



ICEDIG.EU

Innovation and consolidation for large scale digitisation of natural heritage

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 D3.1

Work package **WP 3** / Task: **T3.4** / Lead: **CU**

Date: 31/09/2019

Quality Management Methodologies for Digitisation Operations

Deliverable 3.1

Abraham Nieva de la Hidalga, Myriam van Walsun, Paul Rosin, Xianfang Sun, Agnes Wijers,



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



Description of Work ¹	
Project:	ICEDIG
Work Package:	3
Work Package Title:	Imaging and other physical extraction
Task Number:	T3.4
Task Title	Quality control
Deliverable Number:	D3.1
Deliverable Title:	Quality control methodology for digitisation operations
Deliverable Description:	A report on the quality control methodologies and associated best practice workflows and tools for different collection types to be applied during and after the imaging process.
Deliverable Deadline:	Month 21
Deliverable Deadline Date:	September 2019

Revision Log

Version	Content	Contributors	Reviewers
0.1	<ul style="list-style-type: none"> • Introduction • Quality Management Criteria • Quality Management for Digitisation of Microscope Slides • Quality Management for Digitisation of Skins and Vertebrate Materials • Quality Management for Digitisation of Liquid Preserved Specimens • Quality Management for Digitisation of Pinned Insects • Quality Management for Digitisation of Herbarium Sheets • Quality Management for 3D Digitisation of NHC • Quality Management for associate • Conclusions • Glossary 	Abraham Nieva de la Hidalga (CU), Myriam van Walsun (Picturae), Paul Rosin (CU), Xianfang Sun (CU), Agnes Wijers (Picturae)	
0.2	<ul style="list-style-type: none"> • Revision of Section 7 		Sofie De Smet (MBG)
0.3	<ul style="list-style-type: none"> • Revision of Section 7 		Sarah Phillips (RBGK)
0.4	<ul style="list-style-type: none"> • Revision of all sections, addition of executive summary 		

¹ p. 9-11, ICEDIG Description of Activities [46]



Abstract

This report describes a set of image quality management methods aligned with six kinds of digitisation workflows covered by the ICEDIG project [46]: (1) Microscopic Slides, (2) Skins and Vertebrate Material, (3) Liquid preserved specimens, (4) Pinned Insects, (5) Herbarium Sheets, and (6) Rapid 3D Digitisation of Natural History Collections. The image quality management methods corresponding to each type of digitisation workflow are presented in separate sections. Each section describes: the quality standards that need to be observed, the digitisation workflows, the quality assurance activities to be performed during digitisation (focused on verifying that production processes deliver standard compliant products), the quality control activities that are performed after digitisation (focused on verification of products after processing), and the software and hardware tools which are used to support these quality management activities.



Executive Summary

Digital specimens composed of images and metadata are considered viable research surrogates for the physical specimens from which they have been derived. The images which form part of digital specimen objects can have various degrees of quality, however images with the highest quality are more useful for research, because different kinds of uses and derivatives can support different purposes. This deliverable describes recommended quality management methods which support producing high quality images. The recommendations in this document complement those produced for the metadata associated to the specimens and should support the creation of research quality digital specimens for each of the collection types studied by the ICEDIG project.

The report presents generic quality management criteria which was developed for the digitisation of printed materials (e.g. pictures, documents, and books) which apply to some extent to collections materials which are close to 2D representations (such as microscope slides, and herbarium sheets), as well as to supplementary materials commonly digitised with specimens (labels, physical registries/catalogues, field/lab books, etc.). These are complemented with recommendations for digitisation of collections containing specimens which are 3D objects.

The recommendations are organised in line with the six kinds of digitisation workflows covered by the ICEDIG project [46]: (1) Microscopy Slides, (2) Skins and Vertebrate Material, (3) Liquid preserved specimens, (4) Pinned Insects, (5) Herbarium Sheets, and (6) 3D Digitisation of Natural History Collections. Each section describes: the quality standards that need to be observed, the digitisation workflows, the quality assurance activities to be performed during digitisation (focused on verifying that production processes deliver standard compliant products), the quality control activities that are performed after digitisation (focused on verification of products after processing), and the software and hardware tools can be used to support these quality management activities.

The report can be repetitive since many of the criteria and recommendations overlap, however, the report is designed this way to produce self-contained recommendations for each of the workflows and minimise the need for going back and forth between sections.

This report will need to be updated and extended, because new techniques and equipment are constantly being used for digitising NHC collections.



Contents

Contents	4
1. Introduction	7
2. Quality Management Criteria	7
2.1 Digitisation Standards	7
2.1.1 Image Resolution	8
2.1.2 Image Colour	8
2.1.3 Image Elements.....	9
2.1.4 Image Quality Criteria	11
2.2 Digitisation Workflow	11
2.3 Quality Management in Digitisation	14
2.4 Tools.....	14
2.4.1 Hardware	14
2.4.2 Software.....	15
3. Quality Management for Digitisation of Microscope Slides	16
3.1 Microscope Slides Digitisation Standards	16
3.1.1 Image Resolution and Colour Standards.....	18
3.1.2 Full-Slide Scans Image Elements	18
3.1.3 High-Resolution Slide Scans Image Elements	19
3.1.4 Example.....	20
3.2 Microscope Slides Digitisation Workflows.....	21
3.2.1 Metadata Required for Digitisation	22
3.2.2 Digitisation equipment	22
3.3 Microscope Slides Quality Management Methods.....	23
3.4 Examples	23
3.4.1 Full-Slide Digitisation.....	23
3.4.2 Full-Slide/Selective High-Definition Digitisation	25
4. Quality Management for Digitisation of Skins and Vertebrate Materials.....	27
4.1 Skins and Vertebrate Materials Digitisation Standards	27
4.1.1 Image Resolution and Colour Standards.....	29
4.1.2 Image Elements.....	29
4.2 Skins and Vertebrate Materials Digitisation Workflows	32
4.2.1 Metadata Required for Digitisation	33
4.2.2 Digitisation equipment	34



4.3	Skin and Vertebrate Materials Quality Management Methods.....	34
4.4	Examples	34
4.4.1	Fast Mass Digitisation Workflow.....	34
4.4.2	Research Quality Imaging (High-Definition) Workflow	36
5.	Quality Management for Digitisation of Liquid Preserved Specimens	37
5.1	Liquid preserved specimens Digitisation Standards	37
5.1.1	Image Resolution and Colour Standards.....	39
5.1.2	Image Elements.....	40
5.2	Liquid preserved specimens Digitisation Workflows	42
5.2.1	Metadata Required for Digitisation	44
5.2.2	Digitisation equipment	44
5.3	Liquid preserved specimens Digitisation Quality Management Methods.....	45
5.4	Examples	45
5.4.1	Fast Mass Digitisation Workflow.....	45
5.4.2	Research Quality Imaging (High-Definition) Workflow	46
6.	Quality Management for Digitisation of Pinned Insects	47
6.1	Pinned Insect Digitisation Standards	48
6.1.1	Image Resolution and Colour Standards.....	50
6.1.2	Image Elements.....	51
6.2	Pinned Insects Digitisation Workflows.....	53
6.2.1	Metadata Required for Digitisation	55
6.2.2	Digitisation equipment	55
6.3	Pinned Insect Digitisation Quality Management Methods	55
6.4	Examples	55
6.4.1	Whole Drawer Digitisation Workflow	56
6.4.2	Conveyor Belt-Based Digitisation Workflow	56
6.4.3	Manual Single Specimen Digitisation Workflow	57
7.	Quality Management for Digitisation of Herbarium Sheets	58
7.1	Herbarium Sheets Digitisation Standards	58
7.1.1	Image Resolution and Colour Standards.....	58
7.1.2	Image Elements.....	59
7.1.3	Example.....	61
7.2	Herbarium Sheets Digitisation Workflows.....	62
7.2.1	Metadata Required for Digitisation	63



7.2.2	Digitisation equipment	64
7.3	Herbarium Sheets Quality Management Methods.....	64
7.4	Examples	64
7.4.1	Inhouse Digitisation	65
7.4.2	Outsourced Digitisation	65
8.	Quality Management for 3D Digitisation of Natural History Collections	67
8.1	3D Digitisation Standards.....	67
8.1.1	3D Model Quality Management Standards	70
8.1.2	Model Elements	70
8.2	3D Digitisation Workflows	71
8.2.1	Metadata Required for 3D Digitisation	73
8.2.2	Digitisation equipment	74
8.3	3D Digitisation Quality Management Methods	74
8.4	Examples	75
8.4.1	ZooSphere Workflow	75
8.4.2	CSIRO-ALA Workflow	76
8.4.3	Smithsonian 3D Digitisation	77
9.	Quality Management for Digitisation of Associated Collection Materials... Fout! Bladwijzer niet gedefinieerd.	
10.	Conclusions	79
	Glossary	80
	References	82



1. Introduction

This deliverable describes recommended quality management methods which support producing high quality images. These images, in turn, form part of digital specimen objects. Digital specimens composed of high-quality images and metadata are considered viable research surrogates for the physical specimens from which they have been derived. The recommendations in this document complement those produced for the metadata associated to the specimens and should support the creation of research quality digital specimens for each of the collection types studied by the ICEDIG project.

Quality management methods can be subdivided in two main areas: Quality Assurance (QA), and Quality Control (QC) [21, 97]. QA covers process-based activities designed to ensure that products or services are produced or delivered according to a predefined set of quality criteria. In the ICEDIG context, QA activities are expected to be performed by the digitisation team (digitisation technicians, in-house digitisation teams, and the service providers). Implementing QA activities is expected to minimise the need for re-digitisation, improve processing times, and support collection management.

QC covers product-oriented activities designed to guarantee that products or services conform to a predefined set of quality criteria. QC activities are expected to be performed by the data curation and publishing teams (curators, collection managers, data publishers and data users). QC activities are expected to guarantee that the digital assets produced are fit for publishing and use.

In general, the definition of QA and QC benchmarks can serve to define the criteria to manage the expectations of digitisation processes and guide the acquisition of equipment.

2. Quality Management Criteria

This section describes the quality management criteria for imaging-based digitisation of collections in general, the criteria is derived from different fields, but directly applicable into the digitisation of Natural History Collections. After this, subsequent sections present the same structure, but introduce special considerations which vary depending on the collection type.

2.1 Digitisation Standards

The history of the digitisation of Natural History Collections can be traced back to early efforts for creating computer-based versions of museum catalogues [42]. Advances in digital imaging technology, internet connectivity, computer software for manipulating digital images and image compatible database software, spurred the increased inclusion of digital images of type specimens in online accessible databases [95]. The decrease of digitisation cost and the emergence of new methods has driven a change from imaging only type specimens to imaging entire collections [46, 49].

As the initial target of digitisation were the museum catalogues, several of the early standards were derived from best practices of digitisation which were initially designed for the digitisation of Libraries and Cultural Heritage Collections, such as those from the US Federal Agencies Digital Guidelines Initiative (FADGI) [31], Metamorfoze [52], the US Library of Congress [61], and the UK National Archives [66]. These standards were mostly developed for the digitisation of printed materials (e.g. pictures, documents, and books), and may apply to some extent to collections



materials which are close to 2D representations (slides, packets or herbarium sheets). However, these standards do not satisfy the needs for all digitisation products derived from different collections, particularly those containing specimens which are in fact 3D objects (e.g. vertebrates, pinned insects, liquid contained specimens, seeds, and shells). Moreover, some of the recommendations are dependent on the digitisation equipment available at the time when they were formulated and may not account for higher resolutions provided by modern equipment. For this reason, the recommendations provided in this document should be taken as the minimum and, whenever available, more stringent criteria should be used (e.g. higher resolution), bearing in mind that the main target is producing higher quality digital specimens which can effectively be used for research.

As an example, Table 1 illustrates some of the standard criteria for digital specimens of herbarium sheets (see section 7 for further details). The important issue highlighted with this table is that the quality characteristics of images are dependent of their intended use, display medium, storage format and licensing details. The relation between the standards and the quality management is directly dependent on the intended use of the product being verified. In a QA scenario, the standards can be used to verify products as they are being produced. In a QC scenario, the standards can be used for evaluating finished products. The following subsections describe the general quality management criteria, and then each workflow section will elaborate on them to include the criteria for the specific types of collections and target uses.

Table 1 Example of Image Quality Criteria for Herbarium Sheets (this same table is included in Section 7).

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy**
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Specimen features	400 PPI	24-bit colour		$\Delta E < 5$
Research on specimen	420 PPI to 600 PPI*	24-bit colour		$\Delta E < 5$
Preservation	450 PPI to 600 PPI*	24-bit colour		$\Delta E < 5$

* Minimum resolution recommended, if digitisation devices available allow for higher resolution, that resolution should be used.

**Delta E (ΔE , δE , dE) is a metric for understanding how the human eye perceives colour difference.

2.1.1 Image Resolution

The main characteristic defining the resolution of a digitised specimen is the number of pixels per inch (PPI). FADGI [31], the US Library of Congress [61], and the UK National Archives [66], recommend scanning photos and high value artefacts at 600 PPI. FADGI and the US Library of Congress recommend 600 PPI as the minimum resolution for research purposes but agree that lower resolutions may be used if the resources for producing them are not available.

2.1.2 Image Colour

Colour depth (a.k.a. bit depth) refers to either the number of bits used to indicate the colour of a single pixel, in a bitmapped image or video frame buffer, or the number of bits used for each colour component of a single pixel [107]. In the latter cases of FADGI [31] and the US Library of Congress



[61] recommend 24 bits as the minimum (eight bits for each colour component). Additionally, FADGI establishes Adobe RGB as the recommended colour space [31,32,52].

2.1.3 Image Elements

Image elements refer to visual elements which can appear next to the specimen and which are intended to help in the identification, processing, and quality control of the digital specimens. There are eight elements that need to be considered for imaging NHCs: (1) Specimen, (2) Background, (3) Colour Chart, (4) Scale Bar, (5) Labels, (6) Barcode, (7) Institution Name, and (8) Other Elements. Table 2 provides an overview of the elements to capture for different types of specimens, for more detail refer to each of the specific workflow sections.

Table 2 Overview of image elements

Imaging Workflow	Specimen	Background	Colour Chart	Scale Bar	Labels	Barcode	Institution Name	Other Elements	Conditions
Microscopy Slides	C	C	NR	NR	C	R	NR	O	<ul style="list-style-type: none"> Specimen can be hard to capture without special equipment due to size. Background is generally white to facilitate viewing of slide elements. Labels can be placed on both sides of the slide, requiring additional images per slide. Special type labels are important for classifying specimens.
Skins and Vertebrate Material	R	C	C	C	R	R	O	O	<ul style="list-style-type: none"> Background must maximise the identification of the specimen, avoiding glossy or reflective materials that can hinder border detection. Some specimens may not require colour chart as colour is not a main feature (e.g. bone samples). The placement of the scale needs to consider the depth and angle at which images are acquired.
Liquid preserved specimens	C	C	O	O	C	C	O	C	<ul style="list-style-type: none"> The containers can contain more than one specimen and require additional handling. Sometimes the specimens are removed and imaged outside the container, but this take longer time. Background must be neutral, especially for see-through containers. Barcodes can refer to one container with multiple specimens. Labels can be hard to image due to the shape and placement in container. Paper records which describe the specimens in a container will need to be digitised as well.
Pinned Insects	R	C	R	R	R	R	O	O	<ul style="list-style-type: none"> Background must maximise the identification of the specimen, avoiding glossy or reflective materials that can hinder border detection.
Herbarium Sheets	R	NR	R	R	C	R	R	O	<ul style="list-style-type: none"> Labels can be hard to image due to overlapping with other labels and with the specimen parts. Additionally, some labels can be placed at the back of the sheet, requiring additional imaging.
3D Specimen Models	R	C	C	NR	C	C	O	O	<ul style="list-style-type: none"> Background must maximise the identification of the specimen, avoiding glossy or reflective materials that can hinder border detection. Colour charts can help if model includes colour or texture. Labels may be captured separately, but some cases such as pinned insects, 3D scanning may help in rapid imaging while minimising specimen handling. Barcodes can be part of the label set

R = Required C = Conditions NR = Not required O = Optional



Specimen is the main element to image. Some specimens may require more than one image to display all relevant features, while others can be harder to image because of their preservation or fragility (e.g. liquid samples, microscopy slides).

Background is an element which needs special consideration since it will determine how well the specimen features are displayed. It is necessary to avoid backgrounds which may prevent distinguishing the borders of the specimen. Similarly, glossy or reflective backgrounds should also be avoided.

Colour Chart is recommended for helping with quality control and postprocessing [106]. The information which can be derived from the colour chart can help in verifying the lighting, white balance and colour accuracy of the image. FADGI, the US Library of Congress, and Synthesys3, recommend the use of colour chart (referenced as colour target or colour checker) [31, 54, 55, 61, 79]. There are many types of colour charts, and examples of many of them have been used by different institutions in their digitisation pipelines. However, modern targets such as those from Image Science Associates are preferred over legacy targets (Colour Control Patches from Kodak) because they were developed for digital image creation and are made to very tight tolerances [18, 53]. Object level targets of this type include a ruler and can be used for verification of colour, sharpness and scale.

The **Scale Bar** is recommended to enable the calculation of the dimensions of the specimen [35]. Synthesys3 acknowledges the use of scale bars in different digitisation workflows [54].

Labels are commonly part of the specimen either placed next to the specimen in the storage container or next to it. Labels are generally used as sources for deriving specimen metadata. Synthesys3 indicates that among the digitisation of natural history collections the clear capture of labels is important for further processing and documentation of the digital specimens [54].

Barcodes are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as unique identifiers which are important for further documentation and linking of the physical and digital specimens [54]. There are different types of barcodes available. Line Barcodes (one dimensional) were used in some digitisation projects, however, there were some misidentification problems (because of line features in the specimens). Consequently, current guidelines promote the use of two-dimensional barcodes, and they are being adopted for new projects [118].

Institution Name (or seal) is required to keep track of the institution holding the physical specimen from which the digital specimen was derived. It complements the other identification labels providing a minimum of provenance information. Sometimes, institution name and seal are part of other elements such as the barcode or the scale bar. Synthesys3 recommends the digitisation of institution name as an important element for documentation and linking of the physical and digital specimens [54].

Other elements include additional labels, cards, paper registries and field books, can be captured in a separate image. These can include labels on the base of taxidermy mounts; labels placed on the back of herbarium sheets and microscope slides; booklets attached to herbarium sheets; or liquid container registries which detail the specimens within each container.



2.1.4 Image Quality Criteria

The quality of a digital image grows in proportion to the spatial, spectral, radiometric, and time resolutions (greater resolution equals greater quality). Spatial Resolution describes how much detail in a photographic image is visible to the human eye. The ability to "resolve," or separate, small details is one way of describing what we call spatial resolution [24, 94, 111]. Spectral resolution refers to the bandwidth of the light frequencies captured by the camera sensor, in RGB images the information stored corresponds to different levels in three bands (red, green, and blue) [24, 94, 112]. The radiometric resolution specifies how well the differences in brightness in an image can be perceived; this is measured through the number of the grey value levels. Radiometric resolution determines how finely a system can represent or distinguish differences of intensity (brightness). Radiometric resolution is usually expressed as the number of levels or the number of bits, for example 8 bits or 256 levels that is typical of computer image files. Radiometric resolution is commonly measured through the number of the grey value levels [24, 94, 110]. Time (or temporal) resolution is given by the interval between time samples at which images are captured. Time resolution is important in dynamic image analysis, where time sequences of images are processed [24, 94, 113]. Video, composition of 360° views and 3D images are examples of image time sequences.

An image might be degraded during capture, transmission, or processing, and the evaluation of resolution measures can be taken to assess the degree of degradation, this is the quality of the image. The quality required can vary according to the purpose for which an image is intended [94].

The set of methods for assessing image quality can be divided into two categories: subjective and objective. Subjective methods are those where the ultimate criterion is the perception of a selected group of professional and lay viewers. They appraise an image according to a list of criteria and give appropriate marks [94].

Quantitative methods measuring image quality provide good objective tests and are easy to apply and automate. The quality of the image $f(x, y)$ is usually estimated by comparison to a known reference image $g(x, y)$. A synthesized image is often used for this purpose.

The different QA and QC methods for the digitisation workflows analysed mention both objective and subjective methods [31, 54, 61, 79]. The sections on quality management for the different digitisation workflows will describe the objective and subjective methods in greater detail, including examples.

2.2 Digitisation Workflow

Identifying the digitisation workflows and the artefacts produced by those workflows is required to determine when quality management activities need to be implemented and the data assets targeted by those activities. From a wide perspective, research infrastructures are designed and implemented to support five types of activities: (1) data acquisition, (2) data curation, (3) data publishing, (4) data processing, and (5) data use [29]. The ENVRI and ENVRIplus projects structured these activities into a generic data lifecycle (Figure 1).



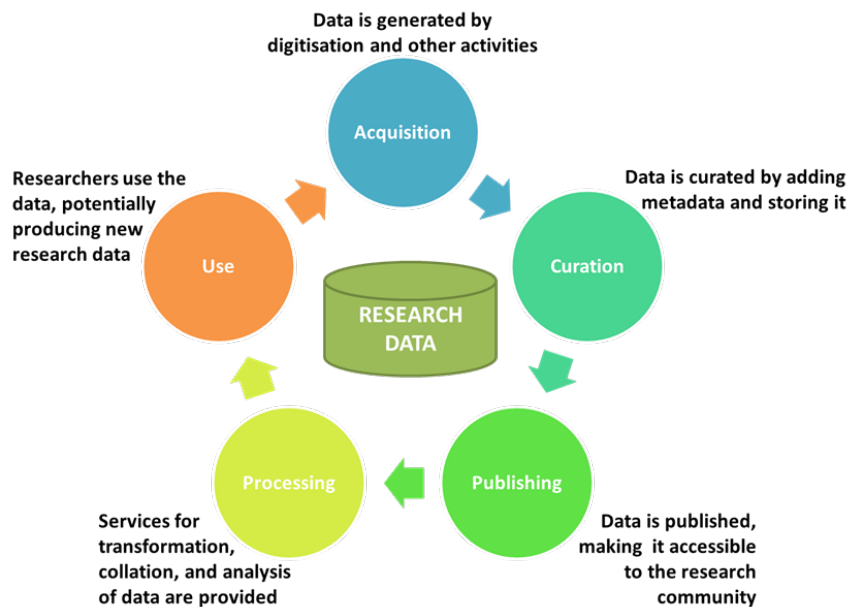


Figure 1 The Data Life Cycle

At a high level, existing and planned Research Infrastructures designed for supporting the digitisation of Natural History Collections exhibit behaviour that aligns with its phases. The requirement analysis for the DiSSCo data infrastructure (T6.2) within ICEDIG, has identified nine generic activities linked to the digitisation of Natural History Collections: (1) Pre-digitisation curation, (2) Imaging Station(s) Setup, (3) Imaging, (4) [specimen] conservation, (5) Image Processing, (6) Image Archiving, (7) Optical Character Recognition, (8) Manual Data Entry and Correction, and (9) Data Transcription. These activities can be aligned with the data lifecycle, as depicted in the Figure 2. Eventually, these activities will be extended to include those related to publishing and use of digital collections.

