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 **D3.4 / R / PU**
Work package **WP 3 /** Lead**: PIC**
Delivery date M18

ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

D3.4 State of the art and perspectives on mass imaging of liquid samples

Myriam van Walsum, Agnes Wijers, Pepijn Kamminga, Steven van der Mije, Karen van Dorp

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

Abstract

The topic of this report is applications of conventional photography and automated imaging of specimens stored in liquids with a special focus on the labels. No automated imaging solutions were found for these types of specimens, which is due to the complexity of this storage type. Specimen imaging by removing it from the container is time-consuming and not applicable for mass-imaging. The main goal of jar and other container imaging is to capture the 'old' label for subsequent data entry and recording of the container.

Imaging labels for data entry and databasing purposes might be achievable in some liquid collections. Distortions, discolouration and cloudiness of the liquid and container can be complicating factors.

We recommend a process based approach of record creation with minimal data, combined with label imaging. This can be done rapidly with minimally trained workers. Especially when data entry is partial or includes interpretation, the images provide potential for future use.

Key recommendations

- ➔ Health and safety always needs to be considered for chemical substances around humans and electric equipment.
- ➔ Opening containers is slow and laborious: not suited for a mass digitisation project.
- ➔ Consider very closely whether combining with curatorial tasks increases or decreases efficiency. Flagging containers for maintenance is a good option.
- ➔ Detailed imaging of specimens out of their containers for research questions is not suited for a mass digitisation project. There are too many variables.
- ➔ Imaging is mostly relevant to capture label data. The labels may be floating freely, attached to inside or outside of container.
- ➔ Often the label is not legible from a single photo due to reflections and curving of container, impeding data entry. A composite image can help with this.
    - ◆ NHM and Picturae are running tests on a solution.
- ➔ Potential imaging solutions need to take into account: reflections, inertia, distortions from glass. Moving camera (or multiple cameras) instead of moving container solves inertia issue.

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

Contents

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

Introduction

This task investigates applications of conventional photography and automated image scanning of both specimens and specimen labels stored in liquids. Due to overlap in content, this report will often refer to ICEDIG deliverable 3.3 (Van Walsum et al 2019). The following points were part of the task description and will be discussed in this report[1]:

1. What is the definition and demand of mass digitisation?
2. What are the current and near-term solutions?
   2.1. Case studies
3. What are the effects of scaling up imaging capacity to millions of specimens a year?
4. What is the efficiency gain of combining imaging and digitisation with curatorial processes like repacking, barcoding, cleaning and sampling?
5. What is needed to minimise risks and maintain high throughput while keeping costs low?
6. What is the information required by researchers?
7. Which existing automated imaging methods fit these collections compared to the information required by 1 and 5?
8. Is imaging beyond 2D useful?
9. Adept recommendations with regard to
   9.1. criteria,
   9.2. specifications,
   9.3. technical solutions,
   9.4. workflows,
   9.5. outsourcing,
   9.6. health & safety

"Given that aquatic species are some of the most threatened under impending climate change (DeWalt et al. 2005, Shah et al. 2014) and their museum records exist primarily in fluid collections, digitization efforts of freshwater species are critical for identifying species and habitats most at risk." Mendez et al 2018, p50

[1] The numbers are used in the headings to refer to the sections where the topic is discussed.

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

Current state of digitisation of liquid preserved specimens

Liquid samples are present in all life science collections. Generally, almost all the different taxonomic collections have their own liquid collections, storage protocols and digitisation status. Because the portion of the collection preserved in alcohol is often poorly differentiable from the dry collection in the database, it is difficult to obtain actual numbers of specimens preserved in liquid. Even more so to get numbers of digitised and imaged samples, together or separately.

Responses to a survey by Synthesys3 in 2017 show that 11 of 22 institutes have no capacity for digitising spirit collections. One institute responded to have 100% of their spirit collection databased, but less than 1% imaged. Two other institutes indicated that some of their spirit collections have been databased and imaged, but that they can't separate the numbers from the overall of the taxonomic group. All other institutes that said to have the capacity for digitising liquid collections indicated that about 1% had actually been digitised.

> *In 2010, the Atlas of Living Australia approved funding for a digitisation project at the Australian Museum and the South Australian Museum. While focussing on entomological collections digitisation through volunteers, extra funding allowed the investigation of digitisation procedures for other collections such as liquid samples (Australian Museum final report 2011, p1).*

Effects of scaling up and efficiency gain of combining with other tasks: minimising risks while maintaining high throughput and low costs [#3&4&5]

A significant amount of liquid samples contains more than one specimen per storage unit. This can be many vials in a larger jar individually numbered, representing separate collection events or it contains simply many individuals with a single registration number (i.e. bulk number). The vials often contain many individuals as well. A different category of liquid collections is the preservation of soft body parts or stomach contents of specimens preserved in the dry part of the collection. In this last case these samples may or may not have the same registration number and be digitised.

