Grant Agreement Number: 777483 / **Acronym:** ICEDIG
Call: H2020-INFRADEV-2017-1 / **Type of Action:** RIA
Start Date: 01 Jan 2018 / **Duration:** 27 months

REFERENCES:
Deliverable **D3.6** / Best practice guidelines for imaging of herbarium specimens
Work package **3** / Lead: **MNHN**
Delivery date **M22**

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

Best practice guidelines for imaging of herbarium specimens

DELIVERABLE D3.6

Authors: Michel Giraud[1], Quentin Groom[2], Ann Bogaerts[2], Sofie de Smedt[2], Mathias Dillen[2], Hannu Saarenmaa[3], Noortje Wijkamp[4], Sarah Philips[5], Steven ven der Mije[6], Agnes Wijers[4], Zhengzhe Wu[3]

1. Muséum National d'Histoire Naturelle, Paris, France
2. Meise Botanic Garden, Meise, Belgium
3. Finnish Museum of Natural History, Helsinki, Finland
4. Picturae BV, Heiloo, Netherlands
5. Royal Botanic Gardens, Kew, UK
6. Naturalis Biodiversity Center, Leiden, the Netherlands

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

ICEDIG.EU

Abstract 2

1. Introduction 3

2. Preparation of herbariums for bulk imaging 5
2.1. Assessing the number of specimens 5
2.2. Barcoding 6
2.3. Mounting/ restoration 7
2.4. Packing and transport 8
2.5. Metadata collection 9
2.6. Freezing 10

3. Major digitisation approaches for herbarium Collections 13
3.1. Overall preparation of collection for mass-imaging 14
3.2. Tasks to prepare an herbarium for mass imaging. 15
3.3. Possible tasks for mass imaging in house, by contractor 18
3.4. Possible tasks for mass imaging in house by contractor 19
3.5. Possible tasks for mass imaging outsourced by contractor 24

4. Case Studies 25
4.1. Meise Botanic Garden (DOE! & DOE2!) 25
4.2. Case Study Royal Botanic Gardens, Kew 28
4.3. Case study Digitarium 31
4.4. Case study Naturalis 32
4.5. Case study Paris 33

5. Conclusion 36

6. Recommendations 37

References 39

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Abstract

This report explains the processes to prepare a herbarium for digital imaging. It is divided into four main sections. The introduction summarizes the aim of mass digitisation and lists the reasons why several institutions started mass digitisation. The second chapter focuses on the preparation phase as it is a critical part of the process. Here the similarities and differences of herbarium collections are examined. Though each collection has its own peculiarities, the main topics of preparation are the same in every digitisation project. Preparation is a complex and often lengthy process. Past experience shows that details are important: assessing the exact number of specimens, addressing the issue of barcoding specimens, getting the appropriate means for transportation, down to the choice of boxes for example, accounting for possible pest infestation and dealing with unmounted specimens. The third chapter addresses the preparation of a herbarium collection for the imaging process itself. Emphasis is on weighing the merits of 1) in-house imaging by ones own staff, 2) in-house imaging by a contractor, and 3) outsourced imaging to a contractor. This section also details the elements of the workflow that are necessary to achieve mass digitisation. The last section presents five case studies of mass digitisation by renowned institutions: Meise Botanic Garden, Meise (BE), The Royal Botanic Gardens, Kew (GB), Digitarium (FI), Naturalis (NL), and the Musée National d'Histoire Naturelle, Paris (FR).

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

1. Introduction

In the past 10 years, mass imaging of herbaria projects has been developed in several institutions. Herbarium sheets are similar in many respects to other two-dimensional objects, such as books and documents. So, expertise on mass imaging of documents can be fruitfully adapted to herbaria. In-house imaging of herbaria is already taking place as are international digitisation programs. For example, the Plant Initiative funded by the Andrew W. Mellon Foundation has concentrated on the imaging and dissemination of nomenclatural type specimens. Such exercises have provided valuable experience of digitisation to institutions.
The change to mass digitisation is nevertheless a significant change in scale for both collections and the contractors who specialize in imaging documents. Mass imaging has its own peculiarities and has to be approached differently to the ad-hoc imaging workflows that will continue to be part of the long-term work of an herbarium.
Mass imaging can be defined as processing a large number of specimens, for a set cost, in a defined time period, the rate of processing samples is therefore key to a measure of its efficiency. Furthermore, quality control is an important pillar in addition to efficiency (Nieva de la Hidalga et al. 2019).

The goal of imaging collections is to create a digital copy of a specimen, increasing the number of potential uses and users through online access. In addition, when a certain number of objects have been digitized, it becomes possible to compare data and images cross validate information and infer additional information.

Another important reason to image collections is as a simple safeguard. Archiving information digitally so that all will not be lost, even if the original object is destroyed. This is part of the curational responsibility of an institution. In the past, even before digitisation, James MacBride's (1892–1976) images of type collections around Europe are the last trace of many sheets of the Berlin herbarium before its destruction at the end of the Second World War (Grimé & Plowman 1986). In more recent times, digitisation projects by Google in the National Museum of Brazil ensured that some of the objects lost to the 2018 fire are still available digitally (Coughenour 2018).

Images of specimens cannot completely replace access to physical collections. Microscopic examination of the specimen will always be required, both to identify certain species, but also to examine cellular morphology. For example, examination of the change in stomatal density with increasing atmospheric CO_2 concentration cannot be achieved without much better images than are available from mass imaging (Woodward 1987). Nevertheless, even low-quality images can be used to evaluate whether specimens will be useful in such a study.

In this report we describe the processes to prepare a herbarium for imaging. There are many decisions that need to be taken and we describe the pros and cons of the different options. The exact methods each herbarium chooses will depend on many local factors including the state of the collection, infrastructure, staff and funding.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

The amount of work required for preparation should not be underestimated. Efforts spent on good preparation will make the whole procedure run more smoothly, reduce stress on staff and will result in a better-quality product. Likewise, poor preparation can lead to systemic problems with the collection and its images that will linger with the collection for many years after the imaging work has finished. Mass imaging is a disruptive process for an institution used to a daily routine, however, it is also a huge opportunity for science and the institutions international profile. Preparation is key to making a success of such a project.

This report is based upon the combined experience of five European institutions with herbarium collections that have conducted mass imaging projects: Paris National Museum of Natural History (France), Naturalis Biodiversity Center (The Netherlands), Meise Botanic Garden (Belgium), Finnish Museum of Natural History along with its shared digitisation centre Digitarium (Finland), and The Royal Botanic Gardens, Kew (UK). Each institution had its own peculiarities and needs with regards to imaging, which depend on local policies, the nature of the building, the nature of the collection and the priorities for imaging.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

2. Preparation of herbariums for bulk imaging

The key to mass imaging is the ability of the equipment to capture images fed by a conveyor belt. Therefore, the chief aim should be to properly organize the feeding of the conveyor belt, so that a high number of specimens can be processed per day. This organization is a question of estimating the time and room required for each step, and an according allocation of resources. Most critical is the identification of bottlenecks in the different continuous processes, as these will have the biggest impact on processing time. These processes may differ somewhat between different institutions, as amongst others the organization of the physical specimens, the architecture of the facilities and the composition of the workforce can vary considerably. Hence, the bottlenecks may not always be the same. Generally, the following key steps can be identified:

- Assessing the number of specimens
- Barcoding
- Restauration/ mounting
- Packing and transport
- Metadata collection
- Freezing

2.1. Assessing the number of specimens

The high fixed costs for the equipment and installation mean that mass imaging is only cost effective above a minimum threshold of the number of specimens. Therefore, estimating the number of specimens is important to decide on an institutional imaging strategy. Collection holders of small herbaria collections have sometimes formed consortia to reach a critical number, as was done in France in the Recolnat project. (www.recolnat.org). Estimating the number of specimens is not only important when tendering external imaging companies but also for project management of staff, equipment and working space. Assessing the number of specimens in the collection is an impactful parameter. The more accurate the numbers, the more realistic the goal can be determined, differences in numbers can cause major differences in the final costs per specimen. The estimated costs will play a crucial role in the acceptance of financing proposals for digitisation.

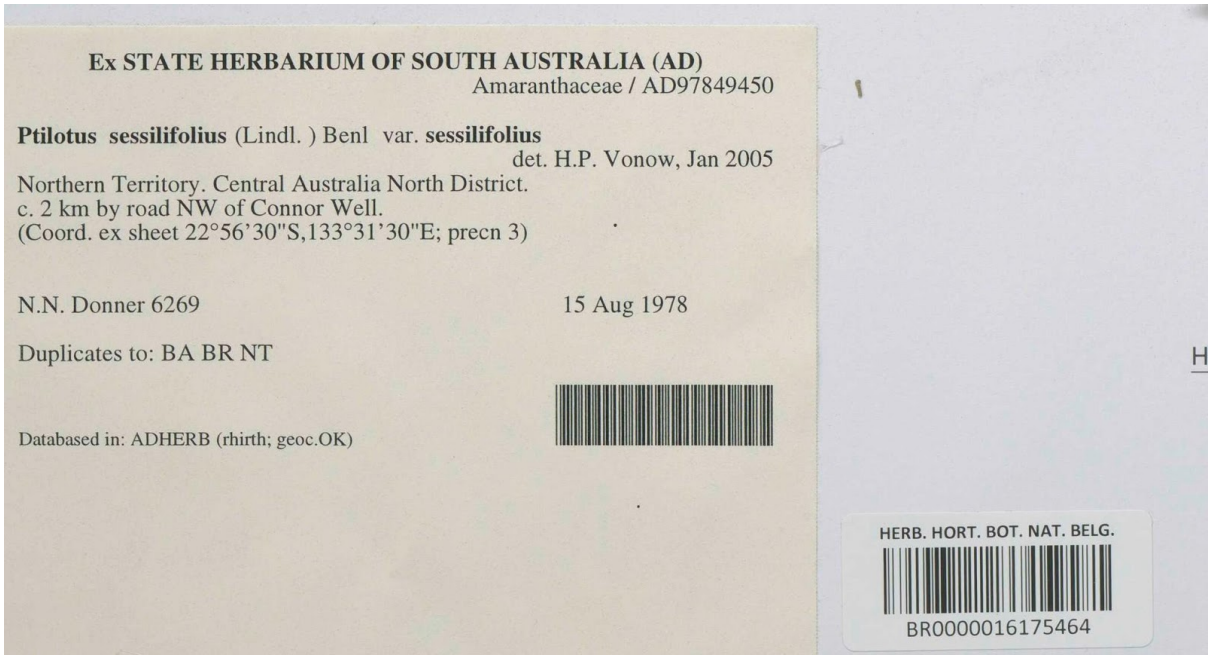
Estimating an accurate number is difficult, whatever the size of the herbarium. Most collections are built out of acquired donations, which are often overestimated in number. Usually, institutions have knowledge of the number by approximation from a count of their storage units (shelfs) multiplied by an estimated number of specimens per unit. Herbaria differ in the size of their storage unit, but for example at Meise Botanic Garden it is generally about 40 specimens, other herbaria are probably of a similar order. Another method that is used is counting all the herbarium specimens stored in 10% of their cupboards and extrapolate this number to the whole collection. One of the results of mass digitisation is that it will give an exact count of the number of specimens.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

2.2. Barcoding

Barcoding is one of the most important steps in the imaging process, because mostly the barcode is the link between the physical specimen and its digital copy. Often, the barcode is also used as the link between the image files made of the specimen and the data digitally captured from it, i.e. the specimen's record in a digital catalogue (Dillen, Groom and Hardisty 2019). Barcodes have to be unique and therefore herbaria often use their herbarium acronym or code as the alphabetic part of the barcode, followed by a series of numbers. Index Herbarium codes are commonly used. Over time, it is possible that institutions changed from one barcode format to another. They may also have used different barcodes for different sub collections. It is important to have a list of all possible barcodes present in the collection, so that the digitisation team can incorporate this in the programming of the software for renaming the scanned sheets.