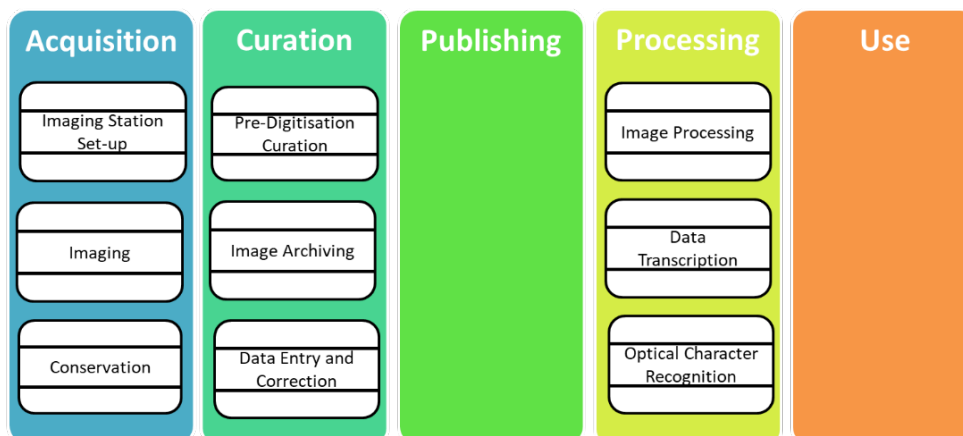


Figure 2 Alignment of Natural History Collections to the Data Lifecycle phases

These activities are derived from the definitions from digitisation workflow activities for different types of collections [50], which will be included in the definition of the Data Management Plan for the DiSSCo infrastructure [49]. The way in which these activities are presented is left undefined to allow greater flexibility for the description of digitisation workflows (in line with their definition by iDigBio [50]). For instance, Figure 3, Figure 4, and Figure 5 show three examples of digitisation workflows (depicted using UML activity diagrams). In these diagrams, the order of activities is switched or some of the activities are excluded.

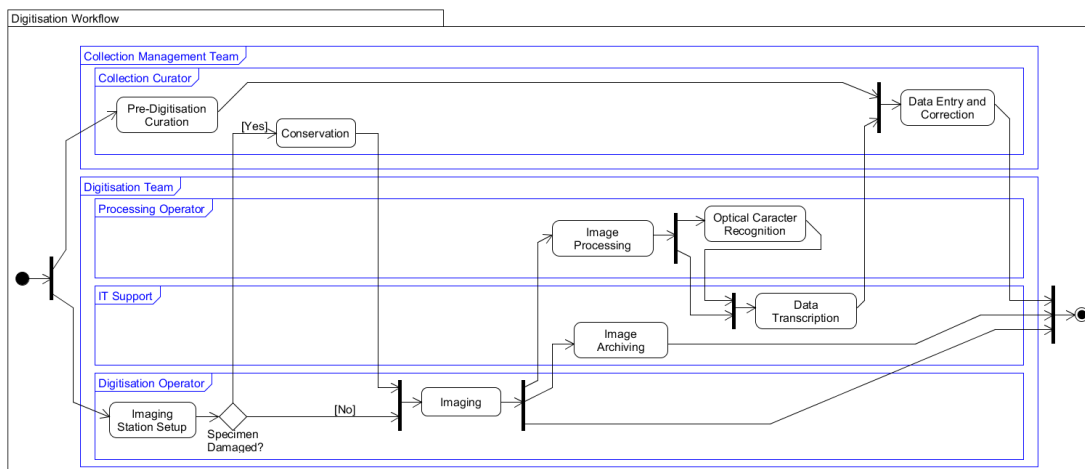


Figure 3 Digitisation Workflow Example 1. All activities in Figure 2 are included, and roles performing them are organised in two teams: Collection Management and Digitisation depicted as individual. Within each group, four roles are defined: Collection Curator, Processing Operator, IT Support, and Digitisation Operator. The activities within each role are executed sequentially, the arrows represent the processing sequence, from left to right. Bars represent forks and merges of processing. For instance, the first bar on the left indicates a fork, signalling that “Pre-Digitisation Curation” and “Imaging Station Setup” can occur in parallel. The last bar at the right shows that the digitisation process ends when all activities are completed (i.e. all processing threads finish).

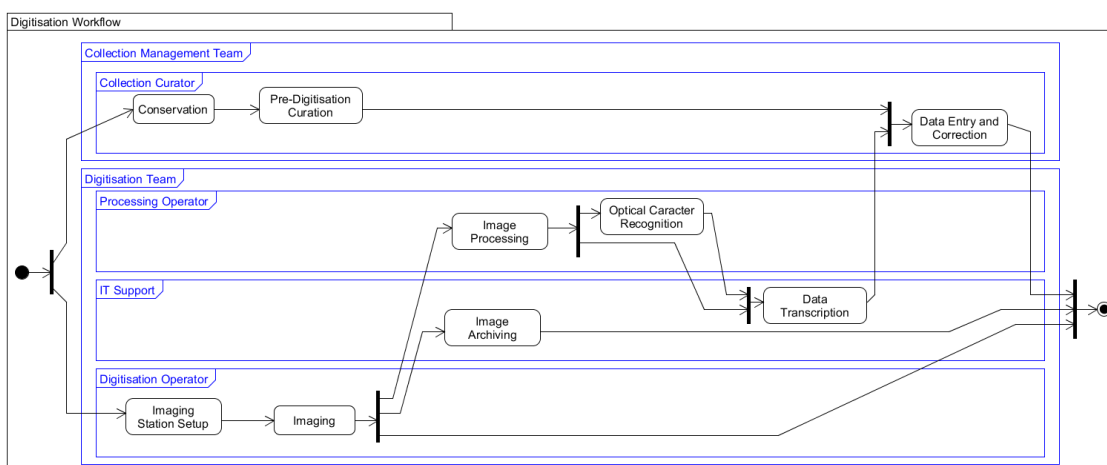


Figure 4 Digitisation Workflow Example 2. The activities and roles in the diagram are the same as those in Figure 3. The difference is that “Conservation” is performed before “Imaging”.

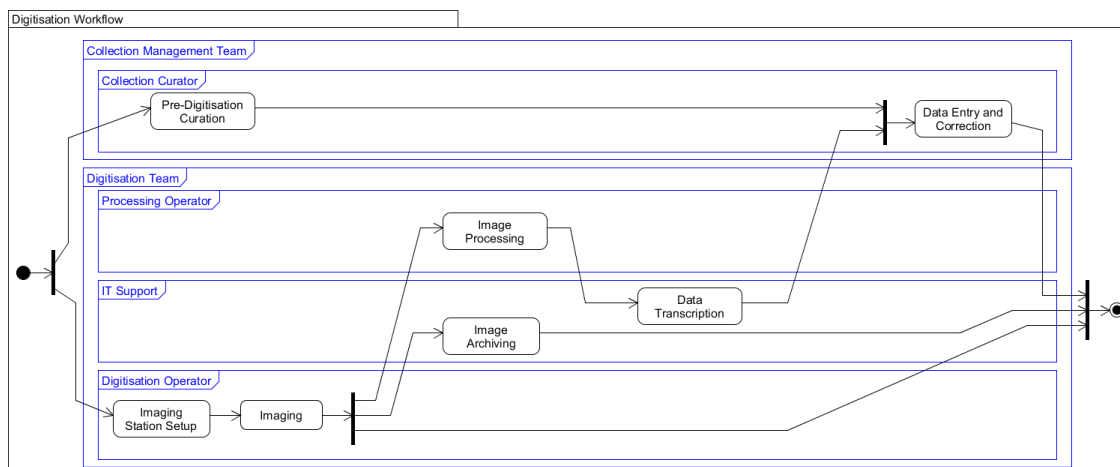


Figure 5 Digitisation Workflow Example 3. This workflow excludes two of the activities included in Figure 3 and Figure 4, Conservation and Optical Character Recognition. The roles assigned to perform them are not changed.

Variability on the implementation of digitisation workflows makes it hard to recommend an end-to-end quality management process. As an alternative, this report will define quality management methods which can be integrated into the workflows and used to verify the resulting data products. In these methods, the quality of a digital specimen, will be evaluated in terms of its intended use. For instance, a derived image intended for online display will have specific quality characteristics which are different from an image intended for OCR, Segmentation or further analysis. In this way, the definition of quality management methods is flexible, and they can be implemented as needed. Defining methods based on products leans towards a purely quality control approach. However, if the methods are combined and integrated into digitisation workflow, for verifying the quality of digital specimens and their derivatives as they are produced, they become also quality assurance methods. The target should be to increase quality assurance in the workflows to help drive down digitisation costs.

2.3 Quality Management in Digitisation

Quality management methods can be subdivided in two main areas: Quality Assurance, and Quality Control [21, 97]. Quality Assurance (QA) covers process-based activities which are designed to ensure that products or services are produced or delivered according to a predefined set of quality criteria. Quality Control (QC) covers product-oriented activities which are designed to guarantee that products or services conform to a predefined set of quality criteria.

In the ICEDIG context, QA encompasses activities which can be embedded in the digitisation workflows. Such activities are devised to detect and prevent error which can surge for the need to re-calibrate instruments, adjust lighting, exposure or scanning speed. Detecting these errors on time can minimise need for rework and excessive handling of specimens. QC encompasses activities which are performed to confirm that the digitisation products are fit for purpose (publishing, further analysis, processing, and derivation). In an ideal scenario, QC methods would be unnecessary, however, the inclusion of third parties requires some overseeing of the work they produce. The examples of QC activities observed are targeted at assets which are final products of the digitisation workflows, as such QC activities connect the creation, processing, and curation of digital specimens to their eventual publishing and use.

2.4 Tools

The tools described here can be used at different stages to evaluate the quality attributes of different products which are part of, or derived from, digital specimens. The tools are divided in hardware and software. However, digitisation systems can be integrated hardware/software systems.

2.4.1 Hardware

- Specimen Handling: Manual loading of specimens (manual positioning of specimen), conveyor-based stations (automated/assisted positioning of specimens).
- Digitisation equipment:
 - Camera based digitisation: Select a modern camera with full sized sensor to achieve the best results.
 - Scanner based digitisation: Derived from digitisation practices in libraries and cultural heritage collections, there are examples of scanner-based systems for natural history collections.



- Other digitisation equipment: Additional examples of equipment which provide additional information, X-Ray, micro CT, Laser, IR scanners, electronic microscopes.
- Lighting equipment: Flashes, fixed lights, LED, fluorescent, or other equipment used for achieving better digitisation results.
- Camera rig: devices used to help in achieving steady images, including mechanical fixed supports (tripods and bases), as well as automated or semi-automated camera positioning assets.
- Equipment for Digital Specimen Management
 - Calibrated monitors: Recommended for visual verification of digital specimens by trained experts. These should be calibrated using special equipment, but can be alternatively calibrated using guidelines, for instance when doing crowdsourcing of QC.
 - Processing Platforms: Physical and virtual resources required for handling of complex image processing: photo stacking, stitching, postprocessing, derivatives production.
 - Storage Platforms: Physical and virtual resources required for preserving digital specimens: quality management, archival, temporary storage of processing results, provenance systems, collection management systems.
 - Publishing Platforms: Physical and virtual resources required for publishing and providing access to digital specimens: portals, blogs, catalogues, crowdsourcing platforms.

2.4.2 Software

The nature of Quality Inspection of images during quality management activities relies in both manual and automated tasks. It is estimated that increasing the amount of automation in Quality Inspection can lead to significant improvements in the overall throughput of digitisation workflows. There are examples of software which has been developed explicitly for quality inspections, such as OpenDICE [32], that implements the image quality measurements recommended by FADGI [31]. OpenDICE enables automatic verification of sampling frequency, tone response, white balance, illuminance, and colour accuracy.

Further research in automation of image quality management is still needed, but there are examples that highlight the potential of building and improving software packages for automating image quality management. A semantic segmentation process developed by Natural History Museum London for microscopy slides was adapted for testing its use in herbarium sheet processing [74]. The process uses a pretrained neural network and transfer learning. The process can also be used to identify the existence of elements such as: Label(s), Barcode, Scale bar, colour chart, and specimen. The outputs of the segmentation process are then processed to evaluate colourfulness, contrast and sharpness of the image. This type of identification does not need processed images and can be done on derived JPG images. Further uses could be passing identified labels to OCR processes, barcode readers, and additional colour analysis software for quality assurance.



3. Quality Management for Digitisation of Microscope Slides

There are several projects involved in the digitisation of microscope slides, however common standards such as those developed for other collections have not been proposed (for instance the standards for herbarium sheets [54, 55, 79], see section 7.1). Consequently, the guidelines presented here are based on existing literature from digitisation projects of natural history microscope slide collections.

3.1 Microscope Slides Digitisation Standards

The requirements for digitising microscope slides should align with the intended use of the digitised slides. Given the size of the collections, one of the purposes of digitisation is to support the cataloguing of the collections. In this case, full-slide images are taken and then analysed for transcription [67, 69]. A second purpose of slide digitisation is to support scientific research with fast retrieval and research quality visualisation of digital specimens [64].

In the first case of transcription purposes, Naturalis Biodiversity Centre (Naturalis) and Natural History Museum London (NHM), have carried out pilots or are in the process of scanning in mass their microscope slide collections [43, 62, 96]. The outputs from these efforts are already published and available for inspection online [67, 69]. Figure 6 shows an example of a microscopic slide from their collection. The image is provided at a resolution which is effectively close to 1200 PPI. This resolution applies for the “raw” (initial scan original) image. Images at this resolution are not usually published but serve as the base from which other image products are derived. The resolution of these derived images will vary according to the intended use of the derived images (such as web publishing, or printing). For instance, Naturalis makes low-resolution images available for quick inspection and download, and a high resolution (~930 PPI JPEG) images, for download by request.



Figure 6 Microscope Slide from Naturalis²

NOTE: The slide in Figure 6 has an actual size of 2,836 X 1,015 pixels. Considering that the standard size of microscope slides is 25 X 75 mm, the resolution is approximately 930 PPI (936 PPI calculated directly considering the actual slide width to be 2,764 pixels).

In the second case of supporting scientific research, part of the collection is imaged twice, once for the entire slide (as in the previous example), and once at high resolution using a microscope-based scanner. This is the current process at the NHM [67]. In this case, a high throughput digitisation workflow is used for collection purposes, while a microscope-based scanner is used for digitising slides at research quality resolution. Figure 7 shows an example of a specimen from a microscope slide from the NHM collection which has been scanned at two resolutions. The image to the left (a) is provided at a resolution close to 2,700 PPI. The image on the right (b) has been scanned at a higher

² Naturalis BioPortal [69] <http://bioportal.naturalis.nl/multimedia/RMNH.INS.869466/term=artro&from=5>



resolution (1mm = 1,135PPI) which is about 28,800 PPI. The right part of Figure 7 presents the specimen at ten times its natural size (1.7 mm). At this resolution, the digital specimen may be considered appropriate for some research purposes. Both images are published together as part of a single record for the specimen, and the higher resolution images are available for download by request explicit request.



Figure 7 Microscope Slide from NHM: a) slide view at normal size (25mm X 75 mm) and b) magnified image of slide specimen (~1.7 mm length shown at 10X magnification)³

NOTE: The slide in Figure 7a has an actual size of 8,688 X 3,592 pixels. Considering that the standard size of microscope slides is 25 X 75 mm, the resolution is approximately 2,700 PPI (2,738 calculated directly considering the actual slide width to be 8,086 pixels). The specimen in Figure 7b has an actual size of 3046 X 4440 pixels. The scale bar included with the image indicates that a length of 200 μ m corresponds to 227 points, or 1,135 points per millimetre. Consequently, the resolution of the image is approximately 28,800 PPI (28,829).

The microscopic slides group from iDigBio tested and documented the use of different laboratory slide scanners for digital collections [64]. Some of the products they tested were promising in supporting mass digitisation producing research quality images of specimens. Figure 8 shows an example of a microscope slide and the 40X view of the specimen contained within. As the example shows, this scanning method has the potential to provide research grade images with high throughput as laboratory scanners can handle loads of up to 320 slides (Hamamatsu Slide Scanner).



Figure 8 iDigBio Microscopic Slides group: sample image using 40x objective and the Motic VM600 scanner with HeliconSoft EDF software [64]

³ NHM Data Portal [67]. Permanent URL: <http://data.nhm.ac.uk/object/5b804af3-5e82-44f6-9861-ed13b4b13f26>. Retrieved: 15:05 18 Sep 2018 (GMT)

3.1.1 Image Resolution and Colour Standards

Table 3 illustrates the standard criteria for digital specimens of microscope slides. The details about the different resolutions and uses are derived from the recommendation from the US Library of Congress [61], and FADGI [31]. These standards can be used in a QA processes to guarantee that products will meet these standards. Similarly, when used for evaluating images produced, then these evaluations form part of QC processes.

The recommendations for bit depth and colour accuracy are derived from the technical recommendation from the US Library of Congress [61]. Finally, the recommended colour space is Adobe RGB (1998), taken from FADGI [31, 52], combined with the resolutions obtained from the review of existing specimens published by NHM [67], Naturalis [69], and the review from iDigBio [64].

Table 3 Example of Image Standards for Microscope Slides

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Research and identification of specimen features	900 PPI*	24-bit colour		$\Delta E < 5$
Research and preservation of type specimens	28500 PPI**	24-bit colour		$\Delta E < 5$

* Minimum resolution (approximated from specimen obtained from Naturalis BioPortal, Figure 6) for full-slide digitisation. If digitisation devices available allow for higher resolution, that resolution should be used.

** Resolution recommended for effective magnification at 5X (approximated from specimen obtained from NHM data portal, Figure 7b).

3.1.2 Full-Slide Scans Image Elements

Image elements refer to visual elements which will appear next to or within the microscope slide specimen and which are intended to help in its identification, processing, and quality control. For full-slide scans, a list of four microscope slide elements was derived from the elements identified for semantic segmentation: (1) Barcode, (2) Labels, (3) Type Labels, and (4) Specimen [23]. The list has been verified by looking at specimen examples from Naturalis and NHM. All these elements are contained within the slide itself. Sometimes labels may be placed on both sides of the slide and may require more than one pass. Figure 53 shows the elements of the microscope slide specimens from NHM⁴ and Naturalis⁵.

⁴ NHM Data Portal [67]. Permanent URL: <http://data.nhm.ac.uk/object/2de299e9-1b6a-4053-ad4b-5e6671ab6f62>, Retrieved: 10:14 19 Sep 2018 (GMT)

⁵ Naturalis Bioportal [69]. URL: <http://data.biodiversitydata.nl/naturalis/specimen/RMNH.INS.867638>, Retrieved: 12:38 18 Sep 2018 (GMT)



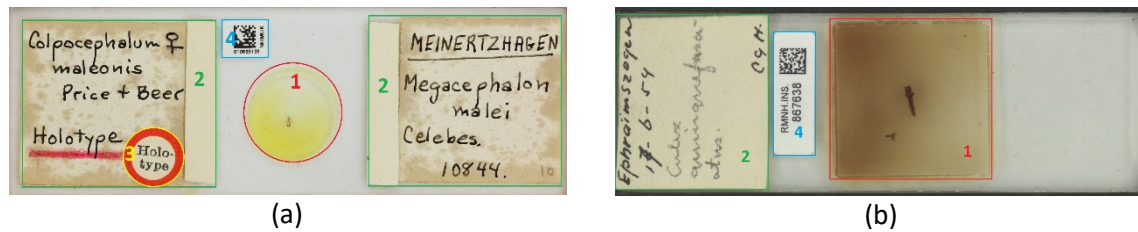


Figure 9 Examples of microscope slide image elements highlighted in two specimens from: (a) NHM⁴ and (b) Naturalis⁵. The numbered elements are (1) Specimen, (2) Labels, (3) Type Label, and (4) Barcodes. Notice that the type label is not always present, even when relevant.

The **specimen** (labelled as 1 in Figure 9) needs to be included and visible. However, the nature of the specimen might require a microscope-based scanner to provide actual research quality images of specimens, full-sized scans of slides can at most provide a view of the region where the specimen is located. Nevertheless, in a full-slide view the identification of the specimen can serve as a pre-processing step prior to high resolution scanning to detect areas of interest, decreasing high resolution digitisation time [64, 86].

Labels and Type Labels are commonly placed next to the specimen attached to the slide. Clear capture of labels is important for further processing and documentation of the specimens [79]. Type Labels provide additional information for identifying type specimens within a collection. The identification of type specimens can also serve as a pre-processing step prior to high resolution scanning to detect specimens of interest.

The **Barcode with Institution Identifier**, identifiers are used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens as well as being machine readable [79].

3.1.3 High-Resolution Slide Scans Image Elements

The elements of a high-resolution microscope slide image are the specimen and scale bar. Figure shows three examples from NHM⁶ and the Peabody Museum⁷.