Imaging of liquid collections can essentially be divided into two categories: whole jar/container imaging, or specimen imaging by removing it from the container. The main goal of jar and other container imaging is to capture the 'old' label for subsequent database entry or text recognition algorithms. Further, the condition of the container can be recorded, and it may give an impression of the specimen. Due to the distortions,

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

discolouration and cloudiness of the liquid and container these images are not suitable for any metric analysis.

Labels are of importance for data recording purposes. These can be found in various locations: top of jar, fixed to outside, fixed to inside or free-floating, attached to specimens or in vials in a larger container. There can be multiple labels present on various locations. Some labels won't be legible without opening the jar. For those labels that can be read without opening, distortion from or reflections of the glass can pose problems.

Imaging of liquid specimens comes with several impediments (Hurst & Siebert 2005). For example:

- Glass containers can be fragile and have sharp edges around breaks.
- The liquids may be various chemicals such as ethanol, isopropyl alcohol, formaldehyde or formalin. These are flammable, volatile and harmful chemicals which need extra precautions such as fumigation hoods and gloves.
- Larger and older containers are often sealed with wax bladder or vaseline, making them very laborious to open, especially in a mass imaging workflow.
- The way a jar is sealed can hold some important clues to its origin, age, or other historical information. Assessing and documenting all these aspects is very laborious and time consuming.
- Potential DNA or other contamination when moving specimens from one container to another requires strict protocols.
- Wet specimen imaging is restricted in time because of the drying out of the specimen.
- Combining liquid sample digitising and imaging with necessary maintenance in some form can be efficient, with a minimum of flagging units in need of maintenance.
- Liquid preserved specimens may have become distorted due to improper preservation or too small containers. Straight from the container, the distorted and reflective specimen may be of no great interest to photograph. Some mounting may be necessary to fix the specimen, or parts like fins, in a flat position. This can be done with pins, but never by placing the pins through the specimen (Hurst & Siebert 2005, p270).

All these considerations make imaging of wet specimens taken from their container time consuming and expensive and not easily amenable to mass digitisation, so it is often only done on very select specimens needed for specific research purposes. This is one of the reasons dry imaging of liquid preserved specimens Taylor (2005, p146) advises that small specimens can be photographed while submersed in a small amount of liquid, but that the required amount of liquid needed to do this for larger specimens is not an ideal situation around electrical items.

The size range of individual fish ranges continuously from sub-centimeter to greater than 2 meters, but the majority lies between 5 to 30 centimeters (Hurst & Siebert 2005).

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

Fish are often distinguishable based on visual differences (p265). Fish shape varies greatly, with many being relatively flat laterally, and symmetrical so that imaging from a single view may be sufficient. Others are more three dimensional in shape or not symmetrically flat (like stingrays), so that multiple views are required to capture all traits (p268).

See ICEDIG deliverable 3.3 for further considerations on the effects of scaling up on risk management and throughput rates (Van Walsum et al 2019).

The role of imaging in mass digitisation [#1]

Two first questions arise with the design of any digitisation project with an imaging component. The decision whether or not to image, as data entry from the object is in many cases faster than imaging, and what to image; researchers and other external users need to drive digitisation priorities. This is also valid for deciding which specimens to image and which details need to be included. See ICEDIG deliverable 3.3 for considerations on the role of imaging in mass digitisation (Van Walsum et al 2019). Imaging of liquid preserved specimens in a mass workflow is difficult for the above reasons. Imaging of labels for data recording is the main point of interest.

Information required by researchers [#6]

As extinction of populations and species proceeds, natural history collections are bound to provide the primary source for new research. The 21st century is marked by big data: connecting complex and large data sets. Data needs to be searchable through good descriptors to be fully made advantage of, while also acknowledging the long history of biodiversity collections, meaning that standardisation and pollution of data is present. During new imaging efforts, it should be tried to plan ahead so that this new dataset can be adequately searchable. See ICEDIG deliverable 3.3 for considerations on information required by researchers (Van Walsum et al 2019).

