data for their research.

Relationships: Collection-holding institutes are often the most qualified agents for digitising natural history collections, having access to technology, human resources, knowledge and expertise. Private collection owners may have direct contact with researchers, curators or collection managers working at collection-holding institutes either to receive help or to assist with taxonomic expertise. The institutes also are often organised nationally or internationally in gremia that include other stakeholders.

6.5 Research community

The research community can be considered both as the collection of all the individual researchers whose research is linked to data residing in natural history collections, and as the fora in which researchers are organised to represent their common interest.

Rights: As long as collections have not been digitised (or published in a paper), data and information residing in a collection is only available to the collection owner. Once collection specimens have been digitised, publication online generally implies that data become available to everyone without restrictions.

Responsibilities: If private collection data are used for research, it is crucial to acknowledge this in the resulting paper so as to stress the value of private collections for research and emphasising the importance of their digitisation. Consortia like CETAF and other bodies focussing on research and formed by researchers can play an important role in removing obstacles that prevent or impede private collection owners in the digitisation of their collection. Possible ways in which the research community can assist collection owners are mentioned in Chapter 4.

Revenues (benefits): Researchers are the main end-users of digitised collection data and therefore benefit most from any improvements in the process of making data from privately owned collections available online .

Relationships: The research community communicates through scientific papers and by visiting (inter-)national gatherings. The main objective behind creating a private collection may differ between collectors ("stamp" collector versus the scientific collector). The scientific collector will be part of the research community, the "stamp" collector not.

6.6 Public administration

Public administration in this context includes environmental agencies, ministries, etc. that act in public interests, regulating nature protection, land use, etc.

Rights: No direct rights.

Responsibilities: They play a very important role not only in making policies, regulating laws and cash flows but also in raising general awareness about nature and nature conservation. As law makers, they can play an important role together with the collection holding institutes in tackling specific obstacles, like an approach for illegally collected specimens or species residing in private collections. Overall, by acknowledging the importance of private collections as part of natural history heritage they can stimulate the digitisation process.

Revenues (benefits): To make informed decisions when enacting policies or laws, public administration makes use of reports and data from the research community either based on specific publications or on assignments carried out by NGOs. The more data are available to support results presented in publications or reports, the better.

Relationships: Direct links between private collections and the public administration are weak. However there is a network of contacts between the public administration and other relevant stakeholders.

6.7 Data aggregators

Data aggregators integrate data originating from different data sources into a coherent data pool. They also act as a data publisher, thus making data publicly accessible. One of the well known data aggregators and publishers in earth and life sciences is the Global Biodiversity Information Facility (GBIF).

Rights: They obtain (some of the) rights to digitised private collection data if shared by collection owners in the process of data publication.

Responsibilities: An important task is to provide information about standards and workflows for registering private collections and also for uploading the data. Offering a multitude of tools and applications (spreadsheet upload, data management workbenches, IPTs, etc.) will help collection owners with variable data skill levels to find appropriate data management tools.

Revenues (benefits): Including private collection data in publicly available data repositories will help to fill the taxonomic, geographic and temporal gaps. Hence, it will also raise the value of these repositories for research.

Relationships: Traditionally the biodiversity data aggregators and publishers are not directly connected to private collectors. Data facilities are usually tailored for institutional use and access for private individuals is limited. However, if a private collector is associated with a natural history society or other academically accepted NGOs, this may help to facilitate access to data aggregators and publishers.

6.8 Industry

Data originating from natural history collections can be used for various purposes in industries. Examples can be found in creative industries, tourism, the food and health industries, agriculture or forestry. Another small but important industry is the digitisation industry that provides equipment, software, and services for digitisation!

Rights: There are no inherited rights for the industry sector. The collection data can only be used by the industry when a commercial license is bought or issued by the original data holder.

Responsibilities: The sustainable use of biodiversity data in industrial applications can be beneficial for society as a whole.

Revenues (benefits): Industry can generate revenue, using biodiversity data for creating certain products or building services.