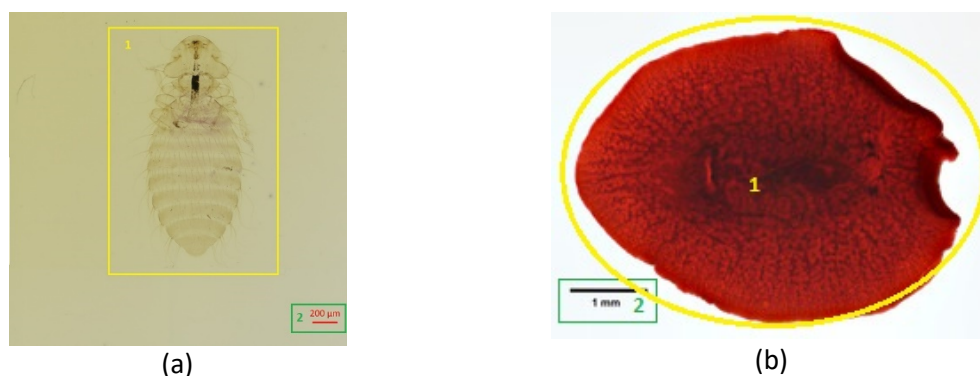


Figure 10 Examples of high-definition scans of microscope slide specimens highlighted in two examples from the collections of the (a) NHM⁶ and (b) Yale University Peabody Museum⁷. Numbers assigned to image elements correspond to (1) Specimen, (2) Scale Bar.

⁶ NHM Data Portal [67]. Permanent URL: <http://data.nhm.ac.uk/object/5b804af3-5e82-44f6-9861-ed13b4b13f26>. Retrieved: 15:05 18 Sep 2018 (GMT)

⁷ InvertEBase Portal [51]. URL: <http://www.invertebase.org/portal/collections/individual/index.php?occid=802125>, Retrieved: 12:38 19 Sep 2018 (GMT)

Specimen, in high-resolution microscope scanning, is digitised as a series of focus bracketed images which are then stitched and overlaid to provide a focused and detailed view of the specimen. Due to the time consumed in high-resolution scanning and the cost of the equipment, the rapid identification of the region of interest plays an important factor in high-resolution scanning.

Scale Bar is recommended to enable the calculation of the physical dimensions of the specimen [79]. In Figure 10a, for instance the scale indicates 200 μ m (0.2 mm), which leads to an estimate of height of 1.7 mm and 1.0 mm width for the displayed specimen[79].

3.1.4 Example

The image in Figure 11 shows a slide and a high-resolution view of the specimen within the slide from the NHM⁸.

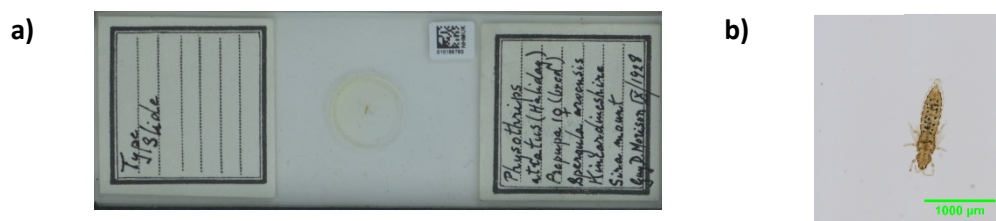


Figure 11 Microscope Slide from NHM: a) slide view at normal size (25mm X 75 mm) and b) magnified image of specimen from slide (~1.4 mm length shown at 10X magnification, the reported magnification of the image is 5X)⁸

Figure 12 illustrates the possibilities of microscope-based scanners. The series shows a 0.5 mm segment of the high definition image from Figure 11b at seven levels of magnification (1X, 2X, 4X, 5X, 10X, 20X and 40 X). The production of digital specimen images with these levels of details using camera only digitisation methods is not possible. However, the camera-based full-slide digitisation and the microscope-based scanning of the specimen complement each other and their use in the mass digitisation of slide collections is encouraged. This is discussed further in the next section on the digitisation workflows for microscope slides.



Figure 12 Magnification series of the high definition specimen image presented in Figure 11b. 5X (highlighted in red) corresponds to the magnification used by NHM [62,96], fourth image from left to right.

⁸ NHM Data Portal [67]. Permanent URL: <http://data.nhm.ac.uk/object/ac2aea0c-de67-4958-829e-3be189d43dc8> Retrieved: 13:58 19 Sep 2018 (GMT)

3.2 Microscope Slides Digitisation Workflows

According to Synthesys3, the digitisation workflows of microscope slide collections of different institutions are not as well differentiated as those of other types of botanical, zoological, or entomological collections [79]. However, examples from Naturalis [43] and the NHM [62, 96] illustrate some of the common tasks which are likely to be found in microscope slide digitisation workflows. Figure 13 shows a high-level view of a generic microscope slide digitisation workflow, derived from the digitisation activities recommendations found in iDigBio, Synthesys3, Naturalis and NHM [50, 43, 62, 71, 96]. Most of the tasks in the diagram coincide with those defined by iDigBio [50,

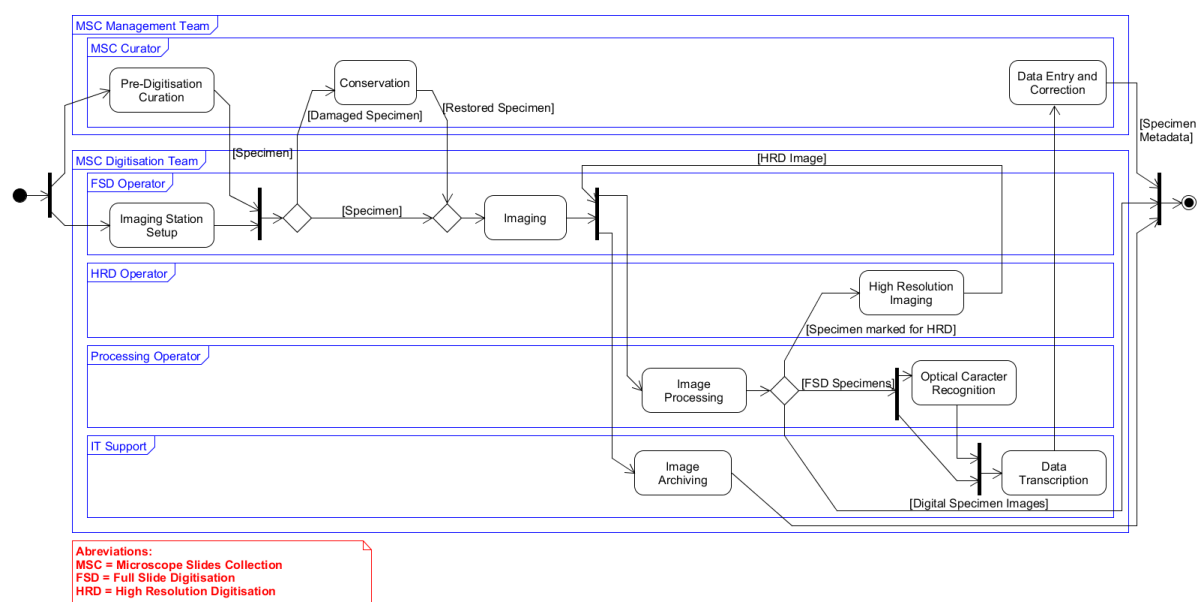


Figure 13 Generic microscope slides collection digitisation workflow

71], while the sequencing and alternative paths are derived from the Naturalis and NHM examples [43, 62, 96]. The diagram includes the activities which can occur in a digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, some digitisation projects may perform only the full slide scan part of the workflow for documenting the collection, leaving the high-resolution scanning of selected specimens for a later date. Similarly, some collections may have already captured data and therefore can omit the “data transcription” and the “data entry and correction” tasks. Table 4 provides a brief description of the workflow tasks and their influence in the production of images.

Table 4 Description of coarse-grained microscope slides workflow tasks, derived from [50, 43, 62, 71, 96]

Workflow Task	Description/Subtasks	Influence on Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators.
Imaging Station Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians. If the workflow includes high resolution digitisation, the setup of the corresponding scanner should also be performed at this time	Equipment should be calibrated to minimise image postprocessing after digitisation.
Conservation	Required conservation/restoration of specimens selected for digitisation.	Some slides may be damaged or fragile. Some may require using additional blank slides as covers or special slide holders.

Table 4 Description of coarse-grained microscope slides workflow tasks, derived from [50, 43, 62, 71, 96]

Workflow Task	Description/Subtasks	Influence on Image Quality
Imaging	Data capture; identification (barcoding); mounting for imaging; digitisation of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta]data capture, so that images are correctly linked to the corresponding specimen records.
High Resolution Imaging	Identification of specimens for high definition scanning, identification of region of interest within slide, high resolution imaging, creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta]data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of derivatives for publishing and distribution; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
Image Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data Transcription	Manual extraction of data from images, for populating/complementing specimen record.	Verify readability of derived image data transcription.
Data entry and correction	Final capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing.

The main variations on the integration of digitisation workflows depend on the metadata required and the type of digitisation stations to be used. The following sections describe how the activities for microscope slides digitisation are affected by these two features [71, 79].

3.2.1 Metadata Required for Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record on the collection management system. Additionally, specific data which is contained with the specimen such as scientific name, field notes and collection name should also be included. In microscope slides digitisation this is normally contained as labels which are placed in the slide next to the specimen. This is relevant for imaging because the image should capture the relevant metadata clearly for its later transcription and verification. The extraction of data from the digitised specimens can entail automated or manual activities which can take place before or after digitisation. Alternatively, some institutions may already have most of the metadata stored in a collection management system, in which case the most important information is the identification data, which is one of the functions of barcodes. There is no convention for naming slide images, however, barcode IDs could be used to name the image specimen files (analogous to the convention used for herbarium sheets).

3.2.2 Digitisation equipment

Digitisation stations can be classified as camera or microscope based. Additionally, the specimen positioning and imaging can be manual or semiautomated. There are no examples of end-to-end automated digitisation stations [71, 79, 98]. Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. Semiautomated stations have various forms which can vary in the amount of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen under the digitisation equipment. The speeds



of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. Highly automated microscope slide scanning systems, e.g. designed for medical purposes, may not be suitable for a large part of NH slide collections due to deviating size and fragility. Naturalis reports a between 2,000 and 2,500 slides per day, using a SatScan semiautomated approach [43], while NHM reports a throughput of up to 5,000 slides per day [44]. The speeds of the high-resolution microscope-based scanners vary in relation to the magnification and the identification of the area of interest to be scanned. An early study compared 31 laboratory scanners and reports that the fastest ones can achieve an average speed of 3.72 minutes at 20X for a 10mm X 10mm area, and up to 25.75 minutes for the same area at 40X. However, these figures are from 2006, and speeds may have improved [37]. For instance, iDigBio has experimented with these types of scanners for microscope slides containing insects [64]. Combining the fast throughput full-slide digitisation workflow with image segmentation and analysis (for rapid identification of regions of interest) tools can support the development of a more efficient methods for high-definition digitisation workflows.

3.3 Microscope Slides Quality Management Methods

In line with the imaging standards for Microscope Slides (3.1), the quality management methods for microscope slides focus on the quality of the digitised specimen and the presence of the recommended image elements, especially labels. The quality management activities should verify the presence and clarity of image elements (sharpness, spatial resolution, and acutance), verification of colours (spectral resolution), and verification of the perception of brightness in the image through the number of the grey value levels (radiometric resolution). Additionally, some elements with known standard dimensions could be used to verify the spatial resolution of the image, for instance barcodes.

This quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are verified should be those that follow the initial digitisation of the specimen and after the production of each image derivative.

3.4 Examples

There are two main variants of implementation of microscope slide digitisation workflows: (1) full-slide and (2) full-slide/selective high-definition digitisation. In both cases, the institution hosting the collection procures and operates the digitisation equipment, designs the digitisation process, performs the digitisation, and manages the digitised collection. There are no current examples reported of outsourced digitisation for microscope slide collections, however schemas in which costly laboratory slide scanners are rented would be a viable alternative. The following subsections the digitisation variants highlight the corresponding quality management activities and their place in the workflows.

3.4.1 Full-Slide Digitisation

The microscope slide digitisation workflow of Naturalis is an example of a full-slide only digitisation workflow. The main reasons for designing and developing this workflow inhouse was the need to rapidly digitise the entire slide collection of Naturalis, focussing on capture of label data [43]. This workflow was designed following the general template of the Digistreets project, with the only variation of omitting the prioritisation step, since the whole collection was the target. Naturalis designed a high throughput workflow which responded to these requirements. Figure 14 shows a



diagram of the glass street production line, which is the instance of the full-slide only workflow. This approach aims to digitise all specimens in the collection within a single project.



Figure 14 Schematic representation of the Glass Street production process. Illustration: Ben van Arkel (from [43]). The numbers on the image are as follows: 1) Location and transport of storage container to the digitisation station; 2) Labelling of slides with barcodes; 3) Fill scanning tray with batches of 100 slides; 4) Place scanning tray in SatScan and capture batch; 5) Return slides from batch to labelled container; 6) Return storage container to labelled storage location; 7) SatScan creates a high resolution image of the digitisation tray; 8) Automated Segmentation of individual slides; 9) Save individual slide image files named using barcode; 10) Manual transcription of label data.

Quality management methods for microscope slides focus on the quality of the digitised specimen and the presence of the recommended image elements (see sections 3.1 and 3.3). The points of the workflow when they are verified should be those that follow the initial digitisation of the specimen and after the production of each image derivative, followed by QC of the data entry. The suggested points for inserting image quality management activities in the workflow described above follow the stage in which slides are inspected before digitisation and those in which images are produced and processed, i.e. after tasks 2, 7, 8, and 9 (Figure 14). Table 5 presents the image quality management activities and their correspondence to different image products. The first column of the table indicates the workflow task that produces the image (or image set) to be verified, the second column indicates the image (or image set) to be verified, and the third column describes the verifications to perform.

For QA, quality verifications should be performed periodically (once per every 10 elements) to minimise rework.

Table 5 Quality Management Activities to incorporate into the full-slide digitisation workflow

Workflow Task	Specimen or Image (set)	Verify
2) Apply Barcode to slide	Individual slides	Presence of elements: (specimen, labels, barcode), these should be verified manually by the curator or trained digitisation technician. Making sure that labels are not occluded and that they are readable.
7) Digitise Slide Tray	High Resolution image of microscope slide batch in digitisation tray (master “raw” image for archiving)	Spatial Resolution: can be verified using the lines on the slide holding tray and the barcodes as reference.
8) Segment full tray image	Set of 100 individual microscope slide images	Spatial Resolution: can be verified using the barcodes as reference. Presence of elements: (specimen, labels, barcode) these can be verified using semantic segmentation methods.
9) Store and rename individual images	Individual microscope slide	Spatial Resolution: can be verified using the barcodes as reference. Presence of elements: (specimen, labels, barcode), these can be verified using semantic segmentation methods. File Naming: verify against barcode

Note: The description of this workflow does not include colour charts or scale bars, consequently the verification of colours (spectral resolution) and brightness (radiometric resolution) are not included.

3.4.2 Full-Slide/Selective High-Definition Digitisation

The digitisation workflow of NHM is an example of a full-slide/selective high-definition slide digitisation workflow. NHM has tested two variants of these workflow. The first variant utilises a setup similar to the one from Naturalis using SatScan for full-slide scanning (see 3.4.1), and a microscope-based scanner (Axio Scan System) for selective high-definition scanning. The second variant tested uses a static SLR camera and light box mount for single full-slide scanning, and a static SLR camera with a 5X zoom lens and light box for high definition specimen scanning [62, 96]. Both variants are instances of the generic microscope slides digitisation workflow (Figure 13). Figure 15 shows a full-slide digitisation station designed and built by the NHM digitisation team.

The quality management verifications recommended for this type of workflow are the same as those for the full-slide digitisation workflow presented in Table 5. These tasks are complemented with an additional verification after the creation of the high definition specimen(s) image(s) (Table 6). The quality of slide images can be evaluated by looking at quantifiable parameters such as colour, brightness, contrast, and focus sharpness (spatial resolution) [119]. Several quality evaluation methods developed for clinical slides can be explored and adapted for microscope slides in natural history collections [7, 78, 89, 119]



Figure 15 NHM Digitisation Station for single full-slide digitisation workflow.

Table 6 Additional Quality Management Activities to incorporate into the full-slide/high-definition digitisation workflow

Workflow Task	Specimen or Image (set)	Verify
High Resolution Imaging (from Figure 13)	High Resolution Image(s) of Specimen(s)	Presence and clarity of specimen(s) (spatial resolution), verification of colours (spectral resolution), and verification of the perception of brightness in the image through the number of the grey value levels (radiometric resolution).



4. Quality Management for Digitisation of Skins and Vertebrate Materials

Skins and Vertebrate material specimens are part of various natural history collections such as Ichthyology, Herpetology, Mammalogy, and Ornithology. Accordingly, these specimens are preserved using different methods such as: study skins (hides), cased skins (preparations without cuts to the abdomen), skeletons, taxidermy mounts, egg and nest lots, and turtle shells [103]. All these specimens can be digitised using 2D and 3D imaging methods. This section focuses on the quality management for 2D imaging methods. For the corresponding methods for quality management for 3D imaging see section 8.

The variety of collections and preservation media means that it is not possible to provide a general standardised method such as those developed for collections such as herbarium sheets (see 7.1). Consequently, the guidelines presented here are based on existing literature and examples from digitisation projects of skins and vertebrate materials within natural history collections.

4.1 Skins and Vertebrate Materials Digitisation Standards

The requirements for digitising skins and other vertebrate materials should align with the intended use of the digitised specimens. Given the size of the collections, one of the purposes of digitisation is to support the cataloguing of the collections. In this case, full-specimen images are taken and then analysed for transcription [67, 69]. A second purpose of specimen digitisation is to support scientific research with research quality digital specimens, as proposed by the Smithsonian Institution and the European Network of Biodiversity Informatics [9, 42].

Naturalis Biodiversity Centre (Naturalis) have digitised large portions of their skins and vertebrate materials in their collections [43]. The outputs from this effort are already published and available for inspection online. Figure 16, Figure 17 and Figure 18 show three examples of skin and other vertebrate materials from Naturalis [69]. Figure 16 shows specimen images which could be used for research to some extent, Figure 17 and Figure 18 provide views of the physical specimens they represent, however, their use for scientific analysis may be limited, which is why they are considered to serve mostly an illustrative purpose. Observe also that this type of collection requires more than one view for displaying different specimen characteristics.



Figure 16 Palatal(a) and lateral(b) views of a common pipistrelle (bat) skull⁹

⁹ Naturalis BioPortal [67], URL: <http://data.biodiversitydata.nl/naturalis/specimen/ZMA.MAM.17421.a>

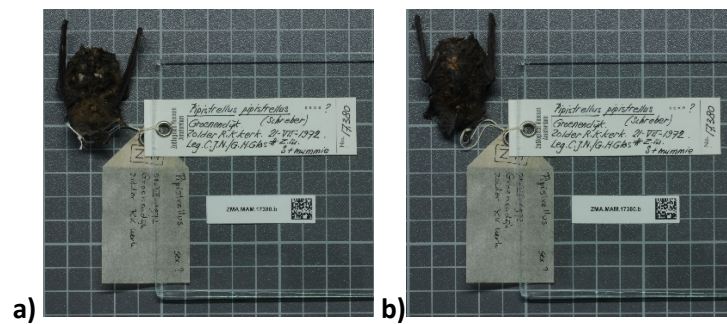


Figure 17 Ventral (a) and dorsal (b) views of a common pipistrelle (bat) skin¹⁰

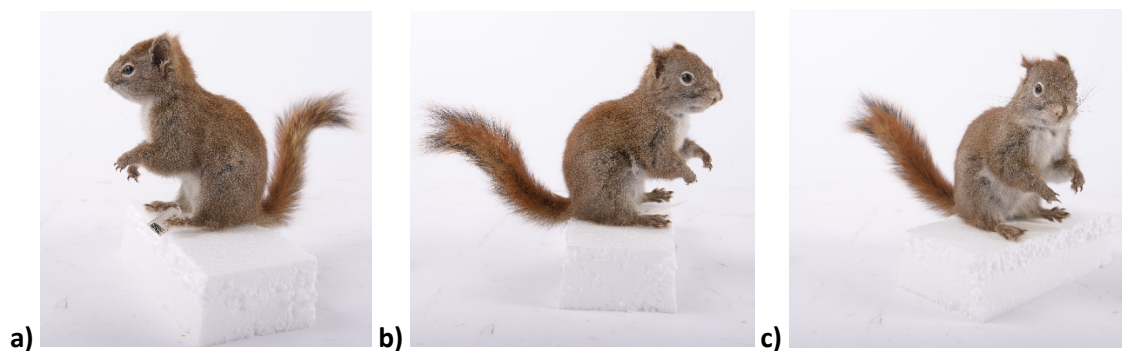


Figure 18 Left-side(a), right-side(b), and 3/4 right-side (c) views of American red squirrel skin (mounted skin/taxidermy mount)¹¹

For instance, the images in Figure 16 are digitised at a 1 to 1 ratio (10cm width by 8cm height), at 600 PPI and allow enlarging and viewing the images with up to 10X magnification with acceptable quality, i.e. in focus and not pixelated (Figure 19). The specimen images in Figure 17 and Figure 18 have a lower resolution (300 PPI), and the greatest magnification which can be achieved is 6X, lower than that of the previous example (see Figure 20).

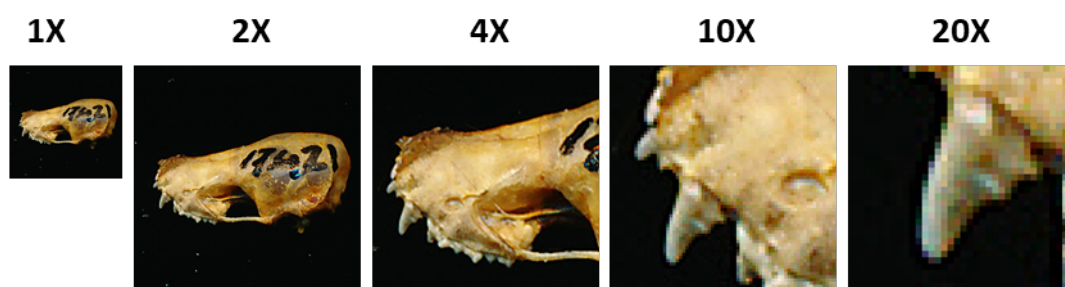


Figure 19 Magnification series of specimen image presented in Figure 16b. Trying to isolate the largest visible tooth. Magnifications beyond 10X are increasingly pixelated.