An issue that occurred during the second mass digitisation project at Meise (2019-2020) is that some herbaria printed a barcode on the original label and sent out a copy of the label to other herbaria without erasing this barcode. This poses a problem for automated barcode recognition software and all these exceptions require additional development work from the digitisation team.



Meises' barcode is attached to the Adelaide label. The Adelaide number is still written on the label.

At Meise, as in most other herbaria, each specimen gets its own unique barcode, even if more specimens are mounted on one sheet (i.e. multi gathering) which was frequently done in the past. The number of barcodes may add up to seven or more on a single sheet. Multi gathering sheets are a problem in the imaging process if they are not barcoded yet before imaging. Asking the operators at the conveyor belt to recognise them and add the correct number of barcodes is very difficult and would constitute a major bottleneck. A possible solution for this multi gathering problem is asking the operators to just add one barcode and to mark these specimens during label transcription afterwards. The decision has to be made by the institution itself, depending on the time, effort and resources it would take if the

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

specimen barcode is added at the conveyor belt or during pre-curation. Good instructions have to be given to the operators where to put the barcode if decided to delegate the barcoding to them. If this step is to be completed by the institution, it is important that the institution allocates enough resources to barcode enough specimens in advance of imaging to avoid the imaging process having to be halted due to lack of prepared specimens.

Another part of the process to consider is the barcode on the cover of the folders containing specimens. These barcodes are added to capture the filing name and other information like country or region present on the folder cover, so that during transcription this information can be linked to all following specimen sheets. Cover barcodes can be placed on each folder, which is the easiest for the operators, but also the most expensive way. Another approach is, as was done in Meise, to only add a cover barcode to a folder when the name on the folder is different from the previous one. During the current digitisation project at Meise, an average of 1 folder for 10 sheets was observed.
In the first mass digitisation project at Meise, the cover barcodes were added by in-house staff, but during the second project it was outsourced to the digitizing company. When outsourcing this part, it needs to be kept in mind that training of the operators takes a while and follow up is necessary because of all exceptions in the collection.

The last part to consider is the barcoding of boxes, drawers, piles, cupboards or other units in which the specimen folders are kept. Very rarely the herbarium may have some information labelled on those units which needs to be transcribed and linked to all following specimen records, but mostly this is for practical reasons. Box barcodes can be used to keep track of what was already processed or what needs to be transported and where. This was key to the very complex transportation, mounting and digitisation process developed for the e-Recolnat project in France. Furthermore, though the digitisation process is highly automated, some issues or errors may arise which necessitate re-scanning. In such cases the box barcode linked to the scan in question facilitates the easy recollection of the specimen, which may otherwise be very hard to find.

2.3. Mounting/ restoration

Most vascular plant specimens are eventually mounted on a, more-or-less, standard sheet of white paper, together with their label(s) and barcode. However, the degree to which collections are mounted varies widely. Furthermore, precious, fragile and old collections may need some restoration before they are suitable for mass imaging. Restoration of the collection before imaging is always a good idea when you have the necessary time, staff and resources. If your collection is well mounted/restored, the imaging process will be much faster (no loose material that can fall off or can cause bad cropping) and the images will look nicer.

It has been a prominent question for French projects, where many herbaria were not completely mounted, if mounting would be added to the digitisation process or not. While it might be possible to bulk image an unmounted specimen, the process would be very slow and cause damage to the specimens, so it is not recommended. The unmounted specimen must be placed on a white sheet, a barcode must be added and the label must be well positioned on the sheet. In the imaging workflow, it is best to decouple the mounting and

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

restoration of specimens from imaging, as this reduces the risk of a bottleneck. This decoupling proved to be efficient in the experience of Naturalis Biodiversity Center. However, this disconnection implies that the specimens have to be handled twice.

Outsourcing the process of mounting is only worth doing if the ratio of unmounted to mounted specimens is high and depends on the rate of mounting that can be achieved. The time of mounting an individual specimen is much longer than the time to mass image it (mounting rate 50 specimens a day to imaging rate 5000 specimens). Therefore, the resources, in terms of material, staff and space, to do the mounting always exceed the resources needed for the imaging.

In all the digitisation projects examined, the rate of mounting was potentially a rate limiting step. There was often a risk of delaying the imaging workflow due to a lack of mounted specimens. For this reason, it is important to implement a mitigation strategy for such an eventuality. It was regularly experienced during the French digitisation endeavours that additional resources needed to be spent to locate mounted specimens elsewhere in the collections for scanning, as mounting of the specimens planned for scanning could not keep up with the imaging.

The condition of specimens affects the rate of imaging. The less a specimen is handled during imaging the better. In part because of that reason, mass imaging has been undertaken more for phanerogams than cryptogams, as the latter are more fragile and/or stored in envelopes that require additional time to open during the imaging workflow. Cryptogams also are only rarely mounted in a flat manner like herbarium specimens of higher plants and bulk imaging is only useful for determination if microscopy is involved.
RBGE made a small film on all aspects of mounting herbarium sheets:
https://www.youtube.com/watch?v=35qgfShomYY

2.4. Packing and transport

In general, the customer is responsible for the packing and transport to the imaging-space. The importance of correct packing is often underestimated, who is not able to handle some boxes?
Experience learns that most institutions are not equipped to carry out practical processes of this kind. In mass-imaging projects the most important issue is the ongoing stream of herbarium sheets to the imaging space; the conveyor belt needs to be fed all the time. It is therefore recommended to organise a test route to get to know every detail in the building that might cause an obstruction or delay; elevator, thresholds, narrow passages and other possible blockages.

Packing material is another topic to think through on forehand. In the Recolnat project special double walled boxes were used to prevent the contents from being damaged after the boxes were stacked on pallets. Certainly, when it comes to transport over longer distances and more pick-up and transfer places it becomes important to use special boxes, though the price can rise considerably.
Even the supply and storage of the empty boxes must be included in the process.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

For a project in Meise, a sectioned trolley was developed so each shelf of the herbarium could be transported and imaged separately, and afterwards returned to the same shelf in the herbarium. This was possible since the imaging space was on-site and in the same building. Not all cupboards in Meise's collection space have the same size or contain the same volume of herbarium sheets, therefore, the transport of each trolley was checked by the staff of Meise.

The herbarium collection of Naturalis is packed in strong specially designed herbarium boxes, that are firm enough to stack. They were loaded on trolleys that went from the collection storage space to the imaging space and after being imaged to a third building where they were stored afterwards.

To ensure that the herbarium sheets are returned to the right place in the collection, encoding boxes is required. Floor, cupboard, drawer, pigeon hole, everything must be recorded in a barcode. Especially when collections from all over the country are brought together to one imaging station, it is crucial to develop a thought-out system. In Picturae's digitisation station in Montpellier, where 42 collections from the South of France (Occitanie – Provence Alpes Côte d'Azur were brought together, the audit trail that certifies correct handling, was all in the hands of the contractor. Due to the different circumstances in the different collection locations, special boxes were delivered directly on site, or stored at the digitisation station till delivery was possible or wanted. Each box that was packed with herbarium sheets received two labels with identical information, one transmission slip inside the box and a second label stuck on the outside of the box. Each label contained the name of the institution, the herbarium code, the name of the collection, as well as the destination and a box number.
Each box contained a bundle of herbarium sheets, a transmission slip (in the box) a label glued to the box and two identical barcodes. Even small details, like closing the boxes in the right way, without hiding the barcodes or box numbers were part of the audit trail. Boxes were then transported to the loading dock of the institution, where they were regrouped by collection and by destination and stacked on pallets. After the authorised staff gave permission, insurance was double checked and the number of cards were entrusted by the IP, the transport could take place.
In Montpellier, with the reception of the collections, all details were checked again: recorded data from the carton label: date and time (automatically), institutional name, institutional code, collection name, barcode number, destination (tying or scanning) possible problem by transport and name of the final digitisation site.

2.5. Metadata collection

Before discussing preparations for metadata collection, it is good to differentiate between metadata from imaging and from transcription. During digitisation it is possible to record certain data about the imaging, like date and time of imaging, type of camera used, location of imaging, the institute that owns the collection, an institutional code, or even a collection name or code. This can be done in the image itself or in a separate file, preferably CSV. In case a separate file is used, the information should be linked to the scan's barcode, to be able to properly link the information to the scan and original specimen. Such a separate CSV file can then also be used to add information by transcription from the specimen labels.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Before digitisation starts, it is important the herbarium decides which metadata they need. For imaging the metadata needs to be recorded at the moment of imaging. Though some metadata, like institute or location may be added later, other information, like a correct time stamp which may be useful in some use cases, such as the links between the specimen and its folder, is much harder to add afterwards.

Most preparation may go in transcription metadata. Obviously, the clearer the labels, the more accurate the transcription will be. However, herbariums generally don't have the capacity to go through the whole collection to add new labels to the specimens. It may be worthwhile, though, to update the folder labels. Those most often contain the scientific name of the specimens inside. When during scanning a file is created in which the sheet barcodes are linked to their folder barcodes (see chapter on barcoding above), it is possible to enter the scientific name and any other information relating to everything in that folder from the folder and copy it to its respective sheet records. A check if all specimens in that folder indeed belong there, would prevent an incorrect name to be added to a specimen stored in the wrong folder. It could also help to combine different folders with the same specimens into one folder, so it only has to be entered once. Transcription from the folders could save a lot of money for other things and the better prepared they are, the better the end results. Provided that the herbarium has the staff available to do such a thing, going through the folders and updating the labels would improve the accuracy of the transcription and would make sure the most recent or preferred name is entered.