Usefulness of imaging beyond 2D [#8]

The most commonly used 3D technique is (micro)CT due to interest for the internal parts that are often associated with this preservation type. Most 3D techniques are slow, which poses a risk for drying out of the specimen as well as deformation during imaging, resulting in a useless scan. The Muséum national d'histoire naturelle (MNHN) comments that CT scanning of specimens while still in liquid is possible, but their movement needs to be fixed with cotton balls or plastic tubes to obtain a good scan result. 3D for liquid preserved specimens can’t be imagined in a mass workflow at this point.

See ICEDIG deliverable 3.3 for further considerations on the usefulness of 3D imaging for three-dimensional specimens (Van Walsum et al 2019) as well as ICEDIG deliverable D3.7

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

(Nieva de la Hidalga et al 2019a) for further information on 3D techniques and considerations for integration in a digitisation project.

Current and near-term solutions [#2.1]

Notwithstanding all the difficulties and challenges surrounding the digitisation and imaging of liquid preserved specimens, several attempts and projects have tried to find solutions.

Case study 1: The Naturhistoriska riksmuseet (NRM, Swedish Museum of Natural History)

The Naturhistoriska riksmuseet (NRM, Swedish Museum of Natural History) responded to the Synthesys survey on digitisation capability that less than 10% of the zoological collection has been imaged, while being unable to differentiate between dry and liquid preserved specimens. This is still more than what ca. 20 other institutes report in the same survey.

Dr. Delling, Collection Manager Zoology at NRM, clarifies that of the approximately 350.000 specimens in the zoology collection, ~30% has been cataloged. The fish collection lies far above this average at 80%, and the herpetological collection is at 35%. Both of these contain a large amount of liquid preserved specimens. Birds, mammals and invertebrates are at 75%, 10% and 17% respectively, but how much of these percentages are stored in liquid is hard to estimate. Invertebrates make up ~73% of the collection, fish 19%, herpetology 6%, with mammals and birds the remaining ~2%.

A lot of pictures have been taken for various purposes over time, ranging in quality from scanned old slides to high-quality stacked photos. The pictures are stored on an image server and are linked to each sub-collection's database: over the next years digitisation capacity will fall caused by the migration to a new database. Digital loans through pictures, especially for insects, are common. Dr. Delling notes that the imaging workflow of wet collections is not that different from dry collections: all sources, including specimen, old labels, old catalogs are captured. No shortcuts are taken, meaning that mass-digitisation is not applicable.

Case study 2: Muséum National d'Histoire Naturelle (MNHN)

The Muséum National d'Histoire Naturelle (MNHN) wet collection imaging protocol describes various procedures, based on the condition of the specimen. These protocols were developed by the e-ReColNat team from the marine invertebrate department at MNHN.

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

Specimens in good condition are taken out of their containers, are wiped dry to limit reflectivity and placed on glass plates above a black surface with the camera facing down on a copy stand.

Photography setup for dry imaging of liquid preserved specimens: the specimen is placed on the glass pane, which is raised from the background, to ensure a blurred background. The camera is facing down on a copy stand and two lights are on either side. (Image copyright MNHN)

Damaged, fragile or small specimens are photographed in their containers/tubes, placed on a black background to limit handling. Sometimes small and bristly specimens are placed in liquid in a petri dish, against a black background. Sealed containers are not opened, and these specimens are photographed frontally, against a black or white background.

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

Small specimen photographed in petri dish. (Image copyright MNHN)

Small specimen photographed in original vial. (Image copyright MNHN)

Imaging station for sealed containers: monitor with live view and camera left, specimen and background right. Black paper is used to limit lighting to relevant parts. (Image copyright MNHN)

Composite image of sealed jar, photographed frontally from 3 sides. Note the free floating label in the right part. (Image copyright MNHN)

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

Case study 3: Naturalis[2]

Starting in 2012 as part of a large digitisation project, Naturalis set up a digitisation team to digitise 100.000 jars across several collections: mammals, crustaceans, Arachnida, Trichoptera and Mollusca. The goals were to establish uniformity in registration and imaging, increase discoverability of specimens and data, and in the process merge three major collections with those of Naturalis. Per collection the workflow was slightly adjusted to the protocols in use in that particular collection. At the end of the project 114.427 units were digitised and imaged, with a storage level barcode label and object barcode label for each jar.

They chose to keep maintenance and digitisation separate. A digitisation team of two team leads and a digitiser were responsible for attaching new labels with matrix code, data entry and imaging, and they flagged containers in need of maintenance for collection staff to process. During digitisation the containers remained closed.

Before entry at the start of the production line, the specimens were sorted alphabetically by the curator, updating taxonomic identifications and making sure the jars were clean and safe for handling. However, in some cases the digitisation team also flagged leaking jars or drying/dried out material. Operators identified bulk storage jars which required a specific workflow.