Relationships: Providing biodiversity data to the industry may be an economically one-sided relationship but it may help society to understand and appreciate the value of data in general.

7. Roadmap

This chapter describes in practical terms what is expected or required from stakeholders in order to start, proceed or facilitate digitisation of privately owned collections.

7.1 Private Collection Owners

Regardless of concrete steps to be taken to enhance or start digitisation, a private collection owner can at any time:

- contemplate and ideally decide as early as possible about the person, museum or institute that he or she wishes to donate the collection to in the future,
- ensure that the collection has an acceptable curation standard i.e. specimen/samples mounted or prepared, labeled, identified and stored in adequate containers,
- if applicable, retrospectively look up coordinates for sampling sites visited in the past for which coordinates on specimen labels are still missing and add these either in the field books, in a separate electronic file or directly in records for specimens that have been digitised.

Looking at collection owners with private collections and considering the aspect of 'registration/digitisation' of their collection, roughly three scenarios can be distinguished:

1. not yet started (0% registered or digitised),
2. partly or completely registered in some form of paper archive,
3. electronically digitised either partly or completely with or without the use of unique identifiers attached to the specimens/samples.

Below, the three scenarios are briefly introduced and although their starting situation is very different, the roadmap of actions is by and large identical.

7.1.1 0% registered or digitised

Assuming a collection owner is convinced of the benefits of electronically digitising his or her collection but has not yet started, the most appropriate approach for digitising a collection depends on a number of factors:

- the scope (what will be digitised, in which order?),
- the who (who will do the actual work?),
- the what (what is needed in terms of hardware, skills, knowledge, etc.?),
- the how (data-entry via software or online; online publication; keep data up-to-date),
- the when (the time factor; how long will it take; how much time is invested?).

Deciding on the best approach does not yet include the actual work, but simply sets the scene for digitising a collection taking into account all of the above variables and possibilities in such a way that it:

- matches the collection owners long-term objective in preserving and managing his/her collection,
- perfectly fits the type of collection (taxonomic group; preservation method),
- is within his or her capabilities,
- fits the collection owner's beliefs and opinion on data sharing,
- is an approach that in terms of timing, protocols, collaborators, etc. the collection owner is comfortable with.

Of course a decision about the approach to start digitisation is entirely up to the collection owner and nobody else. However, other stakeholders can play an important role guiding the private collection owners toward a decision. When it comes to a roadmap, the first and most important decision to take by collection owners is to include others in their decision taking process and to not do it all by themselves. Depending on country, type of collection or cultural aspects these "others" could well be other private collectors, citizen science associations, collection holding institutions in the physical world or simply online information assuming this is available and easily findable. It is up to other stakeholders to not only make sure collection owners are well informed at all times about procedures and protocols be they local, national or international (communication strategy), but also to have those very procedures and protocols available.

7.1.2 Paper archive

To this very day, there are still collection owners who have registered their collection not electronically but partly or entirely in the old fashioned way using an indexed card system. Depending on the exact nature of the system and the information it holds, it can or cannot be used in the digitisation process. If, for whatever reason, it is not practical to use the card index the entire collection will have to be digitised from scratch. If the card index system can be used, one could consider a three step protocol:

1. scan all cards,
2. copy data from the scans into a structured file, and
3. apply unique registration codes to each collection specimen/sample and add the code to the relevant record in the file.

If the collection owner is prepared to copy data from the cards into a structured file himself, the first step can be skipped. The advantage of scanning a card index system is that the subsequent registration can be carried out by anyone, anywhere. A roadmap to be taken by a collection owner with a paper archive is the same as for those who still have to start as described in the previous chapter.