For sheet or specimen transcription a decision needs to be made on which information will be needed after digitisation is ended. For newer collections, clear labels with much information are often present, so much can be entered from the labels without a problem. If the collection is old, with many hardly readable labels and/or with little information on them at all, it may be better to only enter one or a few fields that are most often present or to forego transcription altogether. Such decisions would depend on several things like the budget of the herbarium for transcription, the type of information most used by the researchers of those collections, the type of labels or information present on the specimens, the type of research the herbarium plans to participate in in the future etc. Also, the methods available for transcription, like in house, volunteer crowdsourcing or outsourcing, may influence this decision (for a more in-depth discussion of those methods, see Phillips et al. 2019).

2.6. Freezing

An important part of the curation of herbarium specimens is making sure insects do not get into them, which would destroy the collection (Bedford 1999). A variety of insects, mostly beetles will eat and destroy herbarium specimens (Table 1). The most common decontamination procedure is to freeze the specimens. Two days freezing at -20°C is generally recommended (Florian 1986; Gilberg & Brokerhof 1991). The lower the temperature, the shorter time needed to make sure the specimens are pest free (ref). After a certain amount of time this process needs to be repeated, the amount of time depending on the environmental conditions in the herbarium building.

Often freezing is not directly thought of as part of a digitisation project. However, especially when moving the specimens to other locations for digitisation off site, it is important to

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

consider the risks of transportation without precautions. This is especially the case if collections of other institutions are digitized at the same location. Even if your own collection is clean, another collection may cross-contaminate your own. For example, the client for one project with several herbariums to be scanned at the same location, decided against freezing. Then some of the specimens appeared to be infested with insects. The whole building needs to be cleared out and cleaned by a special company and all specimens decontaminated, not only the ones with the original infestation, but all other specimens in the building at that time as well. This can delay digitisation projects quite a bit. After this a freezer was hired and all specimens were frozen for two weeks before moving on to digitisation.

Therefore, it is recommended to make freezing part of the project, either by freezing them before digitisation or after, or both, depending on the circumstances of transport. Freezing can be done at the digitisation location or in house. Even when digitizing in house, it may be a good idea to incorporate freezing into the project. As the specimens are being moved around anyway freezing at this point may save some time and reduce handling overall.

Depending on the chosen path, the freezing time needs to be incorporated into the planning. To kill insects the core of the stack of boxes on a pallet must be cooled to -20 °C. It is difficult to estimate how much time it takes to bring a collection to a core temperature of -20 °C because each collection is different and therefore packed differently. Once the core temperature is -20 °C, keeping it at this temperature for 2 hours is sufficient, though longer will not hurt.
It goes without saying that this is easier and faster for a loose sheet or bundle of sheets than for an entire pallet. The quantity of material in a box, the type of box and what kind of stretch wrap film is used, plays also a role. A reasonable rule is 14 days at -20 °C or 7 days at -40 C, that is probably a bit too long, but it guarantees the correct core temperature. In order to be completely sure, the temperature of the specimen in the middle of the package should be measured per package (bundle / box / pallet / Danish cart / etc.)

There is the option is a so-called freezing hotel: this is a company that rents out freezer space for certain amounts of time. If one chooses to use a freezing hotel, extra transports to and from this location need to be budgeted. When outsourcing digitisation, an experienced outsourcing partner should be able to help with deciding the best option for the project in question and in organizing this, also depending on the size of the project.

It is also a good idea to keep the humidity low wherever specimens are stored and ensure open windows have screens to prevent insects gaining access to the area. Ultraviolet lamp traps can be used in rooms that are used to process specimens to catch any stray insects coming from the specimens or from outside.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Table 1. Common herbarium pest species

Species	Order	Vernacular name
Liposcelis bostrychophila Badonnel, 1931	Psocoptera	booklouse
Anthrenus verbasci (L., 1767)	Coleoptera	varied carpet beetle
Anthrenus flavipes (LeConte, 1854)	Coleoptera	furniture Carpet Beetle
Attagenus unicolor (Brahm, 1791)	Coleoptera	black carpet beetle
Dermestes lardarius L., 1758	Coleoptera	larder beetle
Lasioderma serricorne (Fabricius, 1792)	Coleoptera	cigarette beetle
Stenobium paniceum (L. 1758)	Coleoptera	drugstore beetle
Thylodrias contractus Motschulsky, 1839	Coleoptera	odd beetle
Trogoderma granarium Everts, 1898	Coleoptera	cabinet beetle

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

3. Major digitisation approaches for herbarium collections

Three major use cases in carrying out mass-digitisation of herbarium collections can be identified: 1) in-house by own staff, 2) in-house by contractor, and 3) outsourced to contractor. They differ in several aspects and have their own constraints. Each of them has its pros and cons, depending on the properties of the collection and institutional needs and priorities. Cost, of course is a major factor, but also space and time need to be considered, in particular the stress placed onto the organisation and people. In the table below we try to categorise the use cases from several angles which may help to choose the right approach if future projects.

Other institutions

In this report, we draw on experiences from institutions who have experience with digitizing their herbarium collections, but mass-imaging of herbarium collections has been exercised also elsewhere, including Harvard University, Moscow University Herbarium, Natural History Museum at the University of Oslo, Agriculture and Agri-Food Canada in Ottawa, and Smithsonian's National Museum of Natural History. We have been in contact with several of them, as colleagues and sometimes also as contractors. These experiences are being reflected in the next sections.

	In-house by own staff	**In-house by contractor**	**Outsourced to contractor**
Scope of tendering	For hardware and support	For services, with details relating to cooperation	For services
Imaging system acquisition cost	100,000€ (see Note 1 below)	0€	0€
On-line data storage at 0.5€/GB/year and 60MB/specimen	5,000€/year	10,000€/year	10,000€/year and up
Staff required	Flexible. Ideally 2 operators, 1 barcoder, 1 part-time quality controller, 1 analyst/ programmer/ data manager	1 project manager, 1 curator to select material, 1 quality controller/ data manager	1 project manager, several curators to select and pack materials, 1 quality controller/ data manager

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Demand on staff time	Initial training required. Can work at own pace.	High demand to work at the pace of contractor.	High demand to pack and unpack material at specific times. Otherwise medium demand.
Transport of collection	No	No	Required, with cost and risk, and unavailability of specimens for a period of time
Space needed at herbarium	Room 10 by 5 m	Room 10 by 5 m	None
Optimal size of collection	500,000	1,000,000	Any, as many collections can be scheduled serially
Daily output max	2,500, but variable in reality	5,000 sustained	5,000 and up, using parallel lines
Cost	~1€/sheet	~0.7€/sheet	~0.3€/sheet
Typical project budget	Small to medium	Medium	Large
Special features	Hardware cost can be shared by moving equipment between cooperating institutions.		If facility is sustained over long run by several institutions, also small collections can be processed.
Institutions using the approach	Harvard, Helsinki, Moscow, Ottawa	Meise, Naturalis, Smithsonian	Kew, Oslo (and other Norwegian), Paris, Recolnet

Note 1: This is a quoted list price from the company Bioshare Digitisation. It consists of acquisition of the hardware components, cost putting the pieces together (3-6 person months), software licence, warranty, installation and configuration of server software for post-processing, and office costs. Additional costs rise from shipping and eventual support and customisation functions. Also not included is a back-end server needed for post-processing and backup of data.

3.1. Overall preparation of collection for mass-imaging

Whatever you choose, 1) in-house by own staff, 2) in-house by contractor, and 3) outsourced to contractor, the considerations mentioned below must be taken into account.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

3.2. Tasks to prepare an herbarium for mass imaging.

Assessment of specimens	Restorating/ mounting specimens
Institutional or governmental policy	Quality Control procedures
Barcoding sheets, batches, folders (filing name)	Decontamination
Replacing folders and separating paper	Marking previously imaged specimens
Desleeving	Image ingestion procedures
Unfolding envelopes	Decision making workflows

The above-mentioned tasks do not stand on their own. Not all tasks will be applicable for every digitisation project, but it is important to think out a thorough plan. It is therefore of great importance to question the collection and institutional situation on forehand. The size of the collection, the state in which it is, the space where it is located, the way in which the objects are encoded, all aspects contribute to making an action plan. Small housing and large numbers make it obvious to image and digitize off-site. Though a small staff, dictates help from outside, it can be desirable to involve the staff in imaging and digitising the collection so they gain more insight and control over the collection.

Assessment of specimens

As stated before, assessing the number of specimens in the collection is an impactful parameter. The more accurate the numbers, the more realistic the goal can be determined, differences in numbers can cause major differences in the final costs per specimen. The estimated costs will play a crucial role in the acceptance of financing proposals for digitisation (see 2.1).

Institutional or governmental policy

Herbarium specimens are both cultural heritage and scientific tools. Institutional norms for handling a specimen depends on the collection management policy. To ensure extra care, some curators will impose strict procedures and ask for trained staff, especially if the digitisation is to be subcontracted.
In house or governmental, in case of national collections, policy, may restrict the transportation of the specimens outside the institution; applying for special authorization can be time consuming and might lead to rigidity in the process of transporting specimens to the digitisation equipment.

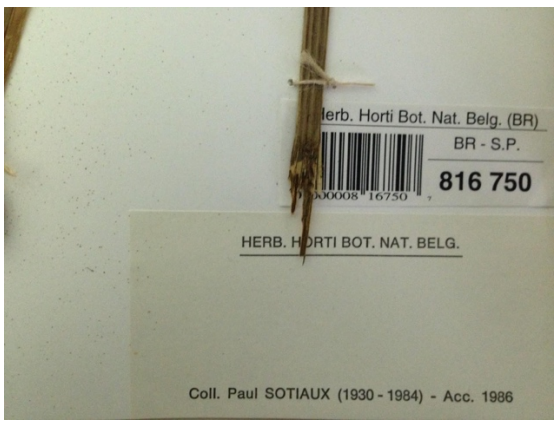
Barcoding

The starting and ending point of the process is the specimen stored in its cabinet and during the process an image is taken. Different people doing different things will be involved, and therefore, each stage of the process has to be documented in order to be able to identify the default in the process and to do the necessary corrections. The audit trail to follow the collection items can only be done by barcoding objects, boxes and pallets.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Applying a barcode or a data matrix on each sheet is the surest way to track each specimen as the barcode or data matrix is immediately recorded in a database. Next to that each pigeon, hole, cupboard and room must be labelled as well. For a project involving several herbaria, the traceability concerns also labelling the original institution to make sure that the specimens are returned to their rightful owner.
RDF tags might help to monitor quality aspects as well, like humidity, but the price of these tags does not does not outweigh the benefits (yet).