¹⁰ Naturalis BioPortal [67], URL : <http://data.biodiversitydata.nl/naturalis/specimen/ZMA.MAM.17380.b>

¹¹ Naturalis BioPortal [67], URL: <http://data.biodiversitydata.nl/naturalis/specimen/RMNH.MAM.60183>

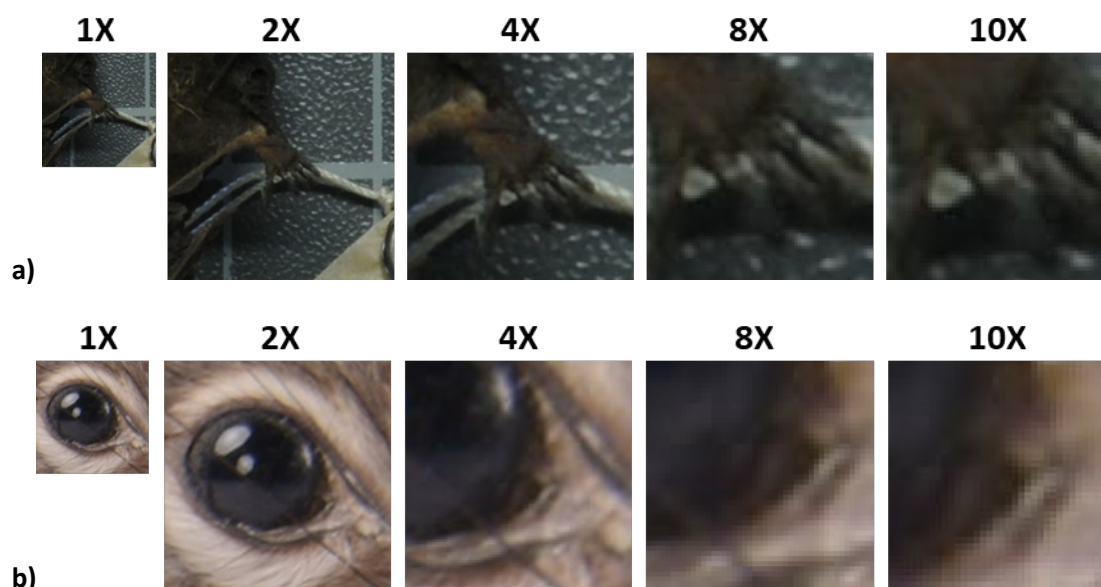


Figure 20 Magnification series of specimen image presented in Figure 17a and Figure 18c. Series (a) is aimed at isolating the fingers of the lower-left limb from Figure 17a. Series (b) is aimed at isolating the right eye from Figure 18c. The nominal resolution for these images is 300 PPI, so the magnification which can be achieved is limited. In both cases, magnification beyond 8X appear out of focus/pixelated.

4.1.1 Image Resolution and Colour Standards

Table 7 illustrates the criteria for digital specimens of skins and vertebrate materials. The details about the different resolutions and uses are derived from the recommendation from the US Library of Congress [61], and FADGI [31], and from calculations derived from existing published specimens. These standards can be used in a QA processes to guarantee that products will meet these standards. Alternatively, standards can be used for evaluating images produced as part of a QC process.

The recommendations for bit depth and colour accuracy are derived from the technical recommendation from the US Library of Congress [61]. Finally, the recommended colour space is Adobe RGB (1998), taken from FADGI [31, 52], combined with the resolutions obtained from the review of existing specimens published by NHM [67], Naturalis [69], and the review from iDigBio [64].

Table 7 Example of Image Standards for Skins and Vertebrate Materials

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Research and identification of specimen features	600 PPI ^{8*}	24-bit colour		$\Delta E < 5$

* Minimum resolution (from specimen obtained from Naturalis BioPortal, Figure 16). If digitisation devices available allow for higher resolution, that resolution should be used.

4.1.2 Image Elements

Image elements refer to visual elements which appear next to the specimen and which are intended to help in its identification, processing, and quality control. For research-quality specimens, a list of

six elements was derived from the recommendations for scanning type specimens: (1) Barcode, (2) Labels, (3) Colour Chart, (4) Scale, (5) Specimen and (6) Background [42]. The list has been verified by looking at specimen examples from Naturalis BioPortal [69]. The elements are not mandatory and can vary depending on the specimen being digitised. Figure 21 and Figure 22 show different sets of elements within specimen images (from the University of Navarra and Naturalis). For illustrative specimens, there is no set number of elements required, except that those included should be clearly visible.

The **Colour Chart** is recommended for helping with quality control and post-processing [106], this can help in verifying the lighting, white balance and colour accuracy of the image [10, 31, 54, 55, 61, 79]. Background must maximise the identification of the specimen borders, avoid glossy/reflective materials which can introduce artefacts.

The **Scale Bar** is recommended to enable the calculation of the dimensions of the specimen and to identify the collection holding the specimen [10, 79].

There are different types of colour charts, and many of them have been used by different institutions in their digitisation pipelines. However, modern targets such as those from Image Science Associates are preferred over legacy targets (Colour Control Patches from Kodak) because they were developed for digital image creation and are made to very tight tolerances [18, 53]. Object level targets of this type include a ruler and can be used for verification of colour, sharpness and scale.

Labels are commonly placed next to the specimen. Clear capture of labels is important for further processing and documentation of the specimens [10, 79].

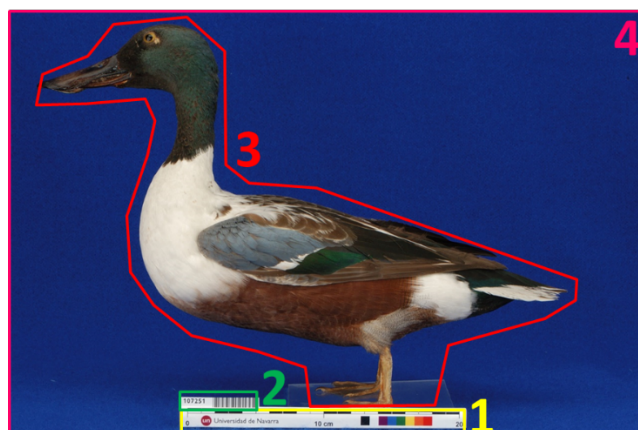


Figure 21 Specimen image showing some of the recommended elements for digitisation of Skins and Vertebrates. (1) Colour Chart and Scale Bar, (2) Barcode, (3) specimen and (4) background¹².

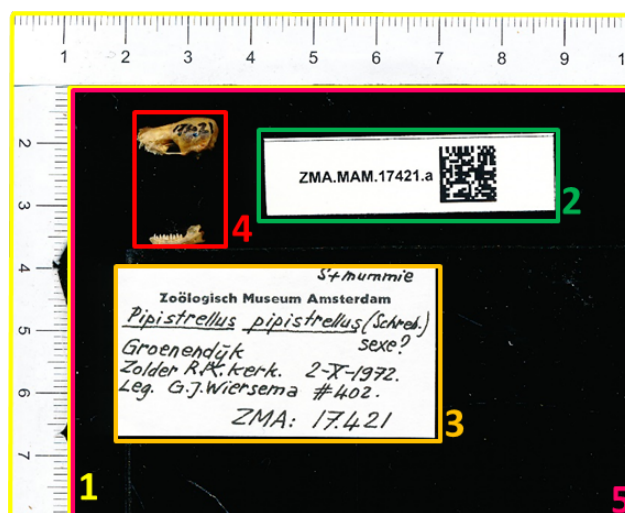


Figure 22 Specimen image showing some of the recommended elements for digitisation of Skins and Vertebrates. (1) Scale Bar, (2) Barcode, (3) label, (4) specimen, and (5) background¹³(no colour chart)

¹² University of Navarra [100], URL: <http://www.unav.es/unzyec/mzna/specs/107251.htm>

¹³ Naturalis BioPortal [67], URL: <http://data.biodiversitydata.nl/naturalis/specimen/ZMA.MAM.17421.a>

Barcodes are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens [10, 79]. There are different types of barcodes available. Line Barcodes (one dimensional), like the one on Figure 21, have been used in some digitisation projects, however, there were some misidentification problems (because of line features in the specimens). Consequently, the use of two-dimensional barcodes is recommended, and they are being adopted for new projects [118].

Specimen: Skin and vertebrate specimens are 3D objects, consequently producing sets of images in different poses is required [10, 79]. To achieve this, exposure of relevant anatomical traits needs to be considered. Figure 23 shows images of a specific trait used in a study of the evolution of a trait in related species. Figure 24 shows the different planes and axes for fish specimens. For specimens with multiple images it is not necessary for all images to include all the elements. However, background, scale, and colour chart should be present in all images showing the specimen, to enable colour postprocessing and estimation of scale. Other elements such as the specimen labels can be included in a separate photo [10, 70].

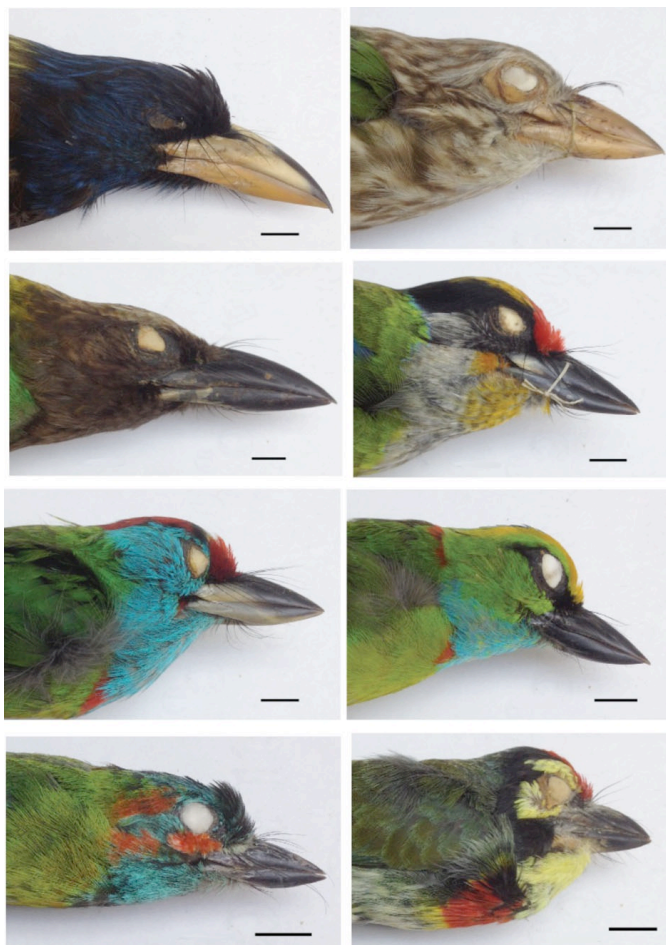


Figure 23 Diversity of size and beak shape in a representative sample of Asian barbets. This image illustrates the use of images of an anatomical trait for comparison of evolution of related species¹⁴.

Background: The background of digital images of specimens can be changed at will and selected to maximise the contrast with the subject. Light and colour can be used: a background can be much lighter or much darker than the relative brightness of the specimen, alternatively it could be of a colour

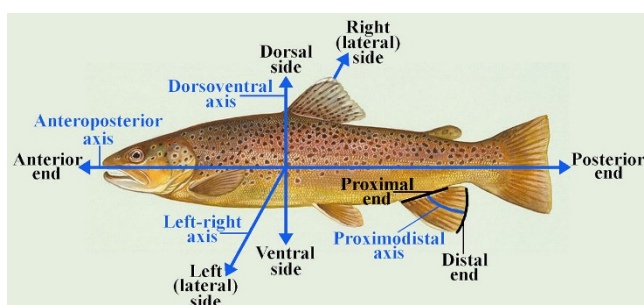


Figure 24 Anatomical directions and defined axes in a fish¹⁵.

¹⁴ Taken from [57]

¹⁵ Taken from [105]

that is not present on the specimen. A featureless background is often preferable, but a background with centimetre or millimetre grid can be suitable too. The main goal is to clearly separate features, especially in the contour of the specimen, from the image background [10].

4.2 Skins and Vertebrate Materials Digitisation Workflows

According to Synthesys3, the digitisation workflows of zoological specimens (which include skins and vertebrate materials) are slower than those of herbarium collections [79], and their speeds are affected by the number of shots per specimen, the varied type and size of specimens, the number of collections which can contain these materials (mammalogy, ornithology, herpetology, to name a few), and the level of documentation of the collection. Additionally, most of the digitisation systems for this type of collection are developed in-house, with custom-made workflows that match the resources available at each institution [50, 70, 77, 79].

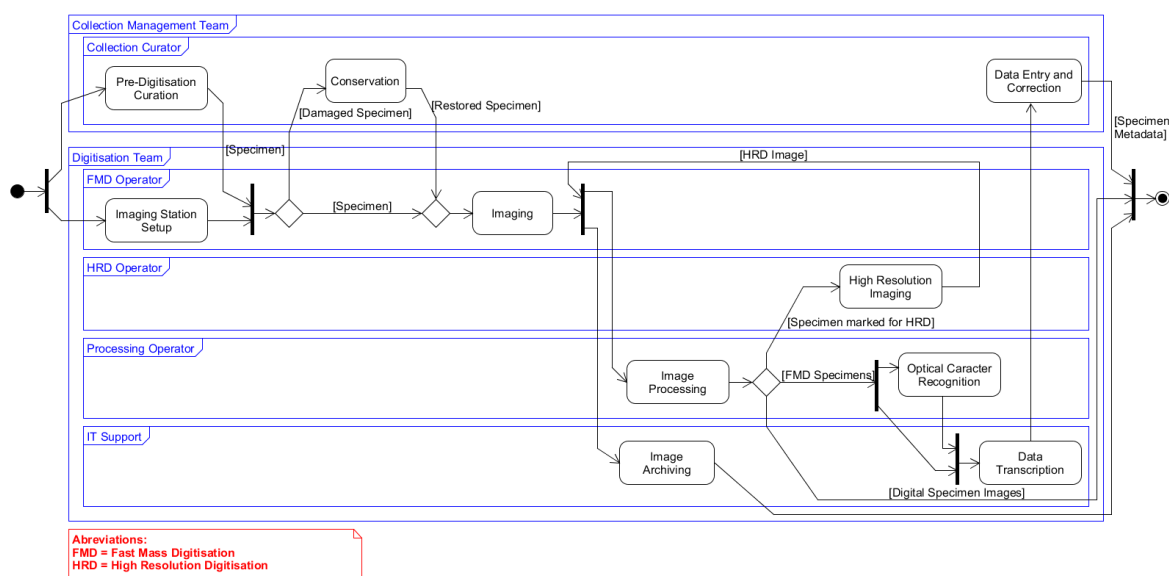


Figure 25 Generic Skins and Vertebrate Materials digitisation workflow

Figure 25 shows a high-level view of a generic skin and vertebrate material digitisation workflow, derived from the digitisation activities recommendations found in iDigBio, Synthesys3 and Naturalis, and the University of Navarra [10, 43, 50, 70, 79]. Most of the tasks in the diagram coincide with those defined by iDigBio [50, 70], while the sequencing and alternative paths are derived from the Naturalis and NHM examples [10,43, 79]. The diagram includes the activities which can occur in a digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, some digitisation projects may perform only the “fast mass digitisation” part of the workflow for documenting the collection, leaving the high-resolution digitisation of selected specimens for a later date. Similarly, some collections may have already captured data and therefore can omit the “data transcription” and the “data entry and correction” tasks [79]. The number of images per specimen, both in fast mass digitisation and in high definition digitisation, also influence the shape of the workflow instances. Table 4 provides a brief description of the workflow tasks and their influence in the production of images.

Table 8 Description of coarse-grained workflow tasks, derived from iDigBio recommendation [50].

Workflow Task	Description/Subtasks	Influence on Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); identification of specimens requiring high-definition digitisation; defining safeguards for identifying and handling specimens treated with hazardous materials (e.g. mercury); transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators.
Imaging Station Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians.	Equipment should be calibrated to minimise image post-processing.
Conservation	Required conservation/restoration of specimens selected for digitisation.	Some specimens may be damaged, fragile, or require to be re-mounted to display all relevant traits.
Imaging	Data capture; identification (barcoding); mounting for imaging (may require different mounts for different poses); imaging of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta] data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of derivatives for publishing and distribution; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
High Resolution Imaging	Identification of specimens requiring high-definition scanning, identification of poses for specimen, high resolution imaging, creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta]data capture, so that images are correctly linked to the corresponding specimen records.
Image Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data Transcription	Manual extraction of data from images, for populating/complementing specimen record.	Verify readability of derived image for data transcription.
Data entry and correction	Final capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing, including file naming.

The main variations on the characteristics of digitisation workflows depend on the metadata required (or already present) and the type of digitisation stations to be used. The following sections describe how the activities for skin and vertebrate materials digitisation are affected by these two features [70, 77, 79].

4.2.1 Metadata Required for Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record on the collection management system. Additionally, specific data which is contained with the specimen such as scientific name, field notes and collection name should also be included. In skin and vertebrate materials digitisation, this is normally contained as labels which are attached to the specimen or to its containing/storage medium. This is relevant for imaging because the image should capture the relevant metadata clearly for its later transcription and verification. As described in the section on image elements (4.1.2), this can be done in a separate image. The



extraction of data from these images can entail automated or manual activities which can take place before or after digitisation. Alternatively, some institutions may already have portions of the metadata stored in a collection management system, in which case the most important information is the identification data, which is achieved by barcoding. For these specimens there is no naming convention, however, many collections use the barcode to name the imaged specimen files (for instance [69]).

4.2.2 Digitisation equipment

Most of the digitisation stations for skin and vertebrate materials are custom-made and camera based. Specimen positioning and imaging is mostly a manual process, and there are no examples of fully automated digitisation stations [70, 77, 79, 98]. Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. There are some components, such as focus stacking rails, from various vendors, which can produce semiautomated stations with different levels of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen under the digitisation equipment. The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. Synthesys3 reported a huge variation in the range of 1,000 to 10,000 specimens per year, with costs in the range from 2.00 € to 16.00 € [79].

4.3 Skin and Vertebrate Materials Quality Management Methods

In line with the imaging standards for skin and vertebrate materials (4.1), the quality management methods focus on the quality of the digitised specimen and the presence of the recommended image elements. The quality management methods should verify the presence and clarity of image elements (spatial resolution, especially the labels), colours (spectral resolution), and the perception of brightness in the image through the number of the grey value levels (radiometric resolution). In this case, some elements can be used to support these verifications. For instance, lines on the scale and barcode elements can be used as reference to verify the spatial resolution of the image, while the colour chart can support the verification of spectral and radiometric resolutions.

Quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are realised should be those that follow the initial digitisation of the specimen and after the production of each derivative from it.

4.4 Examples

The examples of implementation of skin and vertebrate materials digitisation workflows are mainly in-house camera-based systems, in which the hardware and software is sourced from the best available off the shelf components within the institutional budgets. The institution hosting the collection procures and operates the digitisation equipment, designs the digitisation process, performs the digitisation, and manages the digitised collection. These setups can produce research quality digital specimens [10, 77, 79]. The digitisation workflows for these materials can be further divided into fast mass digitisation workflows [43] and research quality imaging (high-definition) workflows [10]. The following examples illustrate these two workflow types.

4.4.1 Fast Mass Digitisation Workflow

The approach taken by Naturalis, called Digistreets, entails the development of a series of workflows which are targeted at mass digitalisation of collections [43]. In this case the requirements of various



collection types were analysed to develop a family of digitisation workflows with a common approach, which allowed the sharing of expertise and resources among the digitisation teams and the collection curators. Figure 26 presents a high-level view of the Digistreets approach and Table 1 presents a description of the tasks and a mapping to the generic tasks of the digitisation workflow (as shown in Figure 25 and Table 8).

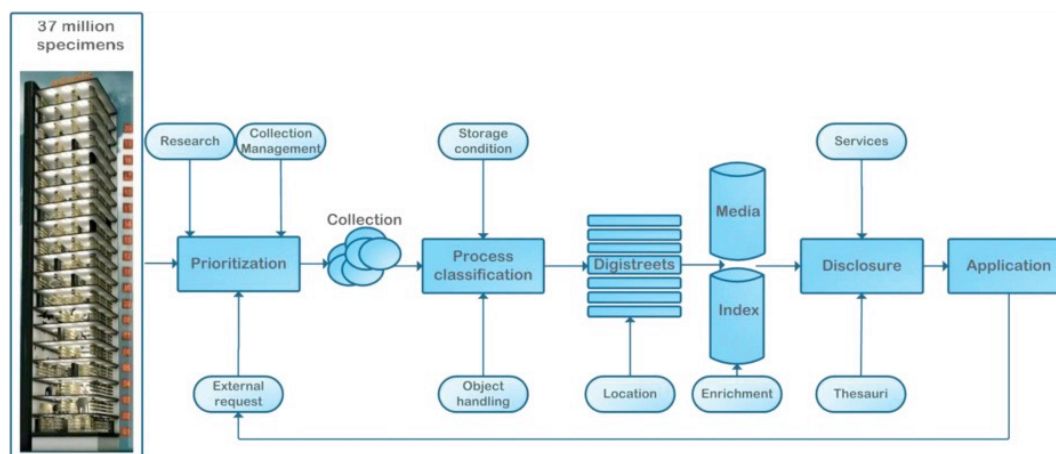


Figure 26 Schematic overview of the Naturalis approach to digitization process organization (from [43])

The Digistreets approach presents a viable solution aimed at minimising rework by sharing expertise while considering the requirements of different collections and materials. The workflows developed by Naturalis are suitable for the characteristics of a large set of institutional collection types. The design is aimed at managing the specimens, processes, images and data aimed at digitising large institutional collections. The examples published are examples of both display (fast low resolution) and research quality (detailed high resolution) specimens (as shown in the examples above Figure 16, Figure 17, Figure 18).