7.1.3 Electronically digitised

Quite a number of collection owners have already started digitising their collection. Results from the European survey carried out in 2018 showed that roughly two-thirds of the

collection owners already had some information about their collection digitally available. These collection owners already decided on an approach for digitising their collection. When looking at specimen/sample digitisation the following groups can be roughly distinguished:

1. partly digitised (1-99%)
    a. each specimen with unique identifiers
    b. specimens without unique identifiers
2. completely digitised (100%)
    a. each specimen with unique identifiers
    b. specimens without unique identifiers

From the collection owner's point of view there is probably little to no reason for changing their approach once digitisation has started. However, every collection owner should be made aware that a private collection sooner or later ends up in a regional or national collection-holding institute. In order to optimise the transition from a private collection to a community owned institution, it is important that collection owners take this change of ownership into account when digitising their collection. The more the protocols used in private collections deviate from those used in regional or national institutes, the more time and effort will be required in the future to integrate the collection and corresponding data. When it comes to a roadmap for this group of collection owners it is important they decide which institute will receive their collection and contact it, not only to discuss the future transition of the collection but also, and more importantly, its digitisation.

7.2 Citizen Science Associations

One aim of citizen science associations in this context is to generate interest and enthusiasm in the study of specific nature related themes or groups. Citizen science associations represent the interests of their members and as such can play a key role in connecting private collection owners to other stakeholders, in particular collection-holding institutes. When it comes to the digitisation of private collections citizen science associations can facilitate this process by spreading awareness and information among its members and beyond through mailings, workshops or meetings. In particular, a roadmap for citizen science associations should include:

- searching for and making available relevant information as well as being familiar with (inter-)national programs linked to digitisation,
- informing members,
- looking for ways to have an up-to-date inventory of collections kept by members,
- if feasible looking for means to stimulate digitisation of private collections,
- contacting collection holding institutes to discuss common strategies.

7.3 Non-governmental organisations

Like citizen science associations, non-governmental organisations are organised around expertise and studies of specific groups of organisms. Contrary to citizen science associations, non-governmental organisations (rather than representing the citizens' interests and facilitating them) use data gathered by citizens, whether observation or collection based. It is in the interest of non-governmental organisations to ensure that citizen scientists keep publishing data online and that these data are accurate and verifiable. A roadmap for non-governmental organisations regarding the digitisation and online publication of private collections could comprise:

- together with citizen associations and collection holding institutes create and keep up to date an overview of specialists, national or international, that can be approached,
- to stress the importance of digitisation and properly identified specimens/samples at meetings, workshops, etc. attended by private collection owners,
- to contact collection holding institutes to discuss common strategies, for instance to include (the digitisation of) private collections in atlas projects.

7.4 Collection-holding Institutions

For regional and national institutes, digitisation of private collections is or will become part of their core business. This statement is based on the fact that most if not all private collections one day will be offered as endowments to publicly owned collections. If private collections at the moment of transition are not digitised, they will add to the total backlog of digitisation of the institute receiving the collection. More importantly, information may become lost if private collections are digitised at a far later stage when the original owner is no longer available or capable to provide additional information related to the collection. In order to guarantee the highest possible data quality (accuracy, precision and completeness), and ensuring that datasets are standardized in such a way that they can be migrated easily, institutions should actively promote the digitisation of private collections at the earliest possible opportunity. Therefore, it is of utmost importance that collection-holding institutes facilitate and participate in activities linked to the digitisation of private collections - which in the future will become their collections. A roadmap for collection holding institutes regarding the digitisation and online publication of private collections could comprise:

- contacting citizen science associations and non-governmental organisations to discuss common strategies, like a common communication strategy,
- preparing an up-to-date list of private collections that could be donated in the future with the help of citizen science associations,
- actively approaching private collection owners to discuss their plans regarding bestowal and linked to that the state and or plans regarding digitisation,

- preparing policies and protocols required to streamline various aspects linked to digitisation of private collections, for instance the use of unique identifiers,
- deciding on resources to be made available for digitisation of private collections; if necessary this could lead to efforts to find funding.