To deal with the specific hazards of this material detailed work instructions specifically described potentially unsafe situations such as leaks, spillage, and dropped jars, and the required steps to ensure the safety of persons and specimens. At the start the production line was located in a dedicated lab with fume hoods and chemical disposal sinks, but due to limited space and minimal arising issues this lab was made available for other tasks and the production line was moved to the adjacent lab.

Data entry existed of minimum fields for most specimens and full data entry for type specimens. Data entry for non-type specimens was limited to genus, species, locality and storage unit. Two types of linked records were created: one for storage unit, one for sample, as bulk containers can contain many small vials each constituting a sample of a separate collection event. Each record was tied to a separate label with data matrix code. The labels were printed on self-adhesive archival paper, specially tested for use in spirit collections.

The container was imaged to capture additional information for data entry at a later time. Oversized jars were digitised and imaged in storage. Depending on collection, one or more images were captured, sometimes only of the labels. Criteria for imaging were: type material, more data on label than recorded during minimum registration, old material, poorly legible labels.

The photos were captured at medium, yet fine, image quality. The camera was connected to a computer and operated from there. This way the focus point was set

[2] Sources used: Naturalis 2013, 2015, 2018. Heerlien et al. 2015.

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

manually on a label for autofocus to use. Data and image quality were first checked by a team lead, and then by the curator.

Supply of new material was a bottleneck, as well as adjusting camera settings/framing of object/lighting. Also, there were issues with a lagging database. Existing records with conflicting data were also a slowing factor.

Record creation of bulk jars is always problematic. Only one photo of the bulk jar was made and the jar was registered as storage unit. In case of an existing registration number the unit was attached to the first record of the jar. In other cases registered as storage unit with multiple records and if no registration number was available as storage unit with a new number. One lesson learned is that linking the image to a storage unit will be the best solution. For one specific subcollection it was known which registration numbers were in a bulk jar, which helped a great deal. In most cases the information on the label attached to the jar was all that could be registered, with an estimate of the number of specimens in the jar.

Case study 3: Inverse panoramas of jars

Due to the limited value of closed container imaging for capturing the specimen and the inefficiency of opening containers and laying out the contents for imaging as described above, the only workable mass imaging workflow focusses on capturing label data. The benefits of capturing all possible label data, with the option of including the specimen as well, are described in the ICEDIG D3.3 report on vertebrate imaging (Van Walsum et al 2019, section Criteria p25-27 and section Process based approach for fast but basic digitisation p31-33). When labels are so large that they are not legible from a single photo because they curve around the jar, then multiple photos are required and data entry is impeded by having to switch between multiple photos. For this reason, a solution of stitched images would be interesting so that data entry can be done from a single composite image. The principle is that of a landscape panorama picture but pointed inwards instead; hence inverse panorama. After stitching the images, the output would be a virtually rolled out image of the container.

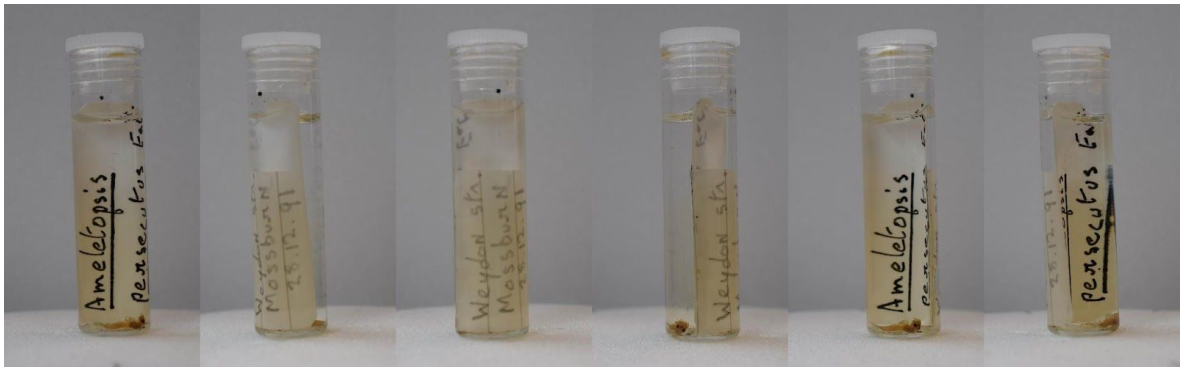
This solution has to account for the various possible curves of containers and aim at legibility of the data. Placing the object on a controllable turntable would be simple, but is likely to cause difficulties with processing due to differential movement of label and specimen relative to container due to inertia. To avoid the inertia issue, either multiple cameras or a camera on rails would be advisable.