Table 9 Description of the Digistreets workflow tasks

Task (Equivalent Generic Task)	Description
Prioritization (Pre-digitisation curation)	a framework for collection prioritization was developed to facilitate the decision-making process for digitising 7 million artefacts in detail. Takes the form of proposals from scientists and collection managers (description of the collection and arguments for digitization).
Collection (Pre-digitisation curation)	Specimens are collected, recording location, storage container(s) and specimens of interest.
Process Classification (Pre-digitisation curation)	Collected specimens earmarked for digitisation are prepared, this may include curation of specimens, and labelling (of location, storage container and specimen), and transport to digitisation stations.
Digistreets (imaging and image processing)	Specimens are digitised according to the predefined quality criteria.
Return of digitised specimens (Imaging: unmounting and return subtask)	after digitisation has concluded and digitised batches are QA verified, the specimens are returned to storage.
Enrichment (OCR, Data Transcription, Data Entry Correction)	The digital specimens are enriched in post-processing with manual and automated process. These processes can include image curation, generation of image derivatives for different purposes, extraction of metadata from digital specimens (e.g. label data, image data, archival-location data, imaging process data, licensing data).
Disclosure (Publishing*)	The digital specimens are published in the Naturalis portal
Application (Use*)	Digital specimens are used by the community, logging/tracking use can serve to justify the initial digitisation of specimens.

* Not included in generic workflow

4.4.2 Research Quality Imaging (High-Definition) Workflow

Ariño's recommendations to produce taxonomic grade images [10] justifies the need to produce research quality specimen images. An instance of this type of workflow is presented by the Item Driven Image Fidelity (IDIF) approach at the Smithsonian [9] (see Figure 27). IDIF relies on using an example collection item to measure at the micron scale the smallest trait to be resolved. This measurement is then translated into the PPI resolution required to photograph the specimen, which then becomes the resolution required for digitising the collection and to record that trait (or group of traits). PPI resolution is then validated through spatial frequency response (SFR) testing [9]. The following image shows the places of the workflow where the IDIF tasks are integrated and the diagram below it shows the example when imaging an identifying trait for a shell specimen.

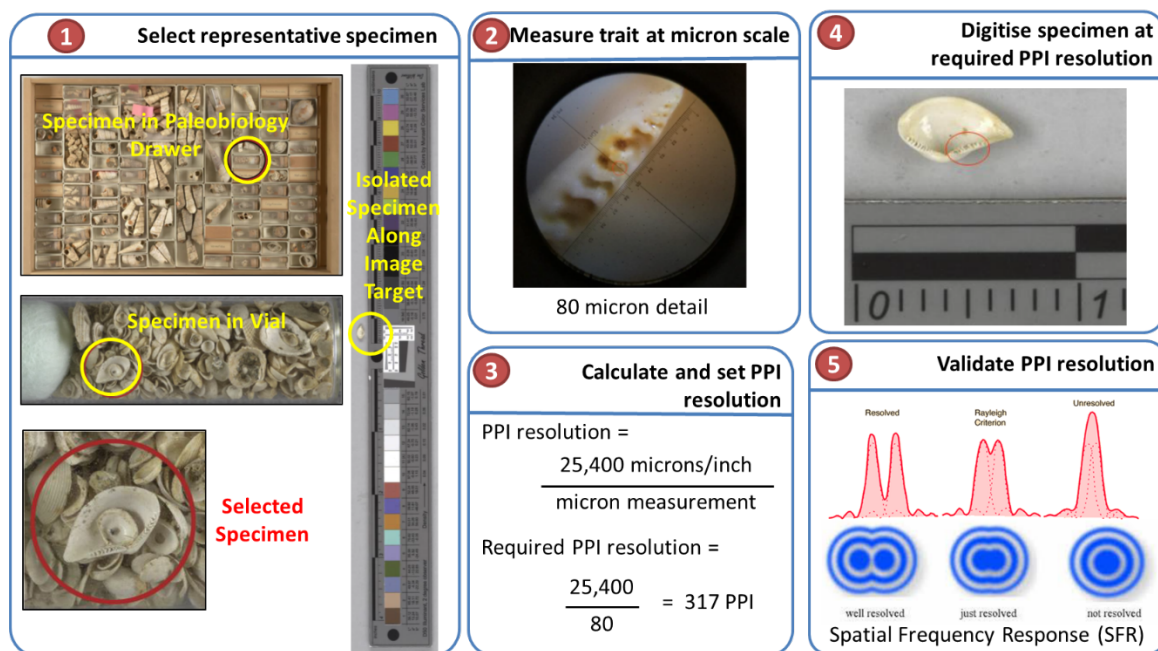


Figure 27 Overview of the Item Driven Image Fidelity process. (1) Select representative specimen displaying trait of interest. (2) Measures the trait of interest at micron scale. (3) Calculate the required PPI resolution for acquiring trait of interest. (4) Digitise specimen at required resolution. (5) Resolution is validated through SFR¹⁶.

The IDIF process is interesting, because its potential for being incorporated into high throughput digitisation workflows. The selection of representative specimens and measurement at micron scale can be performed by the collection curators as part of pre-digitisation curation activities. Then, a member of the digitisation team can calculate the required PPI resolution, image the representative specimen and validate the resolution. Additionally, the validation of resolution can become part of the quality assurance process for checking master images belonging to the same collection.

¹⁶ Diagram composed with images and descriptions from [9]

5. Quality Management for Digitisation of Liquid Preserved Specimens

Liquid preserved specimens encompass specimens preserved in jars filled with various conserving fluids. Liquid preserved specimens can come from all Life Science collections (e.g. animal [8], fungi [90], and plant [91]). This section focuses on the quality management for 2D digitisation methods. For the corresponding methods for quality management for 3D digitisation see section 8.

The variety of collections, preservation media, and containers means that it is not possible to provide a general standardised method for digitisation, such as those developed for collections such as microscope slides (see section 3.1) or herbarium sheets (see section 7.1). Consequently, the guidelines presented here are based on existing literature and examples from natural history collections digitisation projects.

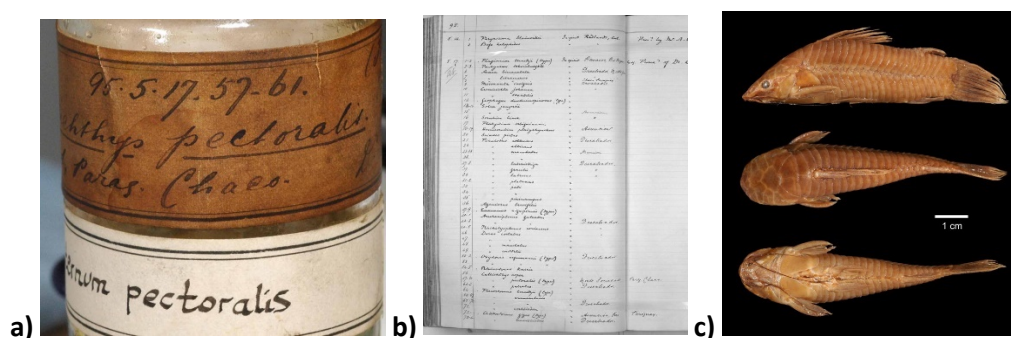


Figure 28 Three types of objects digitised with a liquid preserved specimen. (a) jar labels, (b) register page, and (c), anatomical views (Lateral, Dorsal and Ventral)¹⁷. The three materials are linked to a single specimen of an armoured catfish (*Leptoplocterus pectoralis*). Notice that image c is a composite image (mixed from at least one image per view, i.e. more than one source image)

5.1 Liquid preserved specimens Digitisation Standards

The requirements for digitising liquid preserved specimens should align with the intended use of the digitised specimens. Given the size of the collections, one of the purposes of digitisation is to support the cataloguing of the collections. In this case, the container with labels and specimen is imaged, optionally including images of catalogues, cards or field books. These additional material and then used for transcription and documenting [17] (Section **Fout! Verwijzingsbron niet gevonden.**). A second purpose of specimen digitisation is to support scientific research with research quality digital specimens, as proposed by the Smithsonian Institution and the European Network of Biodiversity Informatics [9, 42].

The Natural History Museum London (NHM) and Naturalis have digitised portions of their liquid preserved specimens collections [67, 69, 83]. The outputs from this effort are already published and available for inspection online. Figure 28, Figure 29, and Figure 30 show examples of liquid preserved specimens published by NHM and Naturalis [67,69]. Figure 28 shows three types of objects which may be digitised alongside the specimen images supporting the transcription and verification of specimen details. Depending on shape and position of the labels, multiple view may be required to capture all label data. The number of views needed to capture research level traits depends on type

¹⁷ NHM Data Portal [67], Permanent URL: <http://data.nhm.ac.uk/object/c21aa2df-5b1e-4560-b7c3-07c1e12a09fe>

of specimen, and whether these traits can be seen at all through the container and potentially obscuring labels. There are two very differing types of liquid preserved specimen imaging: whole container or specimen outside storage container. The former is likely to impact research usability of the specimen image. The latter can be done while the specimen is still submerged in liquid in a clean tank, or outside of the liquid. Then it must be prevented to dry out to limit damage to the specimen. These differences have a great impact on the imaging workflow.

Figure 29 shows specimen images which could be used for research, the specimen in the image requires posing and angles which are different from those of vertebrates. Specimens can also be imaged without removing them from the jar, as illustrated in Figure 30. This kind of image is useful for collection cataloguing.



Figure 29 Multiple views of a liquid preserved octopus (*Octopus fontanianus*) specimen ¹⁸

¹⁸ NHM Data Portal [67], Permanent URL: <http://data.nhm.ac.uk/object/c63ddb85-6db8-416b-8bb8-30fa1c40eba9>



Figure 30 Example of digitisation of liquid preserved specimen in jar published by Naturalis. Japanese fan lobster (*ibacus ciliatus*),¹⁹.

The image in Figure 28c is apparently digitised at a 1 to 1 ratio (600 PPI). However, the image is the result of a composition and there are no physical referents (e.g. scale/ruler) and its use for analysis may be limited. The resolution of the images, however, allows enlarging and viewing the images with up to 10X magnification with acceptable quality, i.e. in focus and not pixelated (Figure 31).

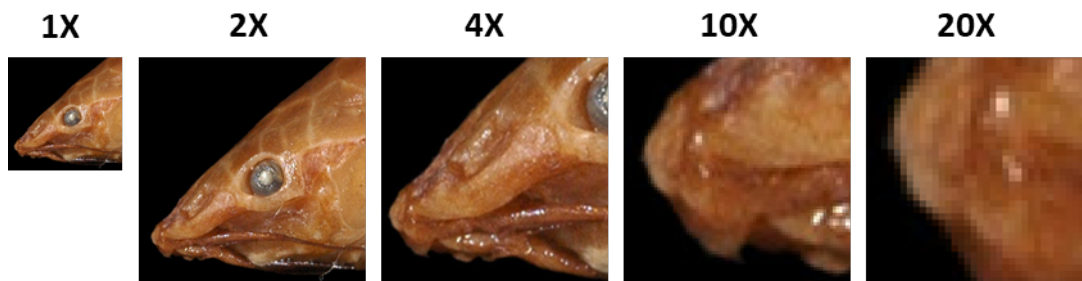


Figure 31 Magnification series of specimen image presented in Figure 28c. Trying to isolate tip of the snout. Magnifications beyond 10X are increasingly pixelated.

5.1.1 Image Resolution and Colour Standards

Table 10 illustrates the criteria for digital specimens of liquid preserved specimens. The details about the different resolutions and uses are derived from the recommendation from the US Library of Congress [61], and FADGI [31], and from calculations derived from existing published specimens. These standards can be used in a QA processes to guarantee that products will meet these standards. Alternatively, standards can be used for evaluating images produced as part of a QC process.

The recommendations for bit depth and colour accuracy are derived from the technical recommendation from the US Library of Congress [61]. Finally, the recommended colour space is Adobe RGB (1998), taken from FADGI [31, 52], combined with the resolutions obtained from the review of existing specimens published by NHM [67].

¹⁹ Naturalis BioPortal [67], URL: <http://data.biodiversitydata.nl/naturalis/specimen/RMNH.CRUS.D.2826>

Table 10 Example of Image Standards for Liquid preserved specimens

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels and Register entries	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Research and identification of specimen features	600 PPI*	24-bit colour		$\Delta E < 5$

*Minimum resolution (from specimen obtained from NHM Data Portal, Figure 28). If digitisation devices available allow for higher resolution, that resolution should be used.

5.1.2 Image Elements

Image elements refer to visual elements which appear next to the specimen and which are intended to help in its identification, processing, and quality control. For research-quality specimens, a list of six elements was derived from the recommendations for scanning type specimens: (1) Specimen, (2) Barcode, (3) Labels, (4) Scale, (5) Background, (6) Container (jar/tank), and (7) Register [50]. The list has been verified by looking at specimen examples from NHM [67]. The elements are not mandatory and can vary depending on the specimen being digitised.

A consideration for liquid preserved specimens is that the shape of the container and the size of the labels can necessitate multiple images to capture full label data (e.g. Figure 32). Additionally, labels can be so numerous or obscured by other elements that they cannot be captured fully. In this case, it can be recommended to also image the catalogue. Figure 32, Figure 33, and Figure 34 show different sets of elements within specimen images (from the NHM and Naturalis).

One of the effects of storing specimens in chemical fluids is the change of colours, consequently, a colour chart is not generally



Figure 32 Recommended elements for digitisation in the container: (1) specimen (2) labels, (3) barcode, (4) Scale and (5) Jar²⁰. The number on the jar corresponds to the digital specimen identifier, and it is different from the barcode number.

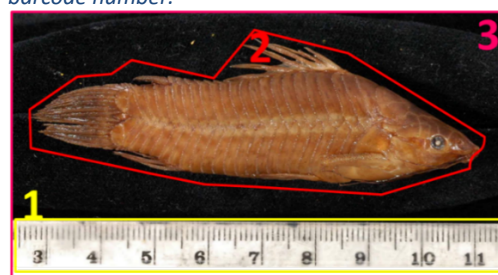


Figure 33 Recommended elements for digitisation: (1) Scale Bar, (2) specimen and (3) background²¹.

²⁰ Naturalis BioPortal [69], URL: <http://data.biodiversitydata.nl/naturalis/specimen/RMNH.CRUS.D.15250>

²¹ NHM Data Portal [67], URL: <http://data.nhm.ac.uk/object/b95f8ab1-bfc2-464d-95f8-9ee7fb07e1a9>

included; however, it could still be helpful to validate the image quality.

Specimen: Liquid preserved specimens are 3D objects, consequently producing sets of images in different poses is required when diagnostic traits are to be imaged [10, 79]. To achieve this, relevant anatomical traits which need to be exposed need to be considered. Figure 35 shows the different planes and axes for fish specimens, but the variety of specimens that are preserved in liquid require customised approaches.

Barcodes are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens [10, 79]. There are different types of barcodes available. Line

Barcodes (one dimensional), have been used in some digitisation projects, however, there were some misidentification problems (because of line features in the specimens). Consequently, the use of two-dimensional barcodes is recommended, and they are being adopted for new digitisation projects [118], as shown in Figure 32.

Labels: are commonly placed on the jar containing the specimen. Clear capture of labels is important for further processing and documentation of the specimens [10, 79]. As shown in the example, capture of labels may be challenging due to the shape of the jar and the placement of more than one label per container. Labels can also be inside the container, free floating or attached to the inside of the container.

Scale Bar: a scale bar is recommended to enable the calculation of the dimensions of the specimen and to identify the collection holding the specimen [10, 79].

Background: The background of digital images of specimens can be changed at will and selected to maximise the contrast with the subject. Light and colour can be used: a background can be much lighter or much darker than the relative brightness of the specimen, alternatively it could be of a

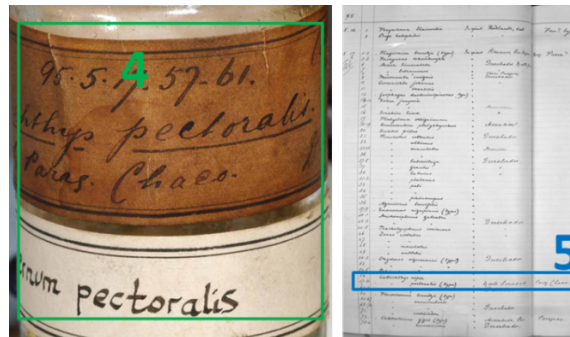


Figure 34 Image showing recommended elements for digitisation of liquid preserved specimens. (4) labels (container), and (5) Register entry (Entry related to specimens highlighted on page)²². The label indicates that the container holds five specimens (57-61), of the same species, also included in the register page image.

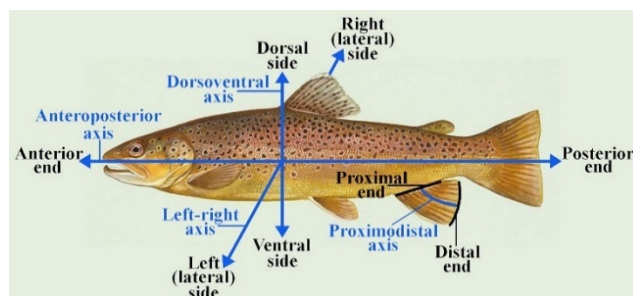


Figure 35 Anatomical directions and defined axes in a fish²³.

²² NHM Data Portal [67], URL: <http://data.nhm.ac.uk/object/c21aa2df-5b1e-4560-b7c3-07c1e12a09fe>

²³ Taken from [105]

colour that is not present on the specimen. The main goal is to clearly separate features, especially along the contour of the specimen, from the image background [10].

Container: The container can be transparent and imaging it can also serve to show the specimen(s) contained within; for this the container, specimen and fluid may require curation before digitisation [17, 56]. Notice that in the example presented in Figure 32, the number in the jar is part of the specimen identifier, and different from the assigned barcode number. The label and register in the NHM specimens are also used to identify the specimens, but their URL identifier is different from these numbers (see Figure 34). Many historic containers are in use in collections. As with the design of labels and the distinguishing features of the text, the container may provide insight in the history of the specimen. It may turn out to be one of the things that are not valued today but may provide specific information in the future.

Register/Catalogue: physical registries or collection catalogues are an important part of the collection that have need to be digitised by institutions. Even if the information from the register has been entered into a collection management system, registers contain relevant provenance information, it is recommended to digitise and preserve them together with the digital specimen [17].

5.2 Liquid preserved specimens Digitisation Workflows

Wet collections have been one of the least imaged collection types. The variety of taxonomic collections that they pertain to, the hazards of glass and the liquids contained in them, the optical distortion of imaging through glass all add to the complexity and slowness of the required workflows. When it is necessary for research purposes to image the specimen outside its container (either dry or in a dedicated tank) the imaging speed is seriously impacted. Digitisation systems for this type of collection are developed in-house, with custom-made workflows that match the resources available at each institution [50, 70, 77, 79].

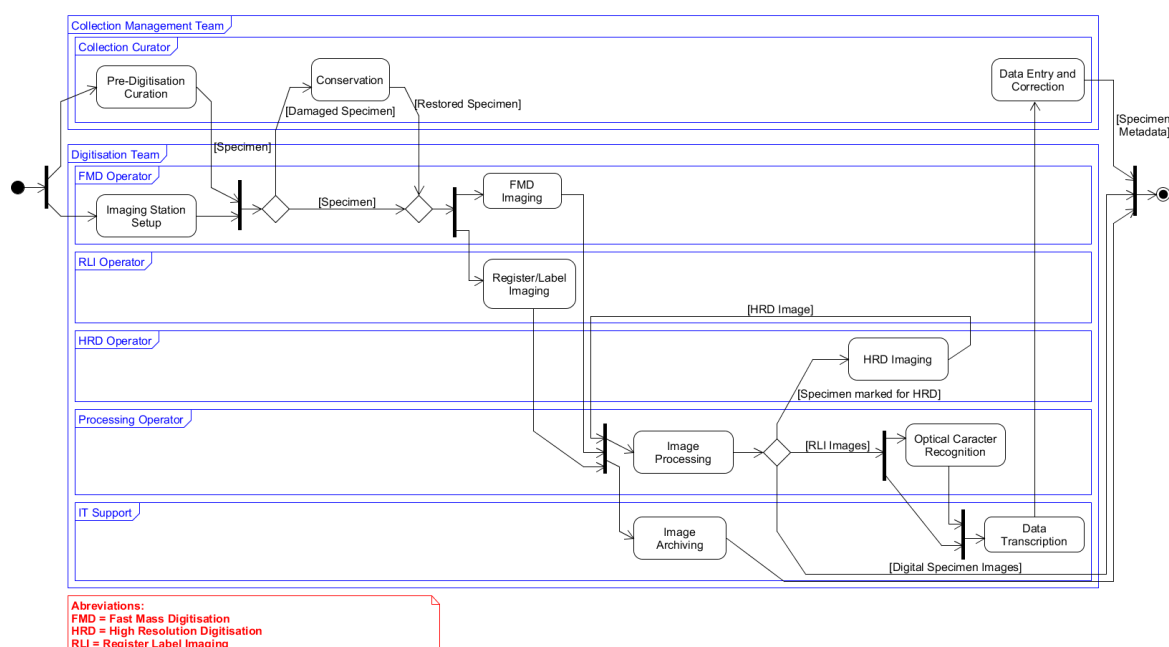


Figure 36 Generic Liquid Preserved Specimens Digitisation Workflow

Figure 36 shows a high-level view of a liquid preserved specimen digitisation workflow, derived from the digitisation activities recommendations from Ariño, iDigBio, and Synthesys3 [10, 50, 70, 79, 85].

The tasks in the diagram coincide with those defined by iDigBio [50, 70] including the alternate paths for identification materials (labels and register entries).

The diagram includes the activities which can occur in a digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, some digitisation projects may require only “specimen digitisation”, because labels and registries have been digitised already. Similarly, some collections may have already captured data and therefore can omit the “OCR”, “data transcription” and the “data entry and correction” tasks [79]. The number of images per specimen also influences the final shape of the workflow instances. Table 11 provides a brief description of the workflow tasks and their influence in the production of images.