7.5 Research Community

The biodiversity research community at large has a vested interest in online data. Whether data is derived from specimens kept in larger museums or small private collections is much less important than the certainty of high data quality and reliable identifications. The already mentioned survey carried out in another task of the ICEDIG project[4] in 2018 showed that the overall size of private collections and the data they contain is impressive and for specific groups or geographic areas unique. The research community is organised in a multitude of programs, platforms and projects both at the national and international level. These platforms (e.g. CETAF, DiSSCo) should be made aware of the importance of private collections as a potential source for biodiversity data. Not only is more communication required but an online tool providing access to high level information of private collections would also be a great asset to raise awareness and to facilitate searching these collections. Once the research community is aware of the (importance of) private collections they may actually be able to contribute to digitisation and online publication for instance by allocating funding, facilities or manpower. More in particular, the roadmap for the research community regarding the digitisation and online publication of private collections could comprise:

- to create an online access portal with information about individual private collections,
- where (inter-)nationally feasible, to look for opportunities to organise stakeholders into setting up a joint approach for aspects related to the digitisation of private collections,
- to develop a communication strategy together with other stakeholders.

7.6 Public administration

Public administration plays a very important role not only in making policies, regulating laws and cash flows but also in raising general awareness about nature and nature conservation. When it comes to a roadmap linked to the digitisation and online publication of private collections, public administrations could tackle:

- as law makers they can play an important role together with the collection holding institutes in tackling specific obstacles like an approach for illegally collected specimens or protected species residing in private collections which are donated to a publicly funded institute.

---

[4] Egmond E & al. (2019) Prioritising scientific and societal needs for data of private collections. ICEDIG deliverable report D2.2. https://doi.org/10.5281/zenodo.2582994

- overall, by acknowledging the importance of private collections as part of natural history heritage they can stimulate the digitisation process.

7.7 Data aggregators

Online access to collection data is provided in a myriad of ways by institutes as well as (inter-)national initiatives in Europe and the world. Each of these portals differ in the metadata they publish, the source of data (observations and/or collections), accessibility and possibilities to add or change data and technical solutions to accomplish this, quality checks, etc. An additional complicating factor is the rights management when private collections, become incorporated in larger regional or national collections. What happens to online data when a physical collection changes address and ownership? Considering the temporary nature of private collections in combinations with the many portals available to choose from, working toward an internationally standardised solution for publishing data from private collections becomes quite a challenge. A pragmatic approach would be to start looking at national levels for solutions to make (meta)data from private collections accessible online. Nevertheless, one should strive for streamlining national initiatives across countries, to anticipate an international solution. When it comes to a roadmap linked to the digitisation and online publication of private collections, data aggregators and online publishers could focus their attention on:

- streamlining efforts to standardise the online publication at a national level,
- notwithstanding national efforts, looking into ways to streamline the online publication of collection data at an international level,
- together with other stakeholders putting effort into optimising messages and channels of communication directed toward all stakeholders involved but particularly the private collection owners regarding the online publication of collection data,
- develop protocols with other stakeholders aimed at keeping virtual data and information synchronised with data and information in the physical collection.

7.8 Industry

Except for companies directly supplying products and services linked to digitisation, industries are not directly involved in digitising or publishing data. Overall, they use data that are directly or indirectly linked to collections. Industry here has a very wide range from food industry to creative industries with no concrete common interests or responsibilities. Therefore despite their vested interests, a roadmap with concrete combined actions has not been drawn up.

8. Recommendations

Digitisation of private collections and sharing the data, first and foremost requires motivation and willingness from the respective collection owners. Current pilots carried out in Estonia, Finland and the Netherlands showed that there are also factors beyond the (direct) influence of collection owners, that affect the rate of success of digitisation and sharing collection related data and information. The most notable challenges that arose as part of pilot projects and recommendations to start tackling these are (in random order):