The Natural History Museum[3] in London is testing with photography of a vial on a rotating stage with a single camera. Currently, the output is a single composite image of a number of photos cropped down to the vial, so that each part of the label data is visible in at

[3] Contributors: Steen Dupont and Ben Price, NHM.

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

least one part of the composite image. In the future, they plan to test the output with their ALICE software for label extraction (Price et al in press).



Current output from test at NHM.

Picturae in Heerhugowaard has conducted a number of tests with recording 360 degree inverse panorama views of cylindrical objects, using jars in two different sizes and with different recording equipment; (1) a DSLR camera with a tilt shift lens, (2) a mobile document scanner, and (3) an iPhone.

Nikon 800D camera with tilt shift lens

A Nikon 800D was used with a tilt shift lens (Schneider PC TS Makro-Symmar 90mm f/4.5). Two sizes of jars (height X cross section) 125mm x 90mm and 110mm x 50mm) were placed on a stepper-motor turntable, to get a precise rotate and step sequence of 36 high-res photos (600dpi) from each jar. The images were cropped into vertical strips, dissecting out the vertical strip exactly facing the camera and then stitched together as a 2D image.

Both the bulk creation of strips and the stitching per recording can be fairly easy realised via ImageMagick[4]. The width of each strip however depends on the number of frames per rotation, the size of the object and the distance of the camera to the object, so full automation will be difficult. Semi-automation per set-up with a number of jars that have more or less the same size, can be realised.

[4] https://imagemagick.org/

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

Output from test at Picturae with Nikon 800D camera. Left: the captured images. Right: the stitched crops of the center of each frame.

This principle can also be applied in a purely mechanical way where no processing is needed: rollout slit-scan photos, a process that has already been explored for a longer time in the cultural field. The principle of slit scan photography is to create 'strips' in a mechanical way. Panning the camera and long exposure make the camera record what happens behind the slit. The image of the slit passes through the viewfinder over time, thereby illuminating one side of the frame at a different time than the other side of the frame. More tests need to be done to understand if this method would be applicable in an automated set-up.

Mobile document scanner

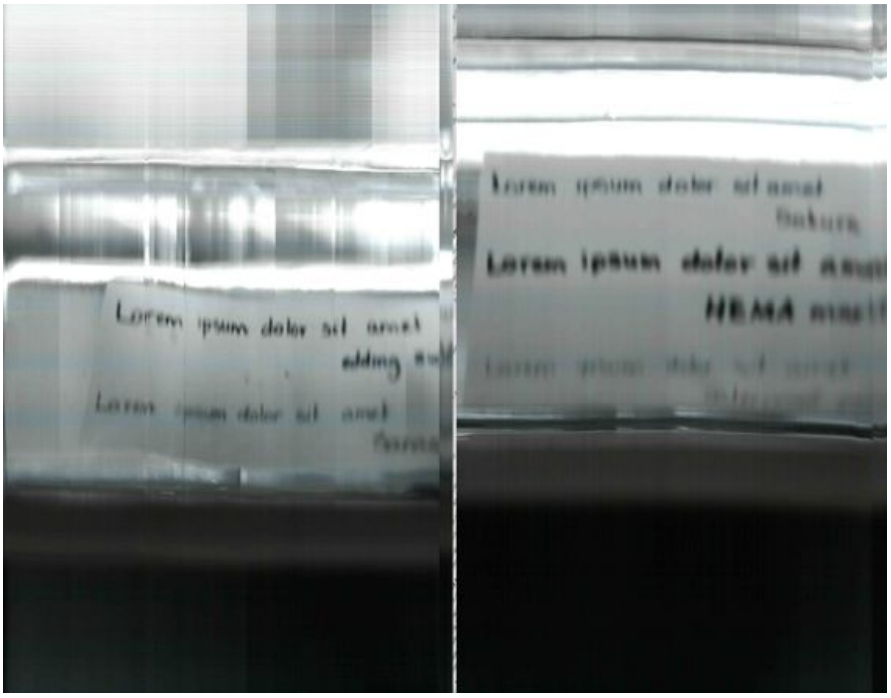
The above mentioned method works with fragile jars on rotating surfaces, therefore a solution was tested with a flexible scanner where you can choose to move the scanner around the jar or to fix the scanner and rotate the jar.