Covered barcodes: Meise experienced that sometimes specimens already barcoded in the past and these barcodes are sometimes covered by parts of the plant or determination slips. On the conveyor belt, it is not possible to automatically read the barcode and rename the image. Therefore, when this problem occurs, these specimens are kept aside and must be imaged and databased in house.



Barcodes which are partly covered by the plant. These specimens are kept aside and will be re-imaged in house.

Desleeving

Plastic sleeves: when going for mass digitisation, plastic sleeves need to be removed from the specimens to prevent reflection on the image. If this task needs to be done at the conveyor belt, keep in mind that it slows down the process. You can only leave the plastic sleeves on when imaging specimens who all have plastic sleeves because you can then adjust the settings of the camera.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Specimen imaged in a plastic sleeve on the in-house digitisation infrastructure

Unfolding Envelopes

Envelopes: images of specimens who are kept completely in envelopes are only useful for the label information on the sheet. Therefore, they need to be re-imaged in house after the mass digitisation if they were not kept aside during pre-curation. These images can be easily traceable when they are flagged during the transcription of the label information. Thanks to the track and trace system of the external company, they can be easily retrieved from the collection. example of envelopes.

Specimen completely kept in envelope, needs to be reimaged in house.

Restoring and mounting specimens

Unmounted specimens will stop the process of the ongoing feed of a conveyor belt and need to be addressed in another way. In Montpellier a large amount of the collections needed to be mounted and/or remounted before they could be imaged. A complete crew was trained to mount the herbaria sheets and the mounting became part of the workflow.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

At Meise, the only unmounted specimens that are stored in their collection are duplicate specimens which will be send out to other institutes once they have a determination. Therefore, these specimens don't need to be imaged and are kept aside on the conveyor belt (for more information on mounting see 2.3)

Quality control procedures

All the stages of traceability and checks are part of the quality control that embraces the whole process. Procedures which are implemented to answer defaults or errors have to be related with the ability to reprocess the specimens if a problem is detected. The best and safest procedure is being able to check each specimen at all the stages of the process. However, this can be found too costly and the periodicity of the checking can be adapted to fasten up the process and checking by random sampling can be implemented but the time span between controls must be consistent with the ability to reprocess a whole set of specimens. If the digitisation is sub contracted out, the institution has to set up its own independent procedure besides the procedure set up by the contractor, however this procedure will generally be based on random sampling with statistical tests to accept or reject the production checked. Accepting a percentage of errors is the responsibility of each institution, considering that the lower this percentage the higher the cost of digitisation are. Costs of quality control are most of the time underestimated, while quality control is a very important phase of the imaging and digitisation process, checks should be implemented at all possible stages in the mass digitisation process.

3.3. Possible tasks for mass imaging in house, by contractor

e.g. Naturalis with Wageningen collection as test case and Paris

Training staff	Health and safety check

Training staff

At the start of the imaging process, once the conveyor belt is set up, it is necessary that all operators are well trained, as well as by the external company how to work with the equipment as also by the herbarium curators in order to know how to handle specimens.

Health and Safety issues.

In the past, specimens were routinely treated with chemicals to prevent insect damage. Mercuric chloride, in particular, has been a standard treatment until the 70's. These chemicals represent a potentially acute and chronic hazard; manipulating each sheet can mobilise contaminated dust and exacerbate the problem. Some sheets are labelled as being treated, but that knowledge is not necessarily correct as over time wrongly stamped sheets have been used. However, information is often lacking and even when present are not necessarily up-to-date. According to local regulations, chemical analyses prior to digitisation is required. If toxic products are detected, depending on health and safety rules, handling requires wearing special equipment, which can increase handling time. During the imaging of the collection at Meise ventilation was installed in the herbarium that reduced the indoor mercury levels to well below recommended thresholds and the handlers needed to wear

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

plastic gloves and lab coats at all times. A fume hood was installed in the imaging room, and specimens were stored there before being imaged.
At Naturalis, extensive attention was paid to the health risks associated with the inherited herbaria. The greatest individual risk was caused by the earlier use of mercury compounds for, among other things, keeping the collections free of mould. Not only the collection managers and researchers who worked with specimens were at risk, but also employees of the herbarium digitisation project and logistic staff. During the digitisation project, rigorous measures were taken, ventilation was installed in the imaging room and all employees were tested for the amount of mercury in their urine, at the start, halfway through the project and at the end of the project. The necessary personal protection was made obligatory (gloves, lab coats) and vulnerability of staff members was assessed. For instance, young women, people with respiratory problems etc. The result was that sufficient precautionary measures had been taken to tackle the problem.

Naturalis had TNO and ARBO-diensten set up an intensive measurement program to determine which actions involved a risk and which restrictions and protection options had to be deployed. This resulted in a series of measures, including ventilation, regulations for collection actions and personal protective equipment. At the end of 2015, the combination of measures taken has reduced risks to far below the limits above which action must be taken. A monitoring program has been implemented that further monitors the risks.
In addition, a project was set up to tackle the risk at source. In this project, which was conducted under the direction of TNO, a type of paper has been developed that can support a substance that forms a connection with the mercury in the evaporated mercury compounds. It has been established in the project that this paper can be produced and meets the requirements of stability and absorbency. A business case has also been developed for the production of the paper. However, production of the absorbent paper is still awaiting producer initiatives.

For every herbarium collection the contaminants will differ. Assessment of the used chemicals should be made before the beginning of the digitisation. Known hazardous chemicals include: mercury contamination, methyl benzoate, naphthalene, jodide and dichlorovinyl-dimethyl-fosfaat (dichlorvos).
Moulds, spores, dust can also cause serious health issues. Care should be taken to prevent these particles to spread in the work environment.

3.4. Possible tasks for mass imaging in house by contractor

e.g. Meise Botanic Garden and Naturalis

Tender preparations	Packing
Transport within buildings	Preparation space
Transport between buildings	Transport equipment in case of contractor from other country

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Tender preparations

Every collection has its own specialities which can be covered if pre curation is done before imaging. If this is not possible, it is necessary to list as many specialties as possible and make an estimation of the number of specimens and add them in the tender so the external company can take this into account when setting the price. If possible, a solution for these exceptions can be written in the tender.
Hereunder a list of possible problems which can be encountered while imaging and possible solutions:

Documentation: In some cases, documentation is added to the herbarium sheets (with or without paper clips). For example, preparations and notes for a publication, the publication itself or pencil drawings are sometimes added to the specimens. It is up to the herbaria itself to decide if this material needs to be imaged as well or not. At Meise, it was chosen not to digitise this material because it brings extra problems such as adding different barcodes or using one barcodes multiple times, while most of the time this material also covers the label data, barcode and or the specimen itself.

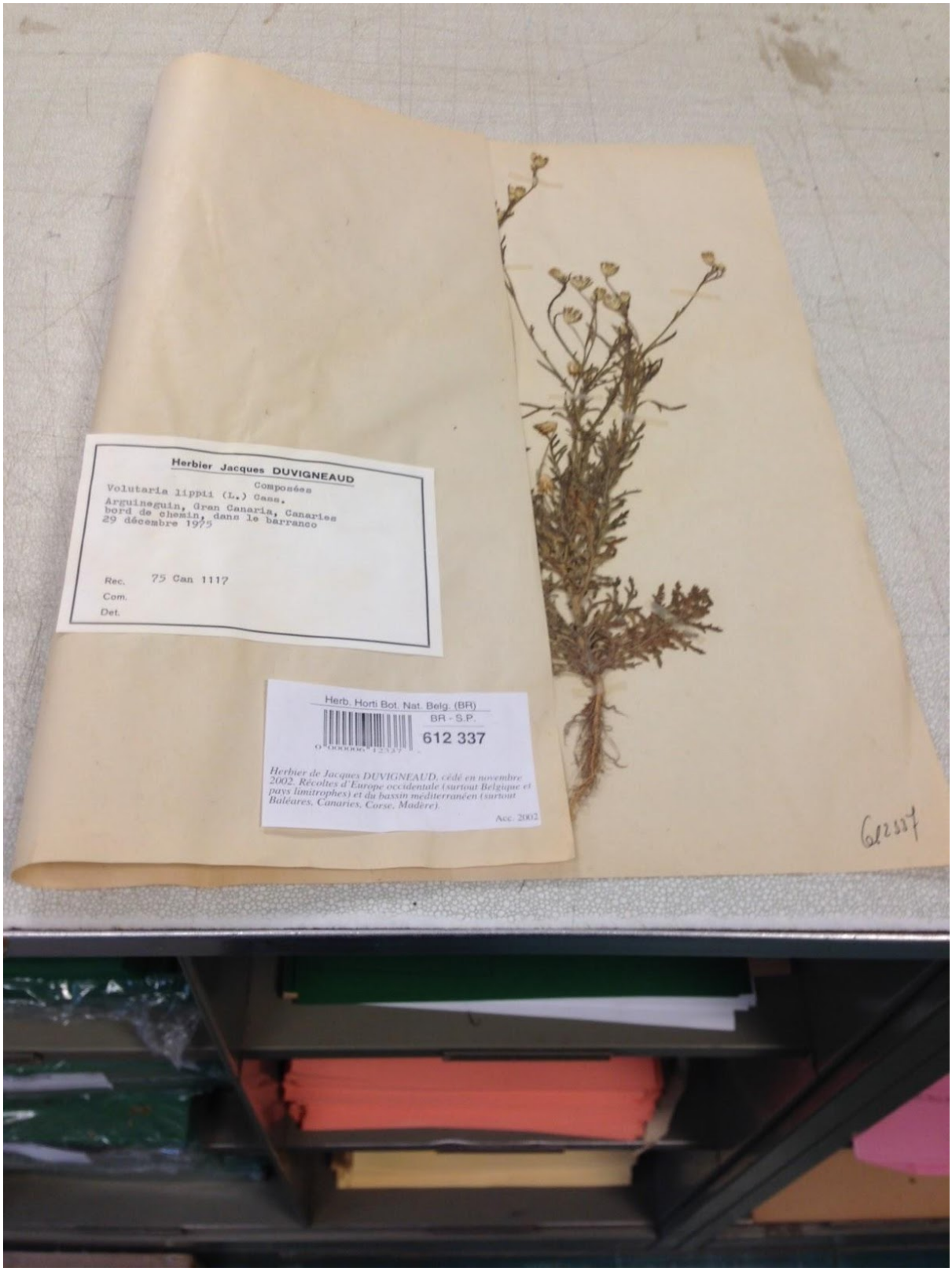
Images: In some herbaria, photos of the plant as it was collected in the wild are added as a second sheet to the herbarium collection. Only in case where there is extra information on the label, these images are digitised at Meise. If not, they are kept aside and will not be imaged again. Another kind of images are black and white images of specimens from other herbaria, which are sometimes the only evidence that is left of this collection. Only if there is label information on the sheet, these images are imaged to make it possible to transcribe the label information.