Table 11 Description of liquid preserved specimens digitisation workflow tasks, derived from iDigBio recommendation and observed practice [50, 70, 77, 79].

Workflow Task	Description/Subtasks	Influence on Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); identification of Specimens requiring posing for different views; defining safeguards for identifying and handling specimens in stored in hazardous fluids (e.g. alcohol, formalin, etc); transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators.
Imaging Station(s) Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians. If digitisation of labels and registers are required, additional stations may need to be installed and operated in parallel.	Equipment should be calibrated to minimise image post-processing.
Conservation	Required conservation/restoration of specimens selected for digitisation. If digitisation of labels and registers are required, additional conservation for ledgers, cards or fragile bounded volumes may be required.	Some specimens may be damaged, fragile, or require to be re-mounted to display all relevant traits.
Imaging	Data capture; identification (barcoding); mounting for imaging (may require removing specimens from the container and using alternative digitisation mounts(e.g. tank containers or dry mounts)); imaging of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta] data capture, so that images are correctly linked to the corresponding specimen records.
Register/Label Imaging	Imaging of identification and data materials associated with the specimens. Scanning of register materials (cards, ledgers, bound books)	Acquisition of metadata from physical specimen records and labels.
High Resolution Imaging	Identification of specimens requiring high-definition scanning, identification of poses for specimen, high resolution imaging, creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta]data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of derivatives for publishing and distribution; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
Image Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.



Table 11 Description of liquid preserved specimens digitisation workflow tasks, derived from iDigBio recommendation and observed practice [50, 70, 77, 79].

Workflow Task	Description/Subtasks	Influence on Image Quality
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data Transcription	Manual extraction of data from images, for populating/complementing specimen record.	Verify readability of derived image for data transcription.
Data entry and correction	Final capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing.

The main variations on the characteristics of digitisation workflows depend on the metadata required (or already present) and the type of digitisation stations to be used. The following sections describe how the activities for liquid preserved specimen digitisation are affected by these two features [70, 79, 85].

5.2.1 Metadata Required for Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record on the collection management system. Additionally, specific data which is stored with the specimen such as scientific name, field notes and collection name should also be included. In liquid preserved specimen digitisation, data are normally contained in labels which are attached to the specimen container (jar) and in the physical registries in the form of cards, ledgers, or books [17]. This is relevant for imaging because the images should capture the relevant metadata clearly for its later transcription and verification. As described in the section on image elements (5.1.2), this requires producing additional images. The extraction of data from these images can entail automated or manual activities which can take place before or after digitisation. Alternatively, some institutions may already have portions of the metadata stored in a collection management system, in which case the most important information is the identification data. For these specimens there is no naming convention, however, many collections use the specimen ID to name the imaged specimen files. If imaging requires manual handling of specimens, then it is important to have metadata about the preservation fluids used to take needed precautions [17].

5.2.2 Digitisation equipment

Most of the digitisation stations for liquid preserved specimens are custom-made and camera based. Specimen positioning and imaging is mostly a manual process, and there are no examples of automated digitisation stations [70, 77, 79, 85, 98]. Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. Sabaj [85] documented the use of custom-made display tanks (“photo-tank”) which facilitate imaging. Similarly, the preservation techniques suggested by Kingham [56] support the preparation of specimens for display which allow digitising specimens in their containers. Additionally, there are some components, such as focus stacking rails and turntables, from various vendors, which can produce semiautomated stations with different levels of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen in a position convenient for imaging. The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. Synthesys3 reported a huge variation in the range of 1,000 to 10,000 specimens per year, with costs in the range from 2.00 € to 16.00 € per specimen [79].



5.3 Liquid preserved specimens Digitisation Quality Management Methods

In line with the imaging standards for liquid preserved specimens (5.1), the quality management methods focus on the quality of the digitised specimen and the presence of the recommended image elements. The quality management methods should verify the presence and clarity of image elements (spatial resolution), colours (spectral resolution), and the perception of brightness in the image through the number of the grey value levels (radiometric resolution). In this case, some elements can be used to support these verifications. For instance, lines on the scale elements can be used as a reference to verify the spatial resolution of the image.

Quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are realised should be those that follow the initial digitisation of the specimen, its related materials, and after the production of each derivative.

5.4 Examples

The examples of implementation of liquid preserved specimen digitisation workflows are mainly in-house camera-based systems, in which the hardware and software are sourced from the best available off the shelf components within the institutional budgets. The institution hosting the collection procures and operates the digitisation equipment, designs the digitisation process, performs the digitisation, and manages the digitised collection. These setups can produce research quality digital specimens [43, 85]. The digitisation workflows for these materials can be further divided into fast mass digitisation workflows [43] and research quality imaging (high-definition) workflows [85]. The following examples illustrate these two workflow types.

5.4.1 Fast Mass Digitisation Workflow

The approach taken by Naturalis, called Digistreets, entails the development of a series of workflows which are targeted at mass digitisation of collections [43]. In this case the requirements of various collection types were analysed to develop a family of digitisation workflows with a common approach, which allowed the sharing of expertise and resources among the digitisation teams and the collection curators. Figure 37 presents a high-level view of the Digistreets approach and Table 12 presents a description of the tasks and a mapping to the generic tasks of the digitisation workflow.

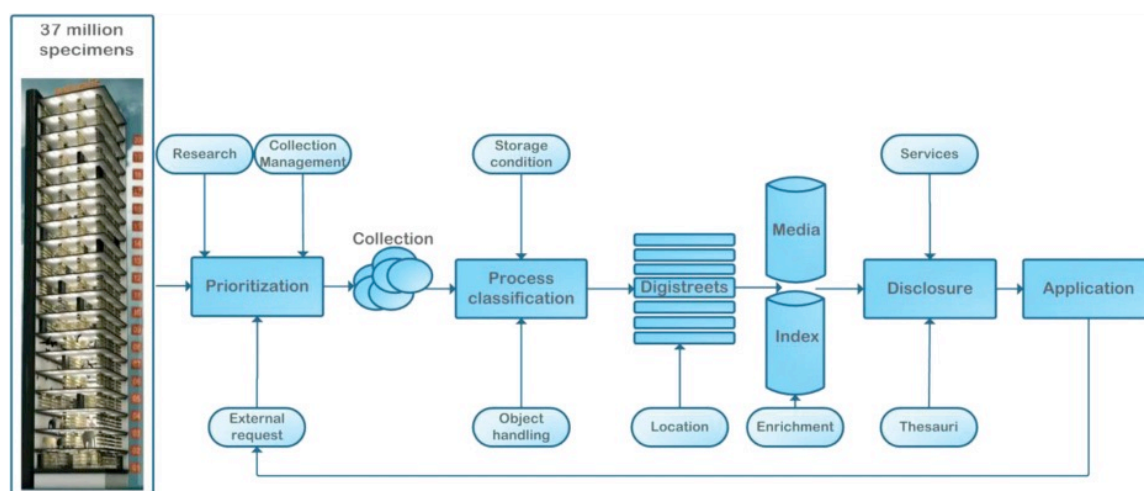


Figure 37 Schematic overview of the Naturalis approach to digitization process organization (from [43])

The Digistreets approach presents a viable solution aimed at minimising rework by sharing expertise while considering the requirements of different collections and materials. The workflows developed

by Naturalis are suitable for the characteristics of a large set of institutional collection types. The design is aimed at managing the specimens, processes, images and data aimed at digitising large institutional collections. The examples published are examples of display (fast imaging documentation) specimens, as shown in Figure 30 and Figure 34 above.

Table 12 Description of the Digistree workflow tasks

Task (Equivalent Generic Task)	Description
Prioritization (Pre-digitisation curation)	a framework for collection prioritization was developed to facilitate the decision-making process for digitising 7 million artefacts in detail. Takes the form of proposals from scientists and collection managers (description of the collection and arguments for digitization).
Collection (Pre-digitisation curation)	Specimens are collected, recording location, storage container(s) and specimens of interest.
Process Classification (Pre-digitisation curation)	Collected specimens earmarked for digitisation are prepared, this may include curation of specimens, and labelling (of location, storage container and specimen), and transport to digitisation stations.
Digistree (imaging and image processing)	Specimens are digitised according to the predefined quality criteria.
Return of digitised specimens (Imaging: unmounting and return subtask)	after digitisation has concluded and digitised batches are QA verified, the specimens are returned to storage.
Enrichment (OCR, Data Transcription, Data Entry Correction)	The digital specimens are enriched in post-processing with manual and automated process. These processes can include image curation, generation of image derivatives for different purposes, extraction of metadata from digital specimens (e.g. label data, image data, archival-location data, imaging process data, licensing data).
Disclosure (Publishing*)	The digital specimens are published in the Naturalis portal
Application (Use*)	digital specimens are used by the community, logging/tracking use can serve to justify the initial digitisation of specimens.

* Not included in generic workflow

5.4.2 Research Quality Imaging (High-Definition) Workflow

Producing taxonomic grade images relies on special workflows which are slower than mass digitisation workflows [10]. The photo-tank immersion method is an example of a research quality digitisation workflow [85] (see Figure 38). The workflow consists of three stages: equipment set up, specimen preparation, and image capture and editing. This method is suitable for live and preserved specimens. The equipment set up stage includes the construction of the tank, selection of background and scale elements, selection of camera, lighting and mounting artefacts. Note that DNA and small particle contamination is a risk: in between specimens the tank and liquid needs to be cleaned. Specimen preparation includes curation, posing and immersion of specimen in photo-tank, including recommendations for posing and handling which require special care to avoid damaging specimens. Image capture and editing includes the imaging of specimen, required exposure, lighting, image processing and storage. Figure 39 shows the use of photo-tank method in the lab for imaging a liquid preserved specimen. The photo-tank immersion method is interesting because it allows photographing specimens in fluid without placing them on a flat surface, preventing distortion and drying out.



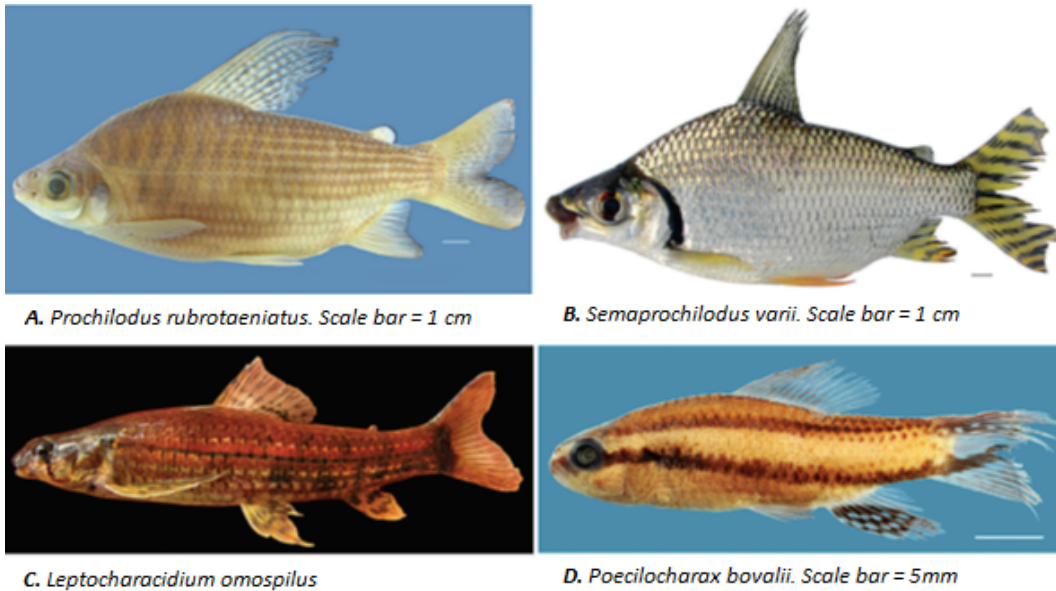


Figure 38 Specimens imaged using the photo-tank method (taken from [85]). A, and D correspond to alcohol preserved specimens, while B and C correspond to live specimens.

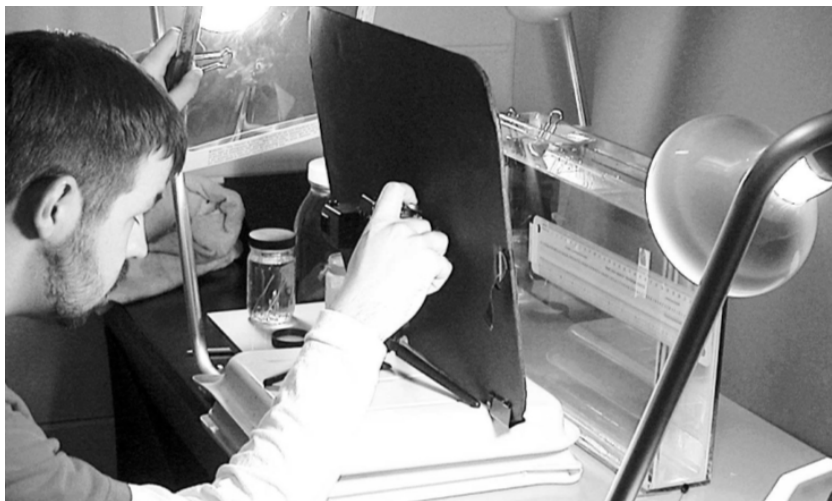


Figure 39 Imaging of liquid preserved specimen (above ruler) using the photo-tank method (from [85]). Notice that the digitisation operator is holding a polarizing filter while capturing the specimen image.

6. Quality Management for Digitisation of Pinned Insects

Pinned insects encompass dried preserved specimens of invertebrates stored in trays and drawers. Because of the similarities in storage and preservation methods, these workflows can include also dry marine invertebrates and shells. All these specimens can be digitised using 2D and 3D imaging methods. This section focuses on the quality management for 2D digitisation methods. For the corresponding methods for quality management for 3D digitisation see section 8.

The variety of collections, preservation media, and containers means that it is not possible to provide a general standardised method for digitisation such as those developed for homogenous collections such as microscope slides (see section 3.1) or herbarium sheets (see section 7.1). Accordingly, the guidelines presented here are based on existing literature and examples from digitisation projects of pinned insects.

6.1 Pinned Insect Digitisation Standards

The requirements for digitising pinned insects should align with the intended use of the digitised specimens. Given the size of the collections, one of the purposes of digitisation is to support the cataloguing of the collections. In this case, specimens, ledgers and labels are imaged and then used for transcription and documenting [14]. A second purpose of specimen digitisation is to support scientific research with research quality digital specimens, as proposed by the Smithsonian Institution and the European Network of Biodiversity Informatics [9, 42].

The Natural History Museum London (NHM), Naturalis Biodiversity Centre (Naturalis), and the Finnish Museum of Natural History (LUOMUS) have active digitisation projects for pinned insect collections [34, 67, 69]. The outputs from this effort are already published and available for inspection online. Figure 40 and Figure 41 show examples of pinned insect specimens published by LUOMUS [34].

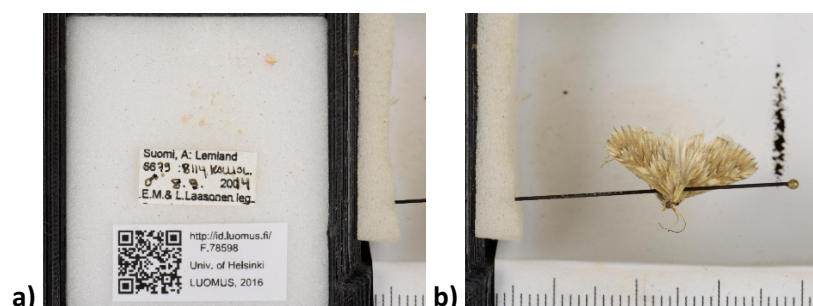


Figure 40 Example of objects digitised with a pinned insect specimen. (a) label(s) and barcode, (b) specimen ²⁴. The three materials are linked to a single specimen of a moth (*Cynaeda dentalis*).

Figure 40 shows some of the elements which may be digitised alongside the specimen images supporting the transcription and verification of specimen details. Figure 41 shows specimen images which could be used for research. Observe also that this type of collections requires more than one view for displaying different specimen characteristics.

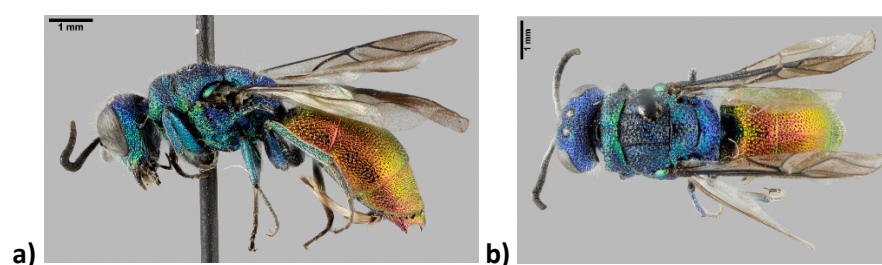


Figure 41 Multiple views of a pinned wasp (*Chrysis horridula*) specimen ²⁵

Figure 42 shows two alternative digitisation methods for pinned insect specimens: full drawer view and single specimen with identifiers and processing elements (scale and colour chart).

²⁴ Finnish Biodiversity Information Facility [34], URL: <http://id.luomus.fi/F.78598>

²⁵ Finnish Biodiversity Information Facility [34], URL: <http://id.luomus.fi/GL.6584>

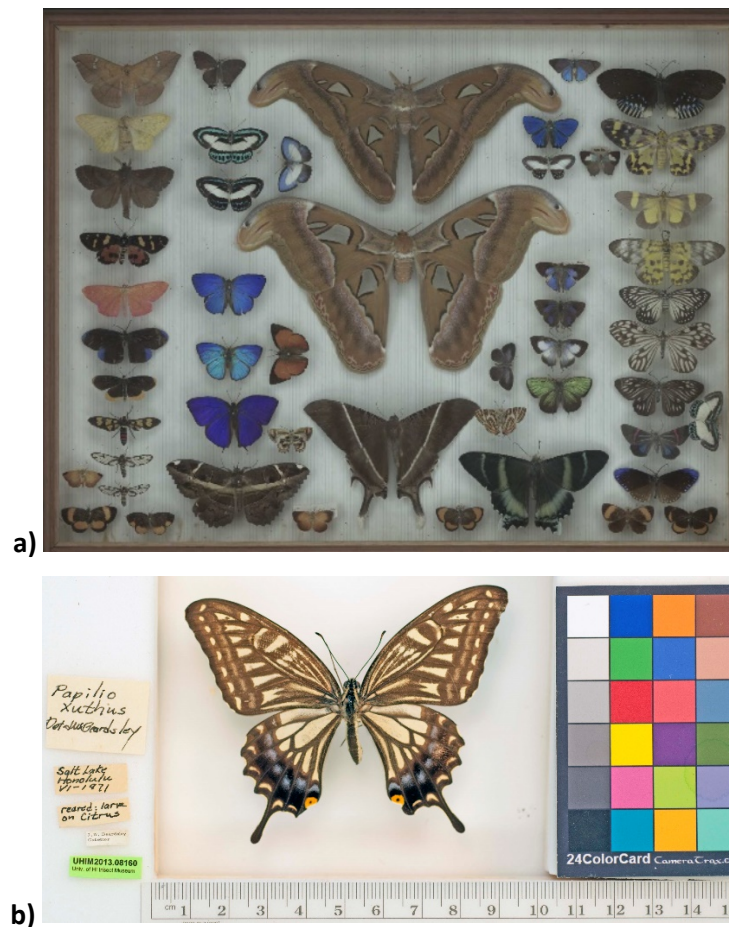


Figure 42 Examples of pinned insect specimens suitable for 2D digitisation. Image “a” shows a butterfly drawer from the Wallace and Banks drawers collection²⁶. Image “b” shows a butterfly specimen published by the Atlas of Living Australia along with identifiers (labels) and processing elements (scale and colour chart)²⁷

The images in Figure 41 are (nominally) stored at 300 PPI, however, their actual resolutions are 4,038 PPI for ‘a’ and 2,337 PPI for ‘b’ (calculated from the 1 mm scale bar). To achieve this resolution at this scale means that the images are photo stacking composites, to deal with reduced depth of field. The resolution of the image in Figure 41a is equivalent to a magnification of 42X (see Figure 43). In comparison, the image in Figure 42b with a calculated resolution of 589PPI (nominal 240 PPI) can be at most enlarged to 10X before being too pixelated (see Figure 44).