For this test the Renkforce W4S "Wireless Edition" mobile document scanner was used,

The advantage of these types of scanners is their microSD memory card slot and integrated WiFi in combination with Apple-equipment, which implicates that it can be used in e.g. collection rooms and thus reduce the transport of fragile objects.

However, the result of this recording was poor, due to the fact that the depth of field of a line scanner is too minimalistic. The shape of the jars did not make it possible to hold the line scanner against the glass and label inside the glass. This will also be common in practice. For jars with a completely smooth surface this solution could work.

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



Output from test at Picturae with line scanner. Both attempts have not led to the desired result.

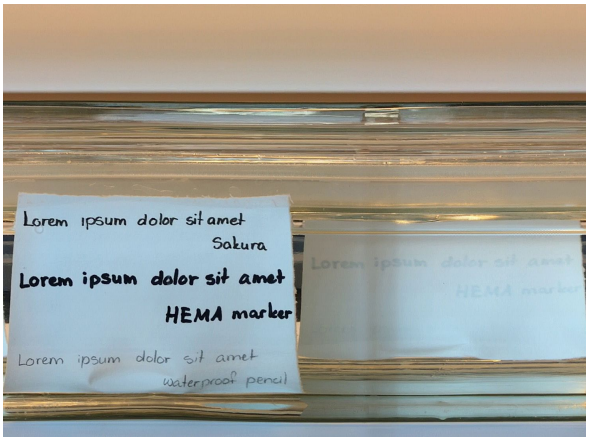
iPhone

The findings with the line scanner made the researchers realise that a line scanner principle in a remote set-up could work. (increase the horizontal resolution in a shorter time /more images per rotation). The iPhone shoots a video of the object as it is rotated 360 degrees. This results in a video with +/- 2500 frames, which have been converted with ffmpeg to individual images. From each image a vertical strip of 1 pixel from the center was taken and all strips were stitched, creating the image.

In future experiments and official recordings, it is of course desirable to work with a video camera that gives higher resolution images. However, it is remarkable that this simple camera already leads to these results.

By recording more frames than the number of steps the turntable takes in one rotation, 1-pixel strips can be used to create the image. When using a lens with sufficient depth of field, practically all jars can be processed without adaptations to settings.

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



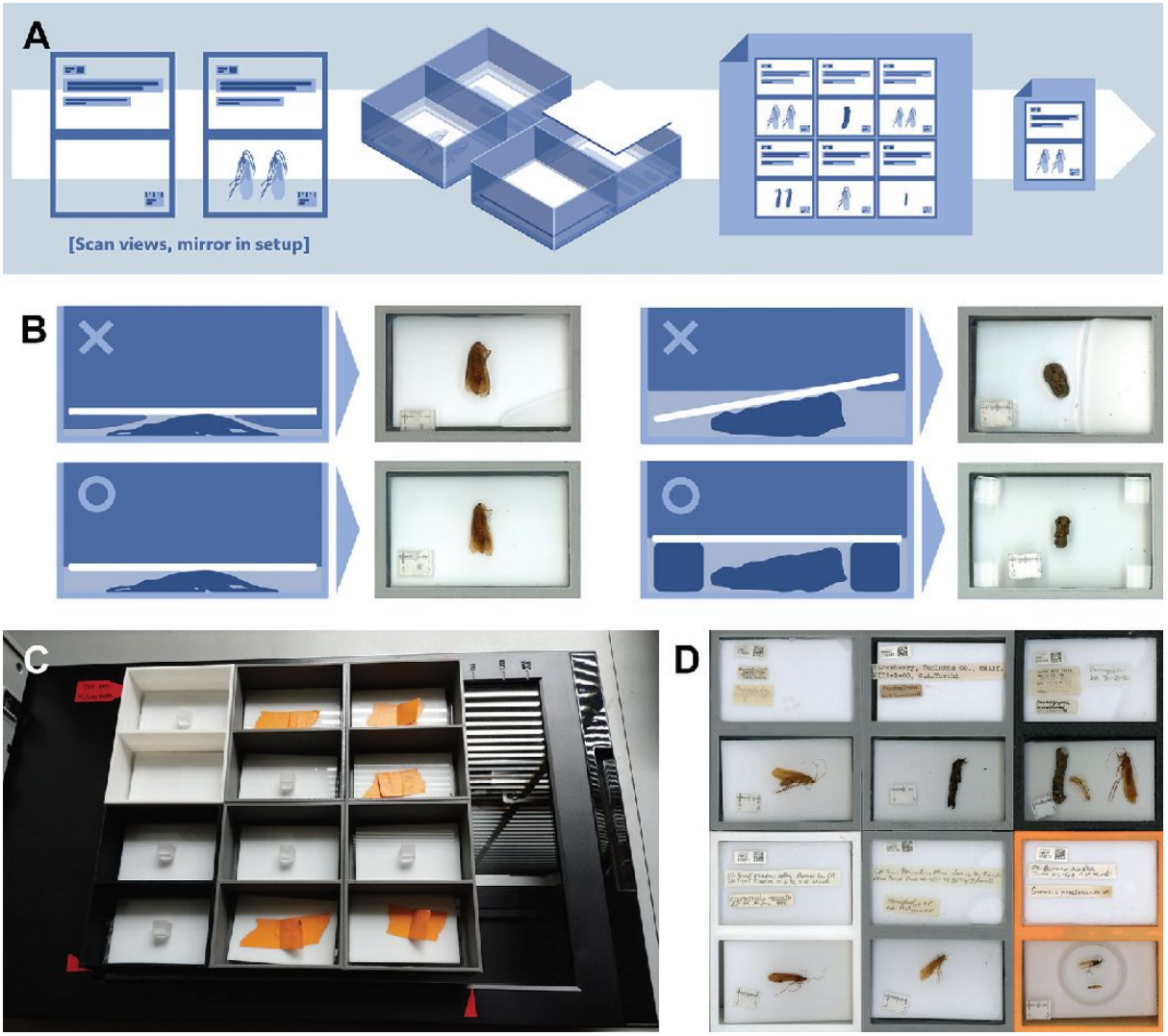
Output from test at Picturae with iPhone. By using only 1 pixel, the back of the label also comes out legibly, in mirror image.