Sheets with multi gathering: It is up to the institute itself to choose how to cope with this problem. In the first digitisation project at Meise, more barcodes were added during pre-curation and they digitised these specimens themselves before inserting them again in the collection. During the second mass digitisation project, they chose for adding only one barcode per sheet at the conveyor belt and tagging these multi gathering specimens during transcription. After mass imaging, these specimens were retrieved from the collection using a track and trace system of the external company, barcodes were added and sheets were re-imaged and tagged with transcription in house.

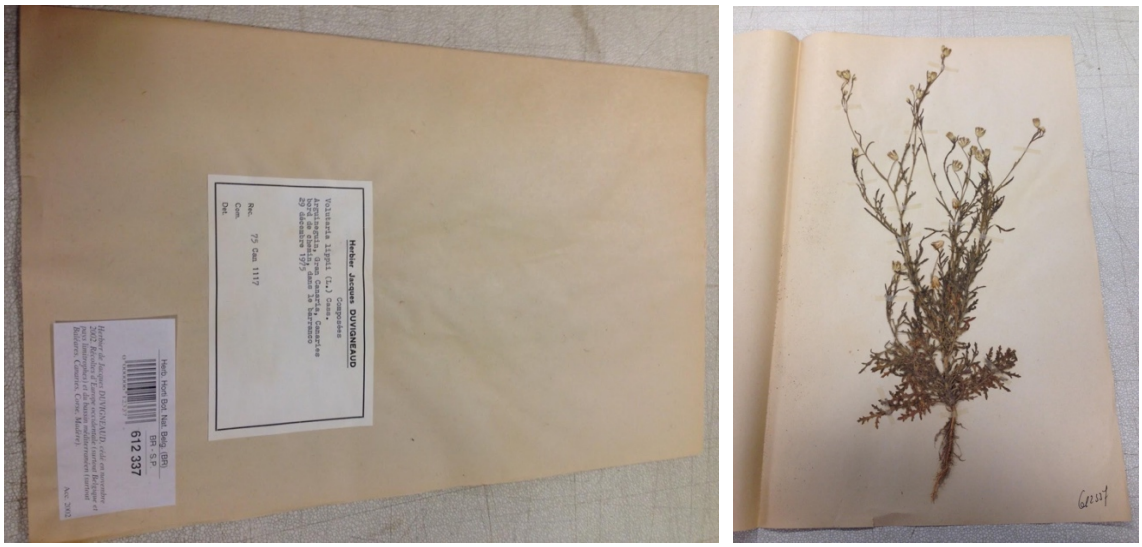
Small size sheets: It occurs that specimens are mounted on relatively small sheets compared to the normal A3 size of the collection. In this case, the external company adds an A3 sheet underneath this specimen so that there is no problem during the quality control afterwards because file size is one of the parameters of the quality control process.

Strangely mounted specimens: It happens that when a new collection arrived in the past, it was mounted like a folder with the specimen inside the folder and the label on the front page and was inserted like this in the collection. It is impossible to image these specimens properly on the conveyor belt. Therefore, the decision was made at Meise that these specimens would be kept aside when they appear at the start of the conveyor belt. These specimens will be remounted properly and will be imaged at the conveyor belt at the end of the project.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483



Specimen that needs to be remounted before digitisation. Labels and barcodes are on the outside of the folder, the specimen inside.

Regular specimen before and after mounting

Paperclips: in the past, different sheets of the same collection were kept together with the use of a paperclip. Or during pre-curation or at the conveyor belt, these paperclips need to be removed before imaging.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Oversized specimens: some collections are mounted on larger sheets than usual. The decision can be made to image all these specimens all at once on the conveyor belt if the size doesn't exceed a certain range (it must still be possible to set up the camera system). For super large collections, the best approach is probably to image them in house. These specimens will be handled after the imaging and transcription by our fixed staff.

Transportation

Even if transport takes place in-house, it is advisable to thoroughly investigate the route from the collection storage room to the digitisation room. Thresholds, corridors, lifts and the everyday flow of employees can get in the way of the transportation of the collection.

Transport within buildings

The advantage of in-house digitisation is that the material cannot be contaminated as quickly, but that also requires the necessary precautions. Windows remain closed and insects can be eliminated with blue lights, not only in the digitisation space, but also on the route to this space. At Meise specimens were loaded onto mobile shelves from the herbarium and unloaded next to the conveyer belt (see case study Meise).

Transport between buildings

As soon as a collection needs to go outside from one building to another, humidity and changes of temperature become a topic to take into account. The, on two sides, open trolleys that were used in Meise, would not fulfil the needs. If the transport has to take place between buildings, a zippered cover can, in this case, offer a solution.
Weather conditions in general should be taken into account during the full digitisation period: long periods of rain may influence the pathway between the buildings; muddy, slippery and may cause high humidity.

Packing

Packing for in house operations has to be adapted to the demands of safety, possible obstacles and the fragility of the collection. Also, the way the herbarium sheets are stored in the collection dictates if further packing is needed. Another topic that may influence the choice of packing is that each specimen can be handled swiftly for the image capture, tailored to fit within the digitisation rate. If there is no need for a temporary storage requirement, traceability become easier as well.
If the specimens have to be digitized in a different building, packing varies according the risk incurred by the transportation. Proper boxes are used to protect the specimens and can be marked to ensure the traceability between the buildings.

Preparation of imaging space

The space has to be designed in a way that the equipment can be safely and swiftly fed. This requires enough space for the staff to move around and be able to take any action in case of malfunctioning.
The space must be suitable regarding the need for electric power and data transmission. As a general regulation, multi-plugs are forbidden and therefore, generally, several electric lines would be necessary. The IT network has to be reliable: failing to send images to the server can slow down imaging.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Some special setting can be necessary to ensure the quality of the images. For example, shutters can be put up in order to make the room as dark as possible. However, this setting has to be compatible with working regulations.

Transport equipment in case of contractor from other country

If the contractor comes from another country, the necessary arrangements should be made to transport the equipment. The most of the time, expensive equipment has to be ensured properly. Also, custom services need to be thought through.

3.5. Possible tasks for mass imaging outsourced by contractor

Contractor site assessment	

Contractor site assessment

As the collection is handed to the contractor, the contract is the only binding document that ensures that the digitisation is duly undertaken. The site used by the contractor has to be checked with a special focus on storage conditions, such as temperature, humidity and integrated pest management. Periodical reports have to be provided by the contractor. The institution must be invited to carry out quality control on the spot, both before the start of the project and during the duration of the project.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

4. Case Studies

4.1. Meise Botanic Garden (DOE! & DOE2!)

At Meise, the preparatory phase of DOE! started in 2015, the mass digitisation of the African and Belgium herbaria. Other herbaria with experience in mass digitisation were visited, the institutions were very clear about the dos and don'ts of digitisation. So, insight on the imaging part of the digitisation was obtained even before the tender was written. Before tendering, also the size of the collection needed to be estimated in order to have a better idea of the possible costs. The method used, is to count 10% of the collection and extrapolated this number to the whole African and Belgian collection.

The location of imaging

The imaging started in June 2016 and was carried out by a Dutch subcontractor, Picturae (https://picturae.com/). After consultation with other herbaria, it was decided to conduct the imaging at Meise in a room next to the herbarium. This in-house imaging has multiple advantages. First of all the collection did not need to be packed and transported and this saved time and money. This also excludes the risk of damage or loss of the collection during transport. The second advantage is that when there are problems during the imaging process, staff is always in the neighbourhood to give support and solve issues. Meise is fortunate to have a room large enough to house the image production line within close proximity to the collection. A local imaging system also has the advantage that the specimens do not have to be frozen or otherwise decontaminated before being put back into the herbarium. Specimens stay in a more or less decontaminated zone and are not more than 24 hours outside the herbarium. An alternative might have been to set up the imaging equipment in the herbarium itself, but this would have meant moving specimens within the herbarium and potentially moving insect contaminations into different parts of the herbarium.

Moving specimens

The Meise collection is housed on three floors, which meant specimens had to be moved by elevator to the imaging conveyor belt. Specimens were loaded onto mobile shelves from the herbarium and unloaded next to the conveyer. For this one person was subcontracted by Picturae. Meise chose to outsource this work because otherwise it would have occupied a full-time technician, who could have been doing more specialist work.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

An example of the trolleys used to transport specimens from the herbarium to the imaging conveyer. A sectioned trolley is used so that each shelf of the herbarium can be transported and imaged separately, then returned to the same shelf in the herbarium

The trolleys have the same number of shelves of the same height as the pigeonholes in the herbarium cabinets. The cabinet number was written on the side of the trolley with a whiteboard marker and on top of the pile of specimens of one pigeon hole a shelf barcode was placed. For track and trace reasons the same shelf barcode was placed in the empty pigeon hole after taking out the first pile of specimens of the top right-hand pigeonhole of the trolley.

Within the two projects one person worked on moving specimens, however, in the second project this person also put cover barcodes on the folders when a taxon name changed. This comes in handy for the transcription of the data and prevented the necessity to scan all herbarium folders which saved money. Meise's staff gave specimen handling training to the

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

conveyor belt operators to ensure the specimens were handled correctly and kept in the correct order throughout the process.

During 13 months 1.2 million specimens passed the conveyor belt, at a speed of 5,000 specimens a day. Each trolley could load approximately 1500 specimens, so at least 3 trolleys a day were processed.

In 2019 a second mass digitisation project started to digitise the remaining 1.2 million herbarium specimens from the General herbarium, the Algae herbarium and the Van Heuck vascular plant collection.