1. Challenge - best level of approach: What is the best level to tackle digitisation of private collections: is it national or are there aspects that can best be tackled internationally? What are national differences in collection ownership (different attitudes, readiness for data sharing, organisation levels, etc.) and how can these be addressed?
   Recommendation: At the national level, collection-holding institutes and citizen science associations (possibly also including non-governmental organisations) should convene and discuss the best approach. Such national deliberations would very much benefit if a similar effort is being made at an international level by organisations representing collection holding institutes (like CETAF) and citizen science associations (ECSA)

2. Challenge - need for a national hub: Should every country have a national hub to organise the digitisation of private collections and ensure that progress is being made? What input is required for this in terms of manpower and facilities? Is this feasible and can it be funded?
   Recommendation: The answer to this challenge should be part of the outcome of the deliberations mentioned in recommendation one.

3. Challenge - European registry of private collections: At the moment, the Global Registry of Scientific Collections (GrSciCol) is not readily available for private collections. There is a lack of universal understanding how private collections should be treated for registering. A global registry for private collections would be an enormous first step to start boosting collection digitisation. With the lack of a system for naming and assigning acronyms of private collections there is a concern of major confusion when going global or even pan-European.
   Recommendation: The answer to this challenge is twofold:
   - Stakeholders involved from the user community (citizen science associations; collection-holding institutes) should look into ways and methods that enable us to have a complete overview of private collections, including those whose owner is not a member of any citizen association.
   - Secondly, stakeholders from the user community should convene with parties or organisations able to provide technical solutions (GBIF; TDWG) possibly instigated by DiSSCo to make GrSciCol readily available for private collections.

4. <u>Challenge - communication strategy</u>: To really start improving the overall rate and speed of digitisation of private collections, an overall communication strategy is needed that allows for local, national and international communication taking into account cultural differences and differences between collection groups or categories etc.

   <u>Recommendation</u>: A first step to be taken is that the international community representing the stakeholders involved (CETAF, ECSA, DiSSCo) works towards producing a document summarising the most important communication strategies available to approach private collection owners taking into consideration national and cultural differences.

5. <u>Challenge - change of ownership</u>: Private collections eventually change ownership. What implications has this for digitisation for instance for registration-codes applied during digitisation?

   <u>Recommendation</u>: A protocol is required describing in detail the consequences of a change in ownership for the digitisation of collections. Based on the subject and knowledge and experience required, this could be best tackled by the Collections Working Group of CETAF.

6. <u>Challenge - acronyms for private collections</u>: For the digitisation of specimens or samples residing in (private) natural history collections, unique identifiers are required if one wants to be able to uniquely identify specimens or samples. For institutes, acronyms are available that together with a number provide unique identifiers. For most private collections acronyms are not available. To avoid a proliferation of (potentially non unique) acronyms thought up by private collection owners themselves, directives are required about the composition of such acronyms.

   <u>Recommendation</u>: As this is an international issue, the international community of stakeholders should look into ways to solve this issue be it a technical solution or a directive about the composition of acronyms.

7. <u>Challenge - incentives</u>: What incentives can be provided for collection owners to start/continue digitisation?

   <u>Recommendation</u>: The current pilots provided some clues for national stakeholders about incentives to start/continue digitisation:
   - when digisiting, tooling should provide assistance in managing the collection using the digitised information,
   - digitisation should help in exchanging information or trading specimens with other private collectors, and
   - digitisation should provide information back to the collector how the data has been useful to do scientific research.

   The current pilot did not provide sufficient time to make a complete overview. Some more work is required to obtain a full list and present this is a document/paper.

8. <u>Challenge - deployment of volunteers</u>: The pilot in the Netherlands showed that deployment of volunteers could play an important role in digitising private collections

to solve the time issue. Despite this success in the Netherlands still a lot of questions are remaining. For instance can volunteers be deployed in all European countries? How do you recruit them and what is required to organise and supervise them?