Case study 3: Flatbed scanning combined with 3D printing

A new approach to digitise wet specimens outside of their container was developed using a flatbed scanner and 3D printed containers (Mendez et al 2018). Where camera-based imaging of specimens in liquid suffer from surface reflections and lens distortion, these are non-issues for flatbed scanners because each point is scanned as the scanner is moved across the object. It does come with its own limitations: cleanness of scanner glass and slide glass impacted the scan quality. Larger specimens resulted in darker images and dark bodied specimens lacked detail.

The required container was based on standard double-wide microscope slides (75*50*1mm) with a custom designed 3D printed box. The design consists of a double compartmented container: one for the specimen, one for the labels. Slide and 3D printed box were attached with silicone adhesive to make the whole leak-proof. The labels were placed text side down and covered with Plexiglas to make them sit flat. Specimens were arranged to not overlap and a 10mm scale bar was included. Specimens and dissections were grouped together inside a ring and for thicker specimens struts were used to keep the Plexiglas background from rocking. Several materials were tested (BPA, white/black/orange unfinished/polished nylon and white/black/grey PLA) for leakage and discolouration after which they only recommend white nylon for use with ethanol which remained stable during the 1.5 year duration of the test.

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



Scanning process images: A scanning process with scan view setup, box covers, full scan, and cropped image B required fill levels of fluid for box covers to prevent bubbles and standoffs (left) to prevent an uneven scan (right) C six boxes set up on scanner; and D resulting full scan. Clear Plexiglas cube standoffs B and slip-joint rings D (lower-right) improve scans. (From Mendez et al 2018, figure 3)

While imaging of wet specimens outside their containers is unlikely to ever achieve speeds to be considered mass-imaging, the system was designed so that a batch of 6 specimens could be imaged at once. One operator manned the scanning station and two operators unloaded and loaded batches. The scan rate was approximately 7 minutes/vial/operator, from unloading through rehousing including curatorial tasks. The speed limiting step was the scanning (including arranging and preview) when using good quality resolution settings.

By using the same containers for many specimens after each other, it is important to have strict protocols to empty and clean the containers for each new specimen. Small parts may stay behind and get disassociated or even contaminate the next batch. Besides that, DNA contamination is also a risk.

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

Recommendations

For further considerations on criteria, workflows, outsourcing or in-house, on-site or off-site see ICEDIG deliverable 3.3 (Van Walsum et al 2019). For health and safety considerations, see Effects of scaling up and efficiency gain of combining with other tasks: minimising risks while maintaining high throughput and low costs [#3&4&5].