Mounting and maintenance of specimens

For the first project Meise conducted a maximum of pre-curation whereas for the second project they chose to minimize preparational work. The choice for changing the herbarium preparation was made only to reduce the workload on their staff. In DOE!, the first mass digitisation project, the staff had to work in the collection 4 h a day, with a team of 15 people during 18 months. Their tasks were (on top of their regular work) to stick barcodes on each specimen, restored the specimens were necessary, put cover barcodes each time a taxon changed, retrieve everything that was mounted completely in envelopes (for in house imaging), retrieving multi gatherings, remove plastic sleeves, and indicate the sheets that did not need to be imaged e.g. images, literature, type specimens (already imaged in previous digitisation projects), already imaged specimens (loans), by putting them in a white folder with a red dot. All those marked specimens could then easily be distinguished by the conveyor belt operators and could be placed on the opposite side of the conveyor belt in a position where they would not be imaged but they would travel along with the rest of the specimens maintaining their order within the collection. After in house digitisation of all these 'exceptions', these specimens needed to be inserted again in the collection. For the second mass digitisation project DOE!2, Meise decided to minimise the pre-curation as it was too labour intensive for their staff as this work came on top of their day-to-day work, which meant that some of their tasks stopped. Minimal pre-curation in this case means that only specimens that were already imaged during previous digitisation projects were marked by putting them in a white folder with a red dot. On a weekly basis, all technicians came together one morning to mount new incoming collections or to do some restoration of old collections. The operators at the conveyor belt had to stick barcodes on the specimens themselves, therefore two operators in the first project became three in the second. One operator to put the specimens on the conveyor belt, one to stick a barcode on each specimen and one to put everything in the same order back in the folders and to check the first image quality. These operators also had to remove plastic sleeves around the specimens as this reflects the light of the camera. More exceptions occurred at the conveyor belt and more training of the operators was necessary. Some collections were kept aside for remounting (Meise has one particular collection which was mounted as a folder with the label data on the front sheet and the specimen inside the folder). The decision was made to stick only one barcode on each sheet although it is a multi-gathering sheet. The reason for this decision is that it is really hard to decipher (at the conveyor belt) if the collection is a multi-gathering or not. This problem is partly covered during the transcription process were the transcribers marked the possible multi gathering specimens and transcribed the label

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

information. This makes it possible to compile a list of specimens which need to be retrieved again from the herbarium to add extra barcodes and image them in house.

Quality control and IT infrastructure

Because the imaging was done in-house, staff were always available to help when problems occurred. Every night after a day of imaging, the TIFF files and their derivative JP2 and jpgs were transferred by Picturae to Meise. A semi- automated quality control was conducted on all incoming images and a visual quality control of 2% of all images was done by Meise's IT staff and quality controller. Once the images were approved, they were sent to our digital archive (the TIFF files are sent to VISA, the Flemish institute for archiving for long term preservation, the JP2 and jpgs are kept in house for internal use and display on www.botanicalcollections.be, the virtual herbarium of Meise Botanic Garden).

Conclusions

DOE!2 is certainly more efficient in terms of time, efforts and pressure on the team. However, more errors can be made by the external company and much more post curation was conducted, for example on multi gathering sheets, retrieving specimens completely stored in envelopes, overall restoration and curation.
When time and resources are available, the best approach is to prepare the specimens before imaging, so that it is not necessary to go through the collection a second time after imaging. Your imaged collection will also look much nicer when it is restorated before imaging.

4.2. Case Study Royal Botanic Gardens, Kew

At Kew funding was only available for a pilot project to establish the viability and best approach to take for a much larger mass digitisation project. This pilot was completed in 2014 in collaboration with the Natural History Museum in London and the digitisation was carried out by Picturae. With only an estimated number of around 67,000 specimens to image, it was not an option to bring the image production line to Kew or the NHM so the specimens were transported by lorry to the Netherlands.

Specimens were sent in two batches to minimise the risk of loss in transportation and to reduce the space needed to store the boxed specimens at the institution before they were collected or could be frozen and returned to the cupboards. The transport was subcontracted by Picturae to Crown Fine Art. who provided cage trolleys for the specimens that were boxed in-house at each institution in a traceable numbered order.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Condition reports were prepared for 0.1% of specimens, including a photograph of the sheet before transportation and after its return from the Netherlands.
Even though all type specimens at Kew have already been imaged, as they are incorporated within the collection it was decided it would need too much resource to separate them out and then re-incorporate them back into the collection; therefore, the risk was accepted to transport them to the imaging site. However, as the types could easily be distinguished by the conveyor belt operators, as they are contained within folders with red edging, they could be placed on the opposite side of the conveyor belt in a position where they would not be imaged but they would travel along with the rest of the specimens maintaining their order within the collection. Other specimens that have been already imaged are less easily distinguished and so rather than spend time trying to identify these minority specimens it was decided to re-image them.

It was considered necessary to send institutional staff to the Netherlands to ensure the complete consignment had arrived safely at the imaging site by physically counting the boxes and checking the inventory. Curation staff also visually checked specimens against the condition reports when they arrived at the imaging site. Whilst no specimens were found to be damaged in transit the condition reporting checking did require additional resources. However, staff were also needed on site to give specimen handling training to the conveyor belt operators and to ensure the specimens were handled correctly and kept in the correct order throughout the process. Specimens needed to be placed back in the correct folder and the order of folders maintained for a smooth transition back into the collection on return to the institution. Each collection has its own unique issues which need to be explained and close on-site cooperation between institutional staff and the commercial company is vital for success. It is recommended that staff be on site for at least two weeks at the start of the process to iron out any issues and clear decision-making authorities decided in advance. The advantage of imaging at the institution would ensure that staff are more readily available to deal with issues throughout the project.

Upon return to the respective institutions, specimens had to be frozen in compliance with IPM procedures for 72 hours. This is a significant bottleneck due to space considerations and restrictions on how many specimens can fit into the current walk in freezer facilities. For any larger digitisation project offsite with a larger throughput of specimens it would be necessary to hire additional freezing capacity.

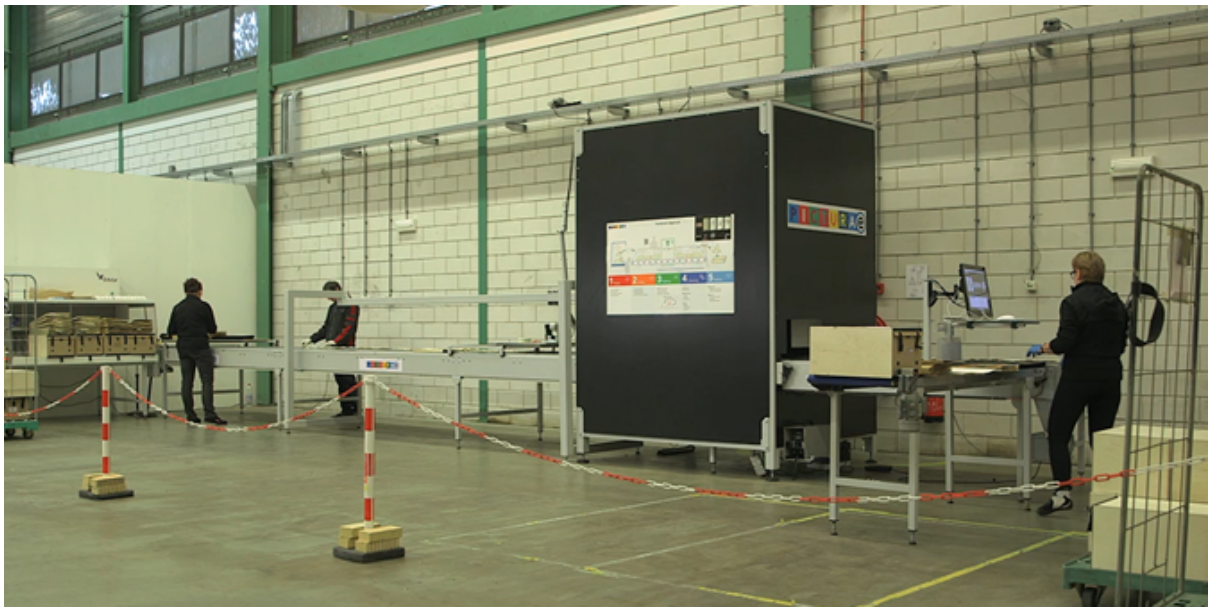
Transportation of material to the Netherlands was particularly labour intensive in the pre- and post-digitisation work involved (condition checking, freezing, packing boxes). As well as curatorial staff time required to complete the additional work, the project incurred direct costs such as transportation, insurance and purchase of boxes. It also increased the risk that the collections were subjected to whilst in transit. For this reason, Kew would prefer to digitise on site in future as this would significantly reduce the cost unless the collection needed to be moved for other reasons e.g. building work.

For this pilot Kew had limited in-house resources so minimal pre-curation was completed on the specimens. It was decided that the folders that contained the specimens would need to be imaged as well, so that the name the specimen was filed under could later on, be

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

captured from the images of the folders. It was necessary to complete this step as the filing name was not always on the specimen label.

The barcoding was to be completed on the conveyor belt however the conveyor belt operators struggled with the identification of more than one specimen on a sheet and did not barcode them appropriately. Therefore, it was decided after the first few weeks of imaging that all sheets would only be given one barcode. The transcribers also found it difficult to flag the multiple specimens correctly on data entry meaning that many multiple specimens would not have been identified in the digitisation process. Evaluation of the costs of the different models showed that outsourcing barcoding and capturing the filing name from images of the folders, was at least 50% more expensive than completing this in-house. A large proportion of these costs relate to the cost of the in-house tidy up of multi specimen sheets, since they need to be reselected from the cupboards, barcoded re-imaged and the labels re-transcribed. Another factor that caused extra costs was the fact that the number of folder images to specimen images was higher than estimated. The ratio of specimen images to cover images was 3/1. In-house folder level information capture and barcoding, prior to the imaging process, needs careful planning and resources to barcode enough specimens in advance of the imaging to ensure that conveyor belt operations are not delayed or stopped due to lack of prepared material.



Working with a commercial company can be a very different working environment than usually experienced by institutional staff and that impact should not be overlooked. Conveyor belt systems are producing up to 5,000 images a day and commercial companies need institutions to make quick decisions for example to approve or reject images. Relevant staff must be available to make these decisions. Staff need to be able to keep up with quality control and have contingency for institutional staff absences e.g. holidays and sickness. As this was only a small pilot, logistics could be more easily handled but for a larger project this has to be considered in more detail. Similarly, in this project images were transferred by hard drive and ingested using formal institutional procedures. However, with

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

a larger scale project transfer and ingestion of images would be a significant consideration as not all institutions will be able to ingest 5,000 images a day.

4.3. Case study Digitarium

(The joint digitisation centre of the Finnish Museum of Natural History and the University of Eastern Finland)

Large-scale digitisation in Finland started in 2008-2009 by creating a national digitisation strategy (Pelkonen et al. 2009). It set goals such as digitising significant parts of all public collections in the next 25 years, and in order to achieve this, developed the necessary research infrastructure for industrial digitisation by 2015. The city of Joensuu was chosen as the site of a new digitisation centre, located in peripheral region where it was possible to get financial support from the EU Structural Funds, and the possibility to work with a university that has as well biology as computing departments. The new research unit, named Digitarium, was placed in a large hall close to the School of Computing where photonics and colour informatics were research priorities. During its existence in 2010-2017, Digitarium received 2.1 million euro from a series of grants from the European Regional Development Fund and the European Social Fund. In addition, Digitarium won 1.5 million euro from competitive FP7 research projects and the national research infrastructure programme, and 0.5 million euro from commercial digitisation services. The workforce of Digitarium consisted of a project manager, IT specialist, two operator/technicians, and a variable number of trainees, interns, and project workers. These developments have been described by Lehtonen et al. 2011; Tegelberg et al. 2012, 2014, 2017; Saarenmaa 2013; and Mononen et al. 2014.