Recommendation: Two separate actions are suggested here:

- Across museums all over the world, volunteers have been deployed. Sometimes experiences, lessons learned, etc. have been summarised in reports for instance in the Volunteers Organisers Toolkit. Getting an overview of best practices for volunteers that are or have been deployed in digitisation efforts would be an important contribution if a more general deployment of volunteers is considered.
- an international consortium of stakeholders, for instance CETAF, can send out a survey/questionnaire to collection holding institutes, citizen science associations and non-governmental organisations to ask them about the deployment of volunteers

9. Challenge - collections and legislation: A question that arises when publishing data from private collections online but also when transferring ownership of a collection concerns the provenance of species and specimens. What is the best practice to deal with specimens/species that were collected without proper authorisation (illegally)?

Recommendation: The best approach is a national one where collection-holding institutes and public administrations work together to find practical solutions for the transition of (possibly) illegally collected specimens from private collections to state owned collections.

10. Challenge - synchronisation of physical and virtual data: Once collection data are published online, they need to be updated regularly to stay synchronised with data attached to the physical specimens; collection owners should be aware of this and changes/updates should be easy to implement.

Recommendation: The solution for this challenge lies in further developing management tools that allow private collection owners not only to easily upload collection data but also to make alterations in online data if changes have been made in the physical collection. In addition, this aspect should feature prominently in communication directed toward private collection owners.

11. Challenge - the best data management system: As indicated in chapter 5 there are a lot of possibilities when it comes to data management systems. Quite a number of management tools are not for free but have to be purchased or require a license fee. Is it necessary to make a list of national online tools (PlutoF, LAJI.FI, Recorder 6, National Biodiversity Network, etc.) that can be used for data entry and management? Does it help if nationally, and ideally also internationally, a limited number of options would be endorsed by stakeholders (organisations) together with publishing a list of arguments for and against their use? What actions should stakeholders take?

Recommendation: A clear point of view is required whether stakeholders (collection-holding institutes, research community, data aggregators) should strive for making

recommendations regarding data management tools, and if so what these recommendations should entail. Rather than making recommendation for specific management systems, a list of required functionalities would offer a better starting point for private collection owners.

12. Challenge - protocol highlighting approaches for digitisation: To facilitate the decision making process for private collection owners when it comes to digitising their collection, a generally and digitally available protocol describing approaches for the digitisation of private collections is needed.

    Recommendations: Stakeholders, for instance the CETAF Collections Group, can take the initiative to draw up a general protocol describing approaches to start (or continue) digitising privately owned natural history collections.

13. Challenge - European focal point: Regarding the complexity of digitising private collections an international focal point is needed to oversee ongoing activities, to encourage stakeholders, to streamline protocols and to serve as a source of information.

    Recommendation: To really make progress in digitising private collections across disciplines, nations, and cultures, a European focal point is required. Such a focal point should not only make sure that the importance of the digitisation of private collections is stressed whenever and wherever possible, it should also act as knowledge center, liaise with stakeholder organisations, actively support ongoing digitisation efforts and look for ways to start new initiatives.

9. Conclusion

Developing an approach to stimulate digitisation of private collections in Europe contains quite a few challenges. Although research institutions generally share the same attitude towards data sharing, for which digitisation is the first step, private collection owners may differ in that regard. Their motivation and practices can vary remarkably and tackling digitisation of private collections can therefore be a challenge. The first and maybe biggest challenge is obtaining a complete list of people who privately own a natural history collection which is far less simple than it may seem. Private collections vary in all possible characteristics like size, subject or level of management and likewise their owners differ in their attitude toward the importance of digitising collection data and publishing data online. Cultural differences across European countries as well as differences in legislation regarding for instance collecting animals or plants make developing a harmonised approach that fits all situations difficult. The approach has to offer guidance regarding the multitude of tools that are available to store collection data either offline or directly online. The pilots carried out in this task helped to get a better understanding of the complexity behind the digitisation of private collections. Stakeholders should cooperate much more intensively to come up with suggestions to support and further streamline this activity across Europe. Additionally, a communication strategy is needed to ensure that private collection owners can have access to the latest information regarding all aspects of collection digitisation.