Specifications and technical solutions [#9.2 and #9.3]

The imaging of liquid preserved specimens comes with a number of particular specifications for potential technical solutions. The first needs to be safety when working with various chemicals electrical equipment. ext, since we have established in this report that mass imaging of specimens taken from their containers is unfeasible, we will focus on specifications for closed containers and specifically label imaging. These containers are often made of glass or other transparent materials, which are characterised by challenging distortions and reflections. Design of solutions need to work with various sizes (from as small as a vial to containers that are over a meter tall) and shapes (round, oval, rectangular). When stitching of photos is considered, the issue of reflections can potentially be solved in post-processing, otherwise special lighting can be used. This can include very diffuse lighting or cross polarisation, in which both the camera and lights are covered with polarising film. To obtain multiple photos of a single container for stitching, it is not advisable to move the container, due to the differential movement of label and specimen relative to container caused by inertia.

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

Conclusion and discussion

As demonstrated above, mass digitisation for the type of collections that are discussed in this report is not very feasible. This does not mean that digitisation of large numbers of objects is impossible, but efficiency must be sought in process and numbers, and not in hardware. This can be approached from either the process or the demand. Both are discussed in ICEDIG deliverable 3.3 (Van Walsum et al 2019, p31-35). Depending on the situation, one approach may be more relevant than the other.

In summary, a lean process is suggested in which minimal records are created (registration number, taxonomic identification, storage location, geographical region) which are initially only associated with the images. The images can be used in a separate phase, potentially at a distant point in time, for the next step in data entry. Many digitisation projects do not capture full label data; by imaging everything it is ensured that all information is captured, even if not searchable. This also allows verification of data without the need to access the specimen. It is acknowledged that in some cases imaging can be so difficult or data entry so easy, that this is not always the most efficient option. Nelson et al. (2012) also found that data entry from the images was generally the most efficient process for (large scale) digitisation.

A second approach was described which considers the demand for imaging and resources for data entry and option for a mass workflow, supported by two decision trees to help design the digitisation project. At the very least, these discussions will help think about how an institute's imaging programs can be designed.

In this report we have described tests for an imaging solution which would result in a single composite image (or, a virtual rollout or inverse panorama) of the container, so that all label data can be read from a single image. This composite image can then be used for manual data entry and the developing fields of OCR and HTR.

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

Glossary

- OCR (optical character recognition): mechanical or electronic conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, or other.
- HTR (handwritten text recognition: also HWR (handwriting recognition). Ability of a computer to receive and interpret intelligible handwritten input from sources such as paper documents, photographs, touch-screens and other devices. The image of the written text may be sensed "off line" from a piece of paper by optical scanning (optical character recognition) or intelligent word recognition.

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

References

Australian Museum (2011). Rapid digitisation project: final report. https://www.ala.org.au/wp-content/uploads/2011/10/Australian-Museum-digitisation-project-final-report.pdf Accessed 12-03-2019.

Heerlien, M., Van Leusen, J., Schnoerr, S., de Jong-Kole, S., Raes, N., & Van Hulsen, K. (2015). The natural history production line: an industrial approach to the digitization of scientific collections. Journal on Computing and Cultural Heritage (JOCCH), 8(1), 3.

Hurst, P., & Siebert, D. (2005). Imaging type specimens of fishes at the Natural History Museum, London. In: Häuser et al. (eds.): Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information: 41-55. Stuttgart.

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

[to do: add D3.3: Van Walsum et al 2019, incl DOI]

Naturalis Biodiversity Center (2012). Procesbeschrijving digistraat Alcohol.

Naturalis Biodiversity Center (2015). Uit het depot, op het web: twee eeuwen nationaal natuurhistorisch erfgoed in het digitale domein.

Naturalis Biodiversity Center (year 2018?). FES NCB Naturalis Zelfevaluatie.

Nelson, G., Paul, D., Riccardi, G., & Mast, A. R. (2012). Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys*, (209).

Nieva de la Hidalga, A., Rosin, P. et al (2019a). ICEDIG deliverable 3.1. Quality control methodology for digitization operations. (Will be available at https://icedig.eu/content/deliverables Q4 2019)

Nieva de la Hidalga, A., Rosin, P., Sun, X., Van Walsum, M. and Wu, Z. (2019b). ICEDIG deliverable 3.7. Rapid 3D capture methods in biological collections and related fields. (Will be available at https://icedig.eu/content/deliverables Q4 2019)

Price, B., Dupont, S., Allan, E. L., Blagoderov, V., Butcher, A., Durrant, J., Holtzhausen, P., Kokkini, P., Livermore, L., Hardy, H. & Smith, V. (2019 in press). ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation.

Synthesys3 (2017). D4.5 - Digitisation on demand - a report on feasibility of a digitisation on demand service for natural history collections.

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

Taylor, H. (2005). A photographer's viewpoint. In: Häuser et al. (eds.): Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information: 41-55. Stuttgart.

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