Digitarium focussed on developing methods and technologies for mass digitisation. Inspired by the approach of MNHN in Paris, Digitarium designed and built a conveyor-driven system for herbarium sheets in 2012, and a smaller version for pinned insects in 2014. Both systems had a circular layout, so that only one operator can run them. Objects were placed on trays so that they could move on the circulatory system. The hardware of the conveyor-driven digitisation systems designed by Digitarium consist of off-the-shelf components and costs 40,000-60,000 euro, but software and computing resources, and the installation project double or triple that amount before the system is ready to use.

The herbarium system was used to image the fern collection of LUOMUS (40,000 sheets), a private collection of Compositae (1,500 sheets), and parts of the Oslo herbarium (250,000 sheets). The insect system was used to digitise a Lepidoptera collection and a Coleoptera collection (about 12,000 specimens each). All these collections had to be transported to Joensuu, which did not cause any harm to the specimens. A good mode of operation was acting as a receiving centre for endowed insect collections, which as part of the process of imaging were filed from the cabinets and drawers of the collector to the unit trays of LUOMUS.

Digitarium built an insect digitisation system for LUOMUS in late 2015, and a herbarium digitisation system in early 2016. Both systems are housed within their respective collections. The insect system is being operated by LUOMUS entomology curators in 2-hour

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

shifts (2-4 shifts/day, depending on other work), and 250,000 specimens have been imaged and their data transcribed until now (September 2019). The herbarium system is being used by one barcoder and two operators 6½ hours a day, and has imaged 442,000 specimens until now.

In 2017 the University of Eastern Finland decided to abandon Digitarium. This was largely due to the increased commercial activities and requests to sell hardware and software, which was far from the core business of the university. The research activities of Digitarium (including the ICEDIG proposal!) were transferred to LUOMUS, and the commercial activities spun out to the company Bioshare Digitisation (part of Sertifer Consulting Ltd).

4.4. Case study Naturalis

Naturalis started a large investment project in 2010, the costs of which were largely covered by a contribution from the Economic Structure Enhancing Fund (FES). The project had a threefold operational purpose: transferring a number of large natural history collections to Leiden and integrating these collections within the national natural history collections; production and publication of collection data from more than seven million specimens, and design of a contemporary laboratory for collection-related scientific research.

In the course of 2013, Naturalis entered into agreements with the universities in Leiden, Amsterdam (UvA) and Wageningen regarding, among other things, the transfer of the large botanical and zoological collections that these universities managed. In these agreements Naturalis accepted the ownership of these collections for the State of the Netherlands, making these collections part of the national natural history collection, for which Naturalis has management responsibility. Earlier, in 2010, Naturalis obtained responsibility for the botanical collections that until then were the property of Utrecht University.

These transfers of ownership corresponded to an important operational objective of the FES investment project: the transfer of these collections to collection depots managed by Naturalis and the integration with the other national natural history collections.
Therefore: All storage units - drawers, jars, shelves, etc. - are uniformly provided with bar codes referring to descriptions of the contents.

In the context of the collection digitisation, the label data of more than 8 million individual collection objects were also digitized and high-resolution photos were made, to which the barcodes refer to each storage unit.
Of some of the other objects - more than thirty million - images were made on storage unit level, which makes it easy to see what is in those storage units.

No distinction has been made between the original national collection and the university additions. All assigned barcodes and digitized data and images can be found via central databases and brought into a comprehensive system via location registration.

The barcoding of the storage units can then also be used to trace the movements.

All herbarium sheets have been digitized to object level regardless of origin and are therefore immediately provided with a complete location registration.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

For the remainder, the collections have been digitized at the level of storage units. All storage units are provided with a QR code, so that movements can easily be traced. This has brought the collection registration of the entire botanical part of the national natural history collection to a uniformly high level of findability.

An important problem with the physical integration of collections was that the university herbaria traditionally used widely different formats of herbarium sheets and storage units. Completely integration of such units would therefore entail extensive and difficult to justify interventions: either trimming large sheets, with a high risk of losing scientific and other heritage values, or aligning storage units and cabinets to the largest format, which would result in a considerable extra seizure of square meters. The problem is solved as follows.

The herbaria from Wageningen University and Utrecht University are "frozen", in the sense that no more additions are made. All growth takes place as an extension of the "Leiden" part of the collection. In this way, it is prevented that even more than in the past, a difference in the sizes of storage units is created, while at the same time accommodating the former deviations.

Physical integration has been achieved by putting cabinets from the previously separate herbaria, in which related species are stored, together. This process was completed with the transfer of all herbaria to the new and renovated Darwinweg depots in 2019.

Approach

In collaboration with Picturae Naturalis set up a semi-automated production line for the digitisation of the herbarium sheets. This approach was ground-breaking and in updated versions is still in use for similar collections. A small film can be found on YouTube on the imaging of herbaria on a conveyor belt for the first time during the FES project at Naturalis: https://www.youtube.com/watch?v=hmG4twyHXkE.

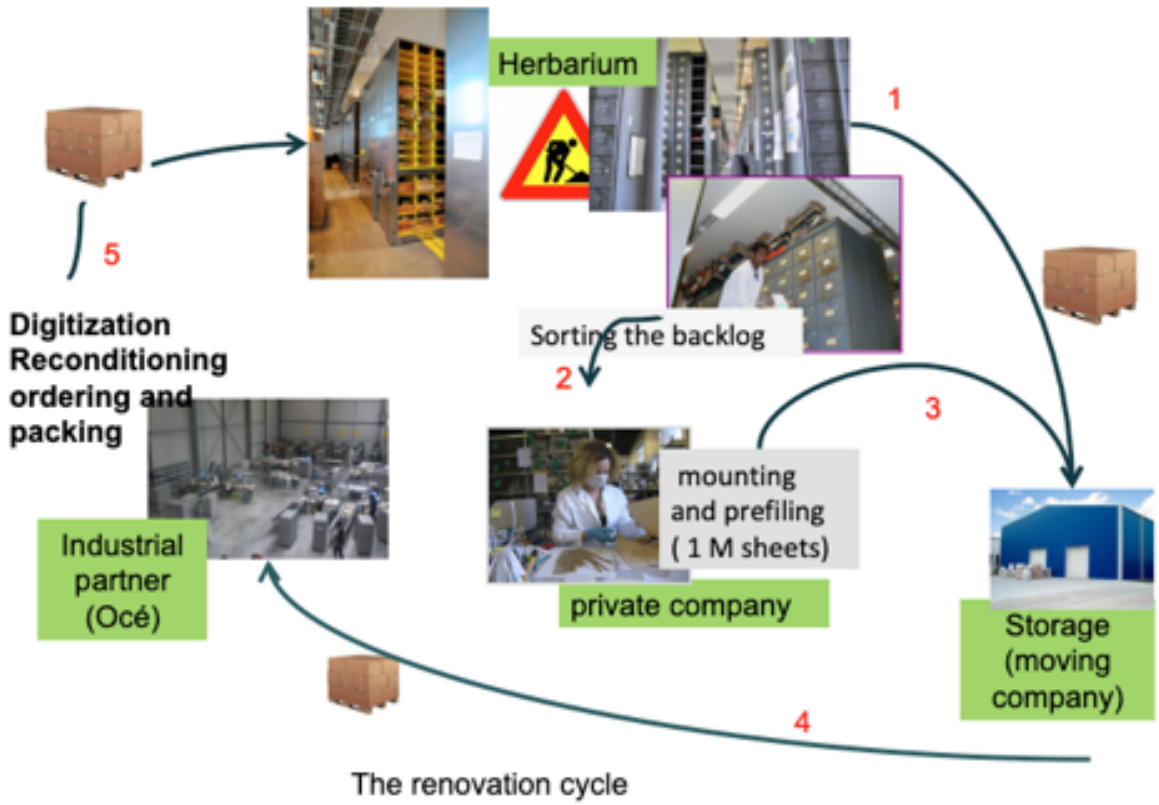
Challenges in the set-up of this project were many, but focused for a large part around health and safety (see 3.2) and logistics. After resolving these initial problems, more than 5 million specimens were digitised (image and data). Naturalis rented an external facility to house the production lines, with the advantage that in-house processes (research, collection management, education) were not hindered by the digitisation process. This outweighed the extra logistic efforts and freezing of the collection.

4.5. Case study Paris

The case of Paris is a mixed case. The digitisation project started in 2008 and ended in 2012. Mounting specimens has been undertaken on site whereas imaging has been done off site, in a building rented especially by the contractor in the outskirts of Paris. Both mounting and digitising have been contracted as two distinct contracts with two different contractors. A third company was contracted for transportation and shelving. The mounting was estimated to 1 million specimens, the digitisation of phanerogams up to 7 million. The whole operation was complex for two reasons: the digitisation was coupled with the renovation of the building. It was the decision by the management to work in the building during the

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

renovation; this decision has been proved a poor one because of amongst others, the nuisance level. The digitisation project was also complex because the way the herbarium was organised was changed at the same time. Previously the collection was organized geographically, with one floor or part of the floor for each continent. The new way to organize the collection was via taxonomy. Planning the digitisation had to wait for the renovation of the building and the sorting of the collection. This complexity required a third site for temporary storage which was provided by the transportation contractor. Too many handling actions came together at one point.



Mounting

Mounting was contracted and carried out in house. The Paris museum provided the contractor with everything needed to do the mounting (paper, glue, etc) and provided the required environment (preparing the space for mounting, internet and telephone facilities). A risk assessment document was set up to deal with health and safety issues, in particular mercury treated sheets, therefore, the contractor's staff received medical monitoring.

Traceability was a key issue in this complex process. The movers carried the specimens to the digitisation facility where they arrived in clearly labelled boxes, 1600 boxes were delivered monthly. Part of them included the unsorted backlog. Boxes received a tracking barcode: the museum provided a "taxonomic file" describing the content of the boxes (box number, family name and APG number, genus name and serial number within family, geographic area). This information was inserted in the contractor's information system (IS) and used along the industrial process (labelling, sorting, quality assurance)

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Herbarium sheets were put in a genus folder using a different colour for the original geographical area. For each genus folder, the operator replaced the old folder with a new coloured one, typed the first letters of the species name and selected the name from the taxonomic list (family, genus, species, authors, ID=taxon number) provided by the museum and finally printed a label with barcode and identification information, which was stuck on the folder.