²⁶ NHM Data Portal [67], documented as part of [84], drawer 13, url: <http://data.nhm.ac.uk/dataset/wallace-and-banks-drawers>

²⁷ Atlas of Living Australia [11], URL: <https://bie.ala.org.au/species/NZOR-4-23691>

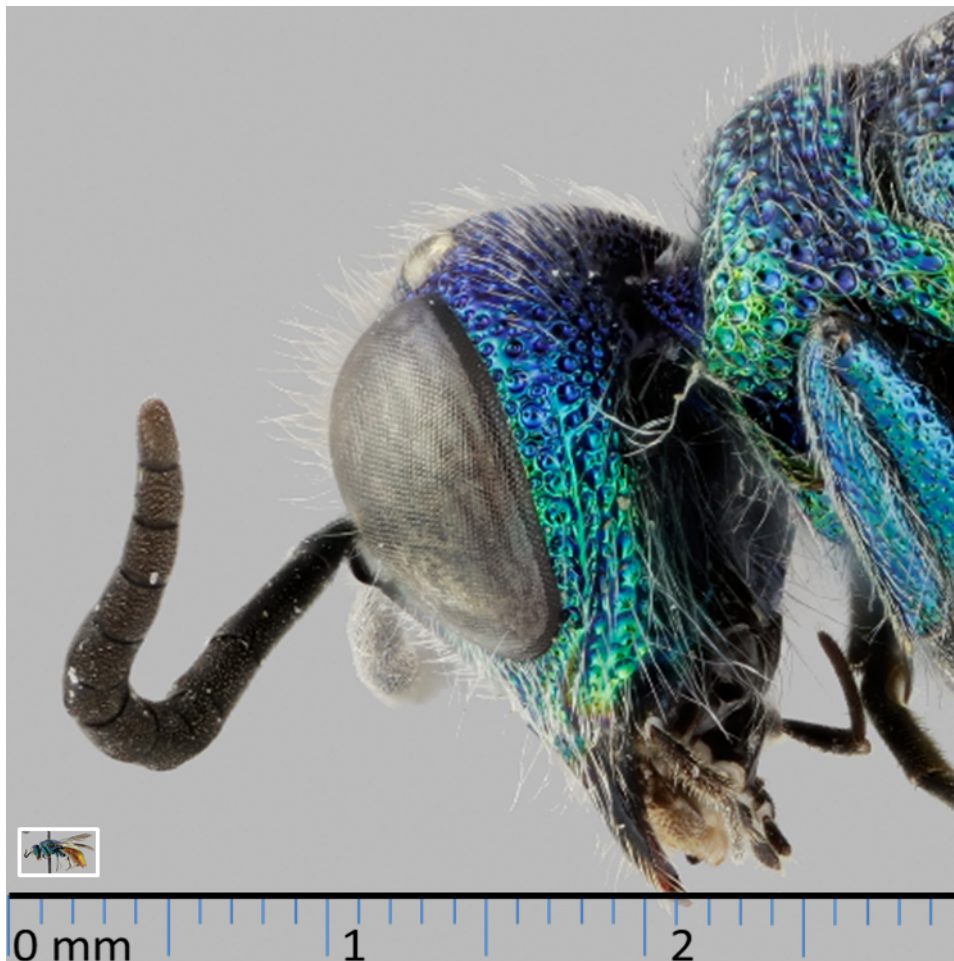


Figure 43 3mm section at 100% of specimen from Figure 41a. The calculated resolution for this image is 4038 PPI, or 159 pixels/mm. This produces a magnification of 42X. The small image in the lower right corner, above the scale bar shows the full image as it looks at 1 to 1 size.



Figure 44 Magnification series of specimen image presented in Figure 42b, trying to isolate the right eye. Magnifications of 10X and up are increasingly pixelated. The image is at 240 PPI, but its actual resolution is 589 PPI (estimated using the scale bar).

6.1.1 Image Resolution and Colour Standards

Table 13 illustrates the criteria for digital specimens of pinned insect specimens. The details about the different resolutions and uses are derived from the recommendation from the US Library of Congress [61], and FADGI [31], and from calculations derived from existing published specimens (Figure 40 – Figure 44). These standards can be used in a QA processes to guarantee that products

will meet these standards. Alternatively, standards can be used for evaluating images produced as part of a QC process.

The recommendations for bit depth and colour accuracy are derived from the technical recommendation from the US Library of Congress [61]. Additionally, the recommended colour space is Adobe RGB (1998), taken from FADGI [31, 52], combined with the resolutions obtained from the review of existing specimens published.

Table 13 Example of Image Standards for Pinned Insects

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels and Register entries	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Research and identification of specimen features	600 PPI*	24-bit colour		$\Delta E < 5$

*Minimum resolution for research quality digital specimens (approximated from specimen from ALA, Figure 42b). If digitisation devices available allow for higher resolution, that resolution should be used (as in the example form LAJI.FI).

6.1.2 Image Elements

Image elements refer to visual elements which appear next to the specimen and which are intended to help in its identification, processing, and quality control. For research-quality specimens, a list of six elements was derived from the recommendations for scanning type specimens: (1) Specimen, (2) Barcode, (3) Labels, (4) Scale, (5) Background, and (6) Container (drawer/tray) [50]. The list has been verified by looking at specimen examples from the Atlas of Living Australia (ALA) and Finnish Biodiversity Information Facility (FinBIF). Naturalis and NHM. Image elements are not mandatory and can vary depending on the specimen being digitised. Another consideration for pinned insects is that labels are sometimes imaged separate from the specimens. Keeping track of which images of labels and specimens go together has an impact on the design of the digitisation workflows.

Figure 45 shows labels, scale, and colour chart next to the specimen, while Figure 40a show the barcode and Figure 42 the drawer.



Figure 45 Pinned insect digitisation elements (same as Figure 42). (1) specimen, (2) labels, (3) scale, and (4) colour chart.

Specimen: pinned insects are 3D objects, consequently it can be necessary to produce sets of images in different anatomical views to capture research level details [10, 79]. To achieve this, relevant anatomical traits which need to be exposed need to be considered when posing the specimen. Figure 46 and Figure 47 show some of the relevant anatomical features of insects, some are common (head, thorax, abdomen leg segments) and other are specialised (e.g. elytra).

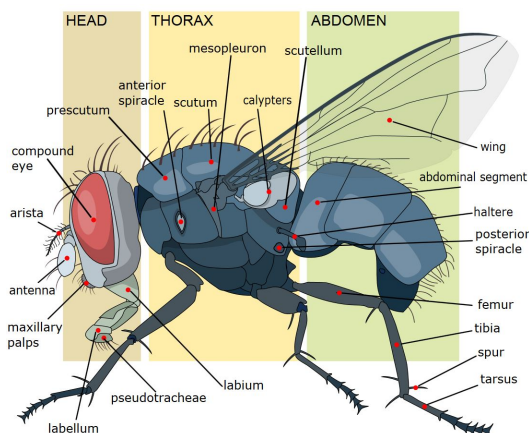


Figure 46 Insect morphology illustrated using a house fly²⁸.

Barcode are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens [10, 79].

There are different types of barcodes available. Line Barcodes (one dimensional), have been used in some digitisation projects, however, there were some misidentification problems (because of line features in the specimens). Consequently, the use of two-dimensional barcodes is recommended, and they are being adopted for new digitisation projects [118].

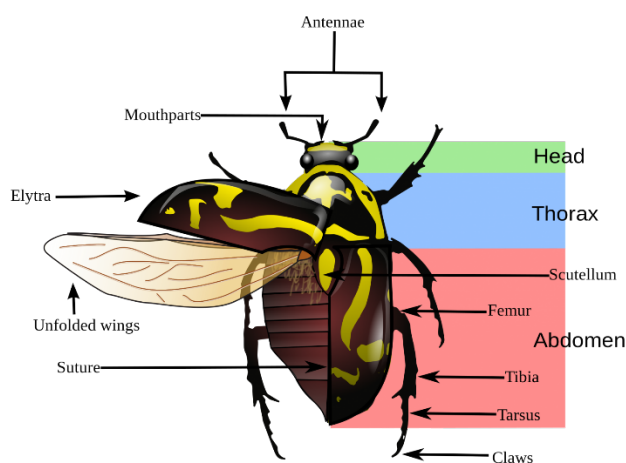


Figure 47 Beetle morphology²⁹.

Labels: are commonly placed on the pin holding the specimen. Clear capture of labels is important for further processing and documentation of the specimens [10, 79]. As shown in the examples capture of labels may be challenging due capture angle and the placement of more than one label per specimen. This is sometimes handled by removing labels from the pin and placing them next to the specimen during digitisation (see Figure 40a and Figure 42b).

Scale Bar: a scale bar is recommended to enable the calculation of the dimensions of the specimen and to identify the collection holding the specimen [10, 79] (see Figure 40b and Figure 42b).

Background: The background of digital images of specimens can be changed at will and selected to maximise the contrast with the subject. Light and colour can be used: a background can be much lighter or much darker than the relative brightness of the specimen, alternatively it could be of a

²⁸ Taken from [101]

²⁹ Taken from [101]

colour that is not present on the specimen. The main goal is to clearly separate features, especially in the contour of the specimen, from the image background [10].

Container(s): Pinned insect drawers can contain specimens from a single species, specimens from related specimens or specimens from various species (see Figure 42a). The number of specimens per drawer can vary, depending on the size of specimens and the type of specimens stored in it. Additionally, within drawers, specimens can be contained in trays which can also vary in their size and the number/type of specimens they contain [79, 50].

6.2 Pinned Insects Digitisation Workflows

Three variants of pinned insect digitisation workflows are often recognised: individual specimen, whole drawer and high definition [79, 50]. 3D imaging workflows focussing on insects are in development, see section 8. The speed of these workflows is limited by the fragility of the specimens, the number of shots per specimen for multiple views and focus stacking, the need to remove or reposition labels and the level of documentation of the collection.

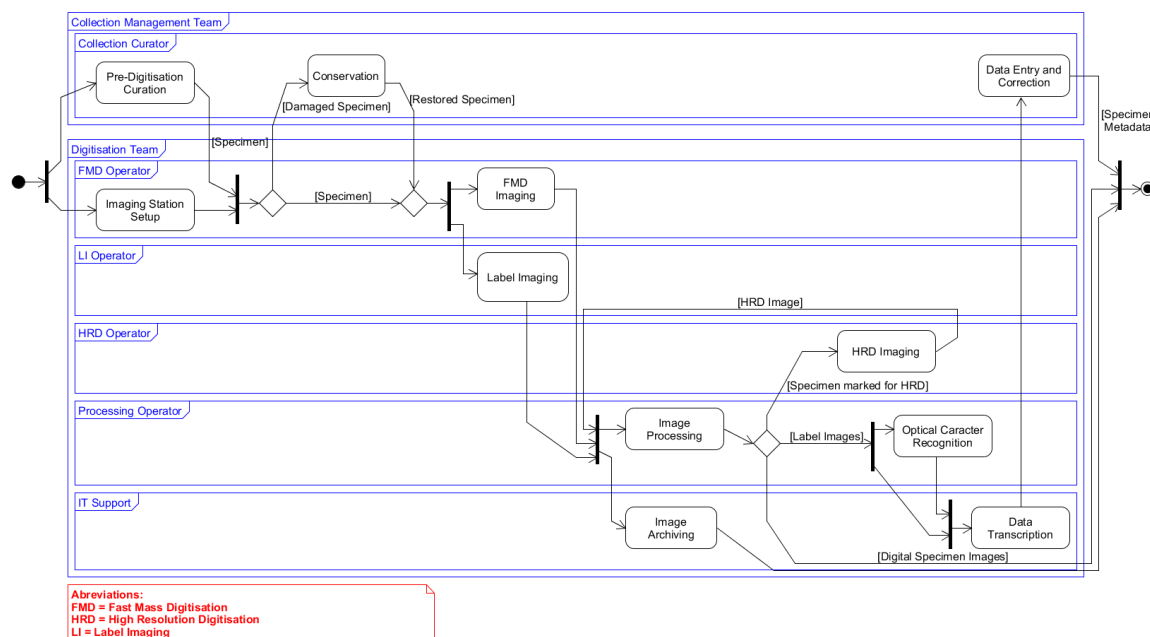


Figure 48 Generic Pinned Insect Specimens Digitisation Workflow

Figure 36 shows a high-level view of a Pinned Insect digitisation workflow, derived from the digitisation activities recommendations from iDigBio, and Synthesys3 [50, 79, 98]. The tasks in the diagram coincide with those defined by iDigBio [50, 70] including the alternate/complementary path for labels.

The diagram includes the activities which can occur in a digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, some digitisation projects may require only “specimen digitisation”, because labels and registers have been digitised already. Similarly, some collections may have already captured data and therefore can omit the “OCR”, “data transcription” and the “data entry and correction” tasks [79]. The number of images per specimen also influences the final shape of the workflow instances. Table 14 provides a brief description of the workflow tasks and their influence in the production of images.

Table 14 Description of pinned insect specimens digitisation workflow tasks, derived from iDigBio recommendation and observed practice [50, 79].

Workflow Task	Description/Subtasks	Influence on Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); identification of Specimens requiring posing for different views; defining safeguards for identifying and handling fragile/valuable specimens; transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators. A major limiting factor is the barcoding of specimens. Adding these to each specimen is laborious, especially when they are added to the pin. In case of drawer-level imaging, this can mean that the specimens don't have a specimen-level barcode because the amount of work required and the overlapping of specimens in trays and drawers.
Imaging Station(s) Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians. If digitisation of labels is required, additional stations may need to be installed and operated in parallel.	Equipment should be calibrated to minimise image post-processing.
Conservation	Required conservation/restoration of specimens selected for digitisation.	Some specimens may be damaged, fragile, or require to be re-mounting to display all relevant traits and labels.
Imaging	Data capture; identification (matching/linking IDs and barcodes); mounting for imaging (may require different mounts for different poses); imaging of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta] data capture, so that images are correctly linked to the corresponding specimen records.
Label Imaging	Imaging of specimen labels (angles, mirrors, depth of field cameras)	Acquisition of [meta]data from physical specimen labels.
High Resolution Imaging	Identification of specimens requiring high-definition scanning, identification of poses for specimen, high resolution imaging, creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta]data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of derivatives for publishing and distribution; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
Image Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data Transcription	Manual extraction of data from images, for populating/complementing specimen record.	Verify readability of derived image for data transcription.
Data entry and correction	Final capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing.

The main variations on the characteristics of digitisation workflows depend on the metadata required (or already present) and the type of digitisation stations to be used. The following sections describe how the activities for pinned insect specimen digitisation are affected by these two features [70, 79, 85].



6.2.1 Metadata Required for Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record in the collection management system. Additionally, specific data which is stored with the specimen such as scientific name, field notes and collection name should also be included. In pinned insect specimen digitisation, data are normally contained on labels which are attached to the pin holding the specimen [79, 85]. This is relevant for imaging because the images should capture the relevant metadata clearly for its later transcription and verification, potentially through OCR. This may require producing additional images from different angles and the use of additional equipment. The extraction of data from these images can entail automated or manual activities which can take place before or after digitisation. Alternatively, some institutions may already have digitised these metadata in their collection management system or have these metadata contained in easily digitised catalogues, in which case the most important information is the identification data. For these specimens there is no naming convention, however, many collections use the specimen ID (barcode) to name the imaged specimen files.

6.2.2 Digitisation equipment

Pinned insect digitisation stations are camera based, specimen positioning and imaging can be manual or semiautomated. Manual stations have dedicated digitisation technicians which position, digitise and return each specimen. Semiautomated stations have various forms which can vary in the amount of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen in front of the camera(s). The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. Currently, the fastest digitisation rates correspond to conveyor belt-based stations [98] that can be 5 to 10 times faster than manual digitisation stations.

6.3 Pinned Insect Digitisation Quality Management Methods

In line with the imaging standards for pinned insect specimens (see section 6.1), the quality management methods focus on the quality of the digitised specimen and the presence of the recommended image elements. The quality management methods should verify the presence and clarity of image elements (spatial resolution), colours (spectral resolution), and the perception of brightness in the image through the number of the grey value levels (radiometric resolution). In this case, some elements can be used to support these verifications. For instance, lines on the scale elements can be used as reference to verify the spatial resolution of the image. The colour chart can support the verification of spectral and radiometric resolution.

Quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are realised should be those that follow the initial digitisation of the specimen, its related materials, and after the production of each derivative.

6.4 Examples

The examples of implementation of pinned insect digitisation workflows have three variants: whole drawer, conveyor-belt and manual. For whole drawer digitisation there are inhouse custom built stations and ready built solutions, e.g. SatScan is a widely used solution. For conveyor-belt based digitisation, Digitalium has demonstrated the adaptation of a solution developed for herbarium sheet digitisation. NHM has tested different configurations for manual stations. In all cases, the institution hosts the collection process, operates the digitisation equipment, and performs the



digitisation. Digitalium can offer an outsourced digitisation solution, for different parts of the digitisation workflow. The following subsections illustrate these workflow variants.

6.4.1 Whole Drawer Digitisation Workflow

Digitisation of whole drawers is a rapid digitisation method which has been used by NHM, Naturalis, and Berlin Museum of Natural History (Berlin MfN), among others. This approach is based on the use of SatScan [63]. The SatScan digitisation station captures sequential “tile” images (200 – 400 per drawer) and its custom software digitally stitches the tile images to produce a whole drawer image (Figure 49). The institutions can then decide whether to publish a zoomable image of the whole drawer, segment the image to present individual pictures of specimens, or allow further processing (annotation, download, etc).



Figure 49 SatScan Digitisation Station from [63]

The whole drawer approach presents a viable solution aimed at rapid documentation of entomology collections. The design is aimed at digitising large institutional collections, as shown in the example above (Figure 42a).

6.4.2 Conveyor Belt-Based Digitisation Workflow

An instance of this type of workflow was developed by Digitalium [98] and is currently being used at LUOMUS and Berlin MfN (see Figure 50). The station can be operated by a single person with minimal training. The operator picks one specimen from the drawer, removes the labels and places them together in a custom-built digitisation tray, next to a 2D barcode. The digitisation trays are then placed on a conveyor belt that moves them in front of different cameras which will take a picture of the labels, and one or two pictures of the specimen. Digitised specimens are then removed from the digitisation trays, then labels are returned to their position, and a new label with the assigned barcode is added.



Figure 50 Pinned insect digitisation station), developed by Digitalium [98], at Berlin MfN.



6.4.3 Manual Single Specimen Digitisation Workflow

The digitisation workflow of the Natural History Museum London is an example of a single specimen digitisation workflow developed in house. Figure 51 shows an example of a digitisation station for single specimens. The operator picks one specimen from the drawer, removes the labels and places them together in a custom-built digitisation tray, next to a 2D barcode. The digitisation tray is then placed on a light box, in front of a camera. The operator then takes a picture of the labels and specimen. If more poses are required, the operator changes the specimen position and takes another picture. Digitised specimens are then removed from the digitisation trays, then labels are returned to their position, and a new label with the assigned barcode is added. Figure 52 shows two prototypes of digitisation stations for specimen multiple views of specimens and labels which do not require removing them from the pin [82]. The positioning of the specimen in the digitisation tray is still a manual process.

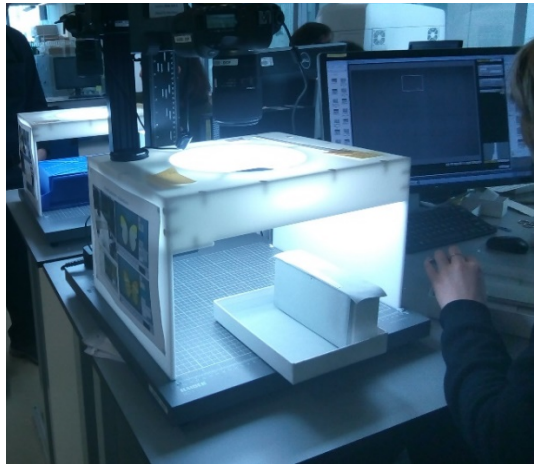


Figure 51 Single specimen digitisation station at the Natural History Museum [6].

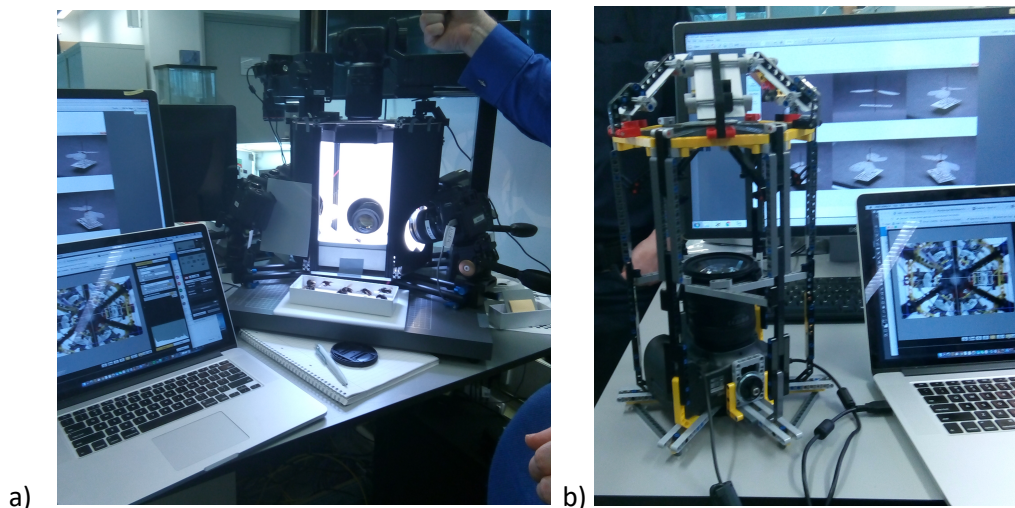


Figure 52 Automated Label Imaging Capture Equipment (ALICE) developed by the Science Informatics team of the Natural History Museum London [6, 82]. Image 'a' shows the setup with six cameras to capture the labels from different perspectives. Image 'b' shows an alternative setup providing six views with a single camera and a mirror array. The software for processing the images and retrieving label information will be the same for both hardware

7. Quality Management for Digitisation of Herbarium Sheets

The common standards for digitisation of herbarium sheets are derived from the recommendations of the Global Plants Initiative [54, 55, 79], recognised by iDigBio, and supported by 329 partner herbaria worldwide³⁰. Additionally, botanical gardens and herbariums have also produced guidelines for digitisation of herbarium specimens. The following sections provide further details and examples. Because of their essentially 2D nature, quality standards for imaging of paper archives are easily applicable to herbarium sheets.

7.1 Herbarium Sheets Digitisation Standards

The Global Plants Initiative guideline for scanning specimens [54] recommends 600 PPI as the optimal resolution for scanning herbarium sheets. This resolution is based on images (TIFF) produced by equipment available when the GPI released its recommendation. The master TIFF image is not usually published, it is used as the base to produce different image derivatives. The resolution of these images will vary according to the intended use of the derived images (such as web publishing, or printing). For instance, Meise Botanic Garden has established a process which produces a set of three images: a high definition master image for archiving (TIFF 450 PPI³¹), a high definition compressed image for derivation of other images (JPEG2000 at 420 PPI), and a low-resolution image for web publishing (JPEG 72 PPI). Similarly, NHM makes an uncompressed low-resolution image available for quick inspection and download, and a high-resolution image (TIFF 600 PPI) for research or derivation of other images (available by request).