Specimen imaging

A data matrix and a barcode were stuck on each sheet: the data matrix was stuck by the contractor for its tracking purpose, the barcode was specific to the museum. A small proportion of sheets already had a barcode; this double coding allowed the contractor to handle the sheets independently of the museum's recording. However, a museum barcode was stuck on each sheet with no barcode.

The specimens were placed three by three on a tray, for a conveyor belt, then scanned. As three conveyors belts were running in parallel, at the end of the chain, the data matrix was read, allowing the system to check that all specimens were back in the right folder. The folders were stored in a box before sorting

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Scanning resolution and image format

Specimens were imaged with a bidirectional scanner which produces 11x17" (A3), 300 dpi images. TIFF files were saved offline (one production day per disk of 1 TB) and JPEG's were made for online use. Filenames were generated from the barcode number read on the images.
Data storage was an issue; one TIFF image weighs 50 MB and one JPEG is 5 MB. This compression rate was chosen to have the same level of details as with TIFF (only colour has slightly changed). This choice was a technical-economic trade-off; for 10 million images TIFF files represent 500 TB, whereas JPEG files represent only 50 TB. Database information was not an issue as it represented less than 100 GB

Quality control

60,000 images were produced each week. 1% of the production (ca. 600 images) was randomly sampled for checking. Samples were distributed among botanical staff to check: focus, data quality, barcode number and barcode location on the sheet. Only two times the complete weekly production had to be reprocessed because of rejection by the control quality team.

Post-operation control. Once the specimens were back into the herbarium, anther random sampling was undertaken to check that the specimens were at their right place and information was complete and correct.

Results

20 people worked on two shifts. The planned rate was 17,000 sheets per day over 24 months. This number was never achieved mainly due to the lack of resource (space and staff) devoted by the contractor to the sorting phase which created a bottleneck that could not be suppressed. Therefore, only 2 conveyors belts did run in parallel instead of 3. Besides, some decrease in the daily production was created by bottle necks after the image capture that let the contractor to shift staff from the image capture to the sorting of specimens.

5. Conclusion

The conveyor belt system has been developed, tested and over the years improved towards an effective tool for mass-digitisation of herbarium collections. Over time, the data capture equipment and conveyors equipment have evolved, the principles of the workflow remain solid. As was shown in above described mass digitisation projects, this digitising method brings the entire collection to an unquestionable level of information.

The planning phase and the workflow design are critical; preparedness is the major issue. Preparedness has many aspects; institutional and governmental policy, H&S, circulation, mounting, specific to each institution. Preparing time is generally underestimated, as well as the fact that collection managers in general know their collection less well than they think; numbers are often over- and underestimated, differences in barcoding, sizes of sheets and other exceptions occur often during the start of the process.
Setting up a workflow is another time-consuming aspect, but a correct method also yields a lot, both in qualitative and quantitative outcome. The parameters that influence the

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

efficiency of the digitisation are numerous and various. The case studies show that though every collection is slightly different the main parameters should be checked. If done well, there is an effective digitisation project for every herbarium.

It is recommended to combine the digitisation of a collection with other planned handling actions, such as re-allocating or (re)mounting. In this way, handling the collection multiple times can be avoided. Though too many actions in one workflow may cause mistakes. This can be overcome with a clear division of tasks and responsibilities and evident agreements on work order.

The choice for in-house by own staff, in-house by contractor or off-site by contractor depends on the availability and suitability of space, available staff, whether the process needs trained staff or not and the sensitivity of the collection. These parameters are balanced with costs, though costs certainly should not be the only lead in the discussion.

Imaging process can improve the digitisation processes, e.g. full databasing can be done from images providing that quality control checks have been sufficient to ensure the exact correspondence between the physical specimen, the associated information and the image. Promising new techniques such as OCR and HTR can soon be used to automate this process. Label segmentation is crucial for these techniques.

6. Recommendations

General

The expected rate of production for a cost-efficient digitisation is around 20,000 specimens per week. Actual numbers reached depend on the number of conveyor belts, number of employees and exceptions in the collection. The constant feeding of the conveyor belt is evident to reach numbers and reduce costs.

Combining imaging with other actions, such as relocating, packing is highly recommended (holistic approach). However, if too many topics are brought together, without developing a thorough plan on forehand, the holistic approach might work against you.

A digitisation project is complex and it is necessary to devote a significant amount of time in designing the workflow and identifying potential bottlenecks. Any project should be accompanied by training of the staff.

Mounting of specimen should be a separate project set up prior to, but preferably aligned to the digitisation project. However, if for any reason (time, cost, conservation issue), mounting specimens is part of the digitisation process, mounting and digitizing rates must be related to one another. In case of enough time and resources, the best approach is to prepare all specimens before imaging, so that it is not necessary to go through the collection a second time after imaging. This will also improve the result of the imaged collection.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Traceability

In any case an audit trail for traceability should be set up on forehand. Traceability of as well specimens, boxes and pallets is ongoing from the moment the collection leaves the collection space until all material is back in place. Barcodes or other unique markings should be implemented besides the identification of specimen by the institution.

When different collections are brought together, it is recommendable to start with a new number system, where old numbers can be kept next to new numbers.

Checkpoints at every stage of the process must be put in place as they do not represent a significant cost.

Staff

Training of staff is necessary in all projects, institutional staff must be trained to understand the technical process, while institutional staff must train the conveyor belt operators on specimen handling to ensure the specimens are handled correctly and kept in the correct order throughout the process. The more exceptions occur at the conveyor belt the more training of the operators is necessary.

The decision-making authorities have to make clear decisions before the project starts, nevertheless, there will be amendments. It is recommended to implement these amendments in the first two weeks of the project and after that stick to the decisions made. This ensures that, in case of major issues, only the very first batches have to be redone.

The impact of different working environments for commercial companies and institutional staff should not be overlooked. Conveyor belt systems are producing up to 5,000 images a day and commercial companies need institutions to make quick decisions for example to approve or reject images. Relevant staff must be available to make these decisions.

Staff need to be able to keep up with quality control and have contingency for institutional staff absences e.g. holidays and sickness, this needs to be organised.
Similarly, transfer and ingestion of images would be a significant consideration as not all institutions will be able to ingest 5,000 images a day.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

References

Bedford, D. J. (1999). Vascular plants. In: Carter, D. & Walker, A. (eds). (1999). Chapter 3: Care and Conservation of Natural History Collections. Oxford: Butterworth Heinemann, pp. 61 - 80.

Coughenour, C. (2018) Inside Brazil's National Museum on Google Arts & Culture. Blog post at https://www.blog.google/outreach-initiatives/arts-culture/inside-brazils-national-museum-google-arts-culture/ Accessed 2019-09-17.

Dillen, M., Groom, Q., Hardisty, A. (2019) Interoperability of Collection Management Systems. ICEDIG Deliverable 4.4. https://doi.org/10.5281/zenodo.3361597

Florian, M. E.1986. The freezing process: Effects on insects and artifact materials. Leather *Conservation News*, 3:1–17.

Gilberg, M., & Brokerhof, A. (1991). The control of insect pests in museum collections: the effects of low temperature on Stegobium paniceum (Linneaus), the drugstore beetle. *Journal of the American Institute for Conservation*, 30(2): 197–201.

Grimé WE & Plowman T (1986) Type photographs at Field Museum of Natural History. *Taxon*, 35: 932–934.

Lehtonen J, Heiska S, Pajari M, Tegelberg R & Saarenmaa H (2011) The process of digitising natural history collection specimens at Digitarium. In: Jones MB & Gries C (Editors) *Proceedings of the Environmental Information Management Conference 2011 (EIM 2011)*. September 28-29, 2011. Santa Barbara, CA. 87–91. University of California.

Mononen T, Tegelberg R, Sääskilahti M, Huttunen MA, Tähtinen M & Saarenmaa H (2014) DigiWeb - a workflow environment for quality assurance of transcription in digitisation of natural history collections. *Biodiversity Informatics*, 9: 18–29.

Nieva de la Hidalga, Abraham, van Walsum, Myriam, Rosin, Paul, Sun, Xianfang & Wijers, Agnes. (2019, September 30). Quality Management Methodologies for Digitisation Operations. Zenodo. http://doi.org/10.5281/zenodo.3469521

Pelkonen V-P, Saarenmaa H & Laurenne N (toim.) 2009. Luonnontieteellisten museokokoelmien digitointi. Strategia ja toimintasuunnitelma 2010-2015. 47 s. Luonnontieteellinen keskusmuseo. (Only in Finnish.)

Saarenmaa H (2013) How much of the legacy collections can be digitised? Chapter 22. In: Hardisty A, Roberts D, The Biodiversity Informatics Community. A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecology*, 13:16.

Tegelberg R, Haapala J, Mononen T, Pajari M & Saarenmaa H (2012) The development of a digitising service centre for natural history collections. In: Blagoderov V & Smith V (Editors). No specimen left behind: mass digitisation of natural history collections. *ZooKeys*, 209: 75-86.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483

Tegelberg R, Mononen T & Saarenmaa H (2014) High-performance digitisation of natural history collections: Automated imaging lines for herbarium and insect specimens. *Taxon*, 63(6): 1307-1313.

Tegelberg R, Kahanpää J, Karppinen J, Mononen T, Wu Z & Saarenmaa H (2017) Mass digitisation of individual pinned insects using conveyor-driven imaging. In: Hereld M (Editor) High throughput digitisation for natural history collections. 2017 IEEE 13th International Conference on e-Science (e-Science 2017). Auckland, New Zealand, 24-27 October 2017. 5 p. DOI:10.1109/eScience.2017.85

Owen D, Groom Q, Hardisty A, Leegwater T, van Walsum M, Wijkamp N & Spasić Irena (2019). Methods for Automated Text Digitisation. Zenodo. https://doi.org/10.5281/zenodo.3364502

Phillips S, Dillen M, Groom Q, Green L, Weech M- H & Wijkamp N (2019). Report on New Methods for Data Quality Assurance, Verification and Enrichment. Zenodo. https://doi.org/10.5281/zenodo.3364509

Dillen M, Groom Q & Hardisty A (2019) Interoperability of Collection Management Systems. Zenodo. http://doi.org/10.5281/zenodo.3361598

Woodward FI (1987). Stomatal numbers are sensitive to increases in CO2 from pre-industrial levels. *Nature*, 327(6123), 617.

Funded by the Horizon 2020 Framework of the European Union
H2020-INFRADEV-2016-2017
Grant Agreement No 777483