7.1.1 Image Resolution and Colour Standards

Table 15 illustrates the recommended standard criteria for digital specimens of herbarium sheets. The details about the different resolutions and uses are derived from the recommendation from the US Library of Congress [61], Synthesys3 [79], Global Plants Initiative [54,55] and FADGI [31,32]. In QA processes, these criteria can be used to assess that the images produced are fit for purpose. Similarly, in QC processes, these criteria can be used to evaluate if the images delivered are appropriate for their intended use.

Table 15 Example of Image Standard Criteria for Herbarium Sheets

Expected Outcome	Image parameter standards			
	Resolution	Bit Depth	Grey Scale Factors	Colour Accuracy**
Web Publishing	72 PPI	24-bit colour		$\Delta E < 5$
Printing	300 PPI	24-bit colour		$\Delta E < 5$
OCR of Labels	400 PPI	8-bit grayscale	Min: 28 steps Min: 5.5 f-stops Y channel noise $\leq 5\%$	
Specimen features	400 PPI	24-bit colour		$\Delta E < 5$
Research on specimen	420 PPI to 600 PPI*	24-bit colour		$\Delta E < 5$
Preservation	450 PPI to 600 PPI*	24-bit colour		$\Delta E < 5$

* Minimum resolution recommended, if digitisation devices available allow for higher resolution, that resolution should be used.

**Delta E (ΔE , δE , dE) is a metric for understanding how the human eye perceives colour difference.

³⁰ <https://www.kew.org/science/projects/global-plants-initiative-gpi>

³¹ Relative to a A3 standard size



The recommendations for bit depth and colour accuracy are derived from the technical recommendation from the US Library of Congress [61], complemented with observations from recent digitisation projects [19, 28, 103]. Finally, the recommended colour space is Adobe RGB (1998), taken from FADGI [31, 32, 52].

7.1.2 Image Elements

Image elements refer to visual elements which will appear next to the herbarium sheet specimen and which are intended to help in the identification, processing, and quality control. There are five elements recommended by the Global Plants initiative [54]: (1) Colour Chart, (2) Scale Bar, (3) Labels, (4) Barcode, and (5) Institution Name. In the case of Herbarium sheets, scanning all the elements may require more than one pass, since they may be in the form of booklets, paper sheets attached to the herbarium specimen, overlapping labels and labels covered by plant parts. Some collections also have envelopes containing loose material associated with sheets, in some workflows these are left unopened, in others these are opened and emptied onto trays and imaged as well. Figure 53 shows the elements of the herbarium sheet specimen³².

The number of images per specimen can vary depending on several factors, such as the inclusion of additional materials (envelopes, vials, microscope slides), information booklets, folders, and physical catalogues and indexes. All of which may be imaged concurrently and linked to the main specimen.



Figure 53 Example of an herbarium sheet and the required elements to capture. (1) Colour Chart, (2) Scale Bar, (3) Labels, (4) Barcode, and (5) Institution Name (many institutions include this as part of the scale bar, the barcode or a stamp).

The **Colour Chart** is recommended for helping with quality control and post-processing [106], this can help in verifying the lighting, white balance and colour accuracy of the image. FADGI, the US Library of Congress, and Synthesys3, recommend the use of colour chart (referenced as colour target or colour checker) [31, 54, 55, 61, 79]. There are many types of colour charts, and examples of many of them have been used by different institutions in their digitisation pipelines. However, modern

³² <http://www.botanicalcollections.be/specimen/BR0000013305871>

targets such as those from Image Science Associates are preferred over legacy targets (Colour Control Patches from Kodak) because they were developed for digital image creation and are made to very tight tolerances [18, 53]. Object level targets of this type include a ruler and can be used for verification of colour, sharpness and scale.

The **Scale Bar with Herbarium Name (and Logo)** is recommended to enable the calculation of the dimensions of the specimen and to identify the collection holding the specimen [79].

Labels are commonly placed next to the specimen attached to the herbarium sheet. Clear capture of labels is important for further processing and documentation of the specimens such as through OCR techniques [79].

Barcode are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens [79]. There are different types of barcodes available. Line Barcodes (one dimensional), like the one on the ones shown on Figure 53 and Figure 54, have been used in some digitisation projects. However, there can be misidentification problems with them, because of line features in the specimens. Consequently, current guidelines recommend the use of two-dimensional barcodes, and they are being adopted for new projects [118]. Regardless of the type of barcode used, it is important to note that barcodes should always include a human readable identifier as collection curators will need to read them without the help of barcode readers.

The number of images per specimen can vary depending on several factors, such as the inclusion of additional materials (envelopes, vials, microscope slides), information booklets, folders, and physical catalogues and indexes. All of which may be imaged concurrently and linked to the main specimen. For instance, at Royal Botanic Gardens Kew (RBGK) produces more than one image if there is a packet on the sheet which cannot be opened without covering any plant material. In such cases, two images are created: one with the packet closed another with the packet open. Additionally, RBGK would

The number of images per specimen can vary depending on several factors, such as the inclusion of additional materials (envelopes, vials, microscope slides), information booklets, folders, and physical catalogues and indexes. All of which may be imaged concurrently and linked to the main specimen. For instance, at Royal Botanic Gardens Kew (RBGK) produces more than one image if there is a packet on the sheet which cannot be opened without covering any plant material. In such cases, two images are created: one with the packet closed another with the packet open. Additionally, RBGK would



Figure 54 Example of an Herbarium Sheet



Figure 55 1 to 1 view of section highlighted in Figure 54

produce two or more images for overlapping labels or det slips, capsules with writing on the top and contents etc. In these cases, RGBK follows the naming practice recommended by GPI. For instance, if the first image is named by the barcode K000229649, the subsequent images are named using lowercase sub-indexes (K000229649_a for the second K000229649_b for the third, etc.).

7.1.3 Example

The image on Figure 54 shows an example of a digitised herbarium sheet downloaded from the NHM Data Portal [67], a fern specimens collected from Tristan da Cunha, in the south Atlantic³³. This image was selected because of its complexity, showing in a single sheet: three specimens, two scale bars, tree barcodes, colour chart, multiple printed and handwritten labels, and text written directly on the sheet.

The example image is a TIFF file with a 600 PPI resolution, uncompressed. The image uses the Adobe RGB (1998) colour space and 24-bit colour depth. The dimensions of the image are 13.063 x 21.120 cm. These dimensions correspond to a life-size herbarium sheet scanned at 600 PPI. The image on the right is scaled to 1/4X or 25% of its real-life size. The highlighted section is shown below as Figure 55 at 1X scale next to a 1cm square for comparison.

Figure 56 shows a comparison of the specimen in the image fragment in Figure 55 at five resolutions (72, 96, 300, 420, and 600 PPI). The derived images keep the dimensions so that when scaled to 1:1 view, 1 inch would contain the corresponding number of pixels.

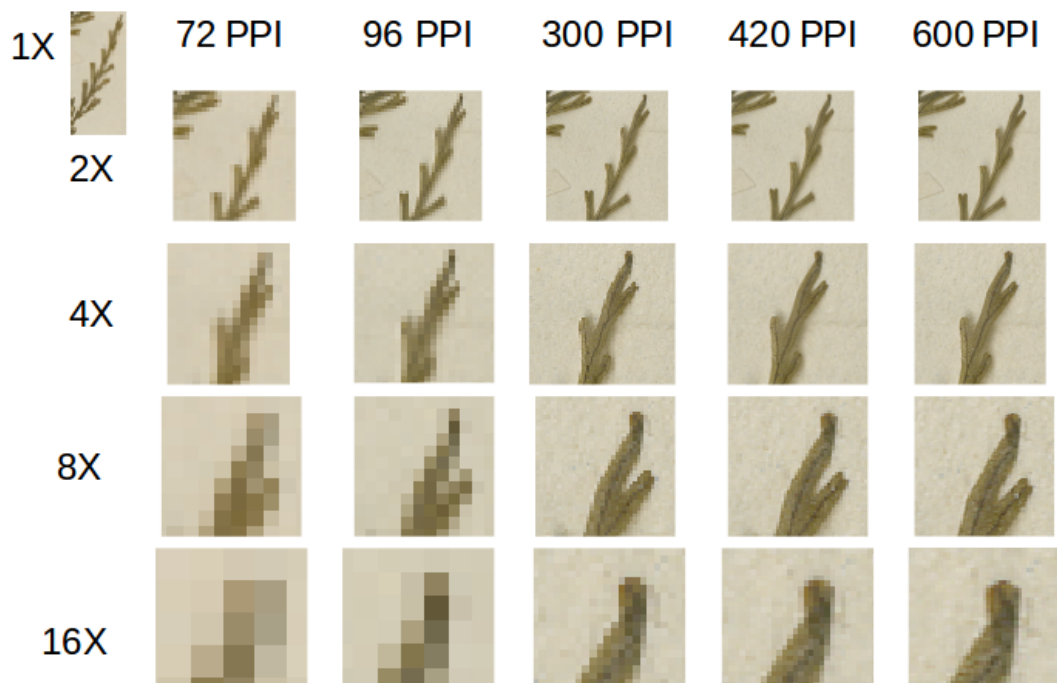


Figure 56 Resolution Comparison on sample image fragment (Figure 55). 72 and 96 PPI resolutions are recommended for publishing online. 300 PPI resolution corresponds to the recommendations for scanning for printing. 420 PPI allows the identification of specimen features (400 PPI is the minimum recommended for OCR). 600 PPI is the minimum recommended for research quality and long-term preservation (see Table 15).

³³ <http://data.nhm.ac.uk/object/32f232e0-9d55-424c-8027-ee28a433b58e>

The first sample on the top left at 1x. The remaining samples are scaled at 2, 4, 8, and 16 times their original size. Four images with 72, 96, 300, and 420 resolutions were derived from the original. 72 and 96 PPI resolutions are commonly recommended for publishing online. These correspond to the display resolutions in Macintosh and Windows systems. 300 PPI resolution corresponds to the recommendations for scanning of photos for printing at the same size. 420 PPI resolution corresponds to the format used by some herbariums for storing their digitised collections. Finally, 600 PPI corresponds to the recommended minimum resolution from the Global Plants Initiative [54, 55, 79].

7.2 Herbarium Sheets Digitisation Workflows

According to Synthesys3, the digitisation workflows of herbarium collections of different institutions are the ones with the most similarity in terms of the steps carried out and the order in which these steps were performed. This likeness between workflows is explained by the similarity in equipment used to digitise the specimens, by the layout of the collections, and by the fact that large digitisation projects of herbarium specimens, such as the Global Plants Initiative have been carried out, setting up and standardizing many aspects of the digitisation workflow [50, 79].

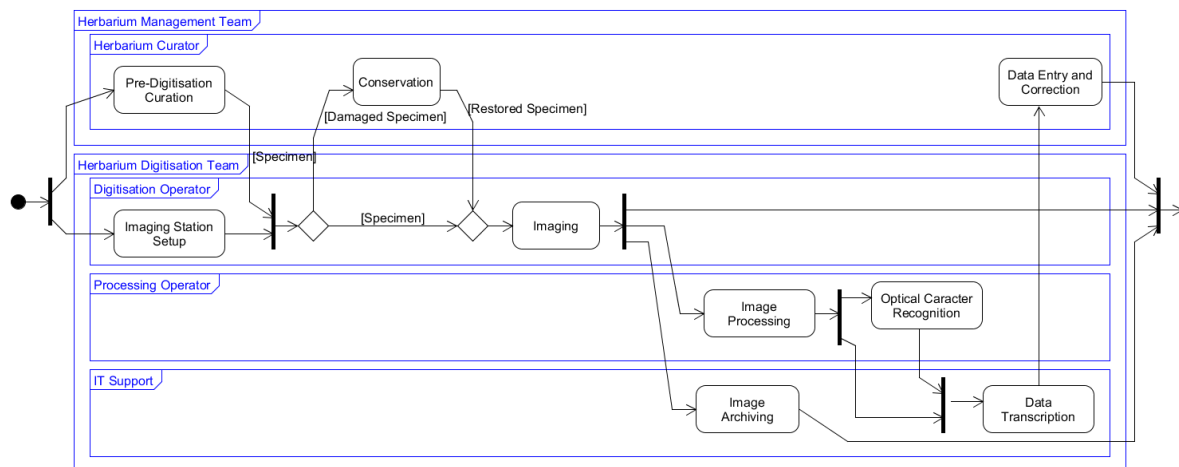


Figure 57 A generic herbarium digitisation workflow

Figure 57 shows a high level view of an example of a herbarium digitisation workflow, built using the digitisation activities recommendations found in iDigBio and Synthesys3 [50, 71, 79]. The diagram includes the activities which can occur in a digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, herbarium collections which have already captured data may not need to include the 'data transcription' and 'data entry and correction' subprocesses.

Table 16 provides a brief description of the workflow tasks and their influence in the production of images. The main variations on the integration of digitisation workflows depend on the metadata required and the type of digitisation stations to be used. The following sections describe how the activities for herbarium digitisation are affected by these two features [71, 79].

Table 16 Description of workflow tasks, derived from iDigBio recommendation [50]

Workflow Task	Description/Subtasks	Influence on Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); conservation if needed (see conservation task below), specifying safeguards for identifying and handling specimens treated with hazardous materials (e.g. mercury); transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators.
Imaging Station Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians.	Equipment should be calibrated to minimise image postprocessing after digitisation.
Conservation	Required conservation/restoration of specimens selected for digitisation. This can be done during pre-digitisation curation as part of the preparation of the collection.	Some sheets may be damaged or fragile, or specimens may need to be repositioned to display all relevant features.
Imaging	Data capture; identification (barcoding); placement for imaging; digitisation of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta] data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of derivatives for publishing and distribution; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
Image Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data Transcription	Manual extraction of data from images, for populating/complementing specimen record.	Verify readability of derived image data transcription.
Data entry and correction	Final capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing.

7.2.1 Metadata Required for Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record on the collection management system. Additionally, specific data which is contained with the specimen such as scientific name, field notes, collector, collection event (date, location, campaign/funding), and herbarium collection (there are different collection hierarchies and schemes) should also be included. In herbarium digitisation this is normally contained as labels which are placed on the herbarium specimen sheet. It is also relevant to mention that often important information e.g. the taxon name under which the specimen is stored in the Herbarium might not be found on the specimen label itself but on the covers and folders the specimens are contained in. Therefore, this information must be manually captured too, or the folders imaged and later transcribed and linked to the specimen. This is relevant for imaging because the image should capture the relevant metadata clearly for its later transcription and verification. The extraction of



data from the digitised specimens can entail automated or manual activities which can take place before or after digitisation. Alternatively, some institutions may already have most of the metadata stored in a collection management system, in which case the most important information is the unique identifier data that links the physical specimen and the digital specimen, which for herbarium sheets consists of a barcode. Additionally, it is a convention to use the barcode to name the imaged specimen files. When more than one image per specimen is required, herbarium holding institutions follow the recommendations of GPI on naming which indicates using sub-indexes for additional images.

7.2.2 Digitisation equipment

Digitisation stations can be classified as camera or scanner based. Additionally, the specimen positioning and imaging can be manual or semiautomated. There are no examples of fully automated digitisation stations [71, 79, 98]. Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. Semiautomated stations have various forms which can vary in the amount of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen under the digitisation equipment. The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. Currently, the fastest digitisation rates correspond to conveyor belt and camera-based stations [80, 98] that can be 5 to 10 times faster than manual digitisation station.

7.3 Herbarium Sheets Quality Management Methods

In line with the imaging standards for herbarium sheets (7.1), the quality management methods for herbarium sheets focus on the quality of the digitised specimen and the presence of the recommended image elements, and that the specimen parts are not covered by those elements. The quality management should verify the presence and clarity of image elements (spatial resolution), verification of colours (spectral resolution), and verification of the perception of brightness in the image through the number of the grey value levels (radiometric resolution). The lines on the scale and barcode elements can be used as reference to verify the spatial resolution of the image. The colour chart can support the verification of spectral and radiometric resolution.

Quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are verified should be those that follow the initial digitisation of the specimen and after the production of each derivate from it.

7.4 Examples

In several herbarium digitisation projects, part of the work has been outsourced to specialised companies, while other projects have followed the traditional inhouse setup. In inhouse digitisation, the institution hosting the collection procures and operates the digitisation equipment, designs the digitisation process, performs the digitisation, and manages the digitised collection. In outsourced digitisation, the institution hosting the collection contracts the services of different providers for parts of the digitisation workflow. The following subsections illustrate these variants highlighting the corresponding quality management activities and their place in the workflows.



7.4.1 Inhouse Digitisation

The digitisation workflow of the Royal Botanic Garden Edinburgh (RBGE) is an example of an inhouse digitisation workflow. The main reasons for designing and developing this workflow inhouse was the variation of funding and scale of digitisation campaigns [41]. For this, a custom and modular system which could be adapted to cater for the needs of different digitisation projects was required. The RBGE designed three concurrent workflows which responded to these requirements. Figure 58 shows the specimen, data and image workflows side by side. Notice that the image quality control activities are part of the image workflow and occur after digitisation.

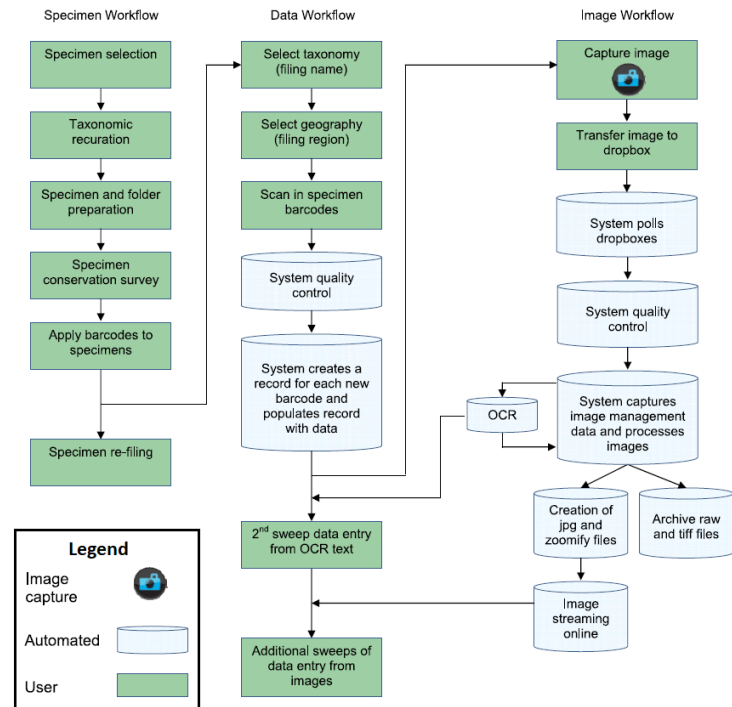


Figure 58 Diagrammatic overview of the digitisation workflows at the Royal Botanic Garden Edinburgh (from [41])

The workflows developed by RBGE are suitable for the characteristics of the different funding streams and the diversity of digitisation projects. The design of modular and integrated workflows to manage the processes, images and data has served this purpose. This approach is different to the mass digitisation projects such as the one by Muséum national d'histoire naturelle (MNHN) in Paris which aims to digitise all specimens within a single project.

7.4.2 Outsourced Digitisation

Picturae is a digitisation service provider for several museums and collections with an experience of more than ten years. The digitisation workflows developed by Picturae (Figure 59), illustrates how the image quality management activities are integrated [86]. The workflow is designed to perform image quality assessment at digitisation time, stopping and rewinding the conveyor belt if an error is detected. The means for verifying the quality built into the workflow allow the rapid verification of every image. This workflow has been applied and tested in large digitisation projects for institutions such as Naturalis (Netherlands), the Muséum national d'histoire naturelle (France), and Meise Botanic Garden (Belgium).



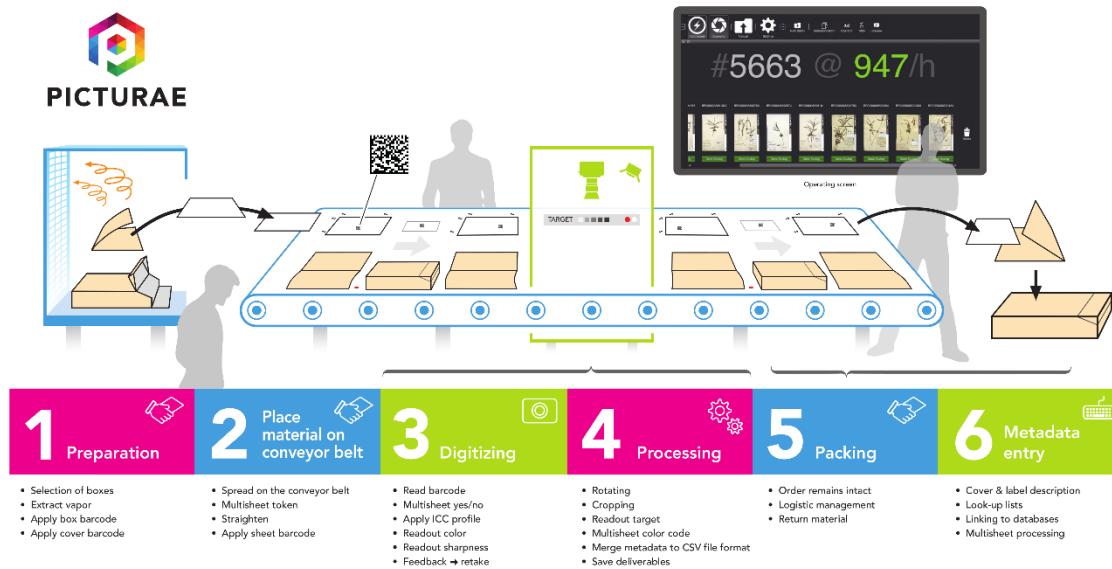


Figure 59 The diagram shows the Picturae digitisation workflow for herbarium collections (Courtesy of Picturae³⁴). The digitisation task (3 Digitizing) includes image quality assurance checks colour and sharpness. Errors in any of these can prompt the need to stop and re-image specimens if an error is detected.

The outsourced approach has demonstrated its effectiveness in mass digitisation projects, there are alternative examples of such projects involving other providers such as the digitisation of the Moscow University Herbarium [89] or the Natural History Museum of Norway [20]

³⁴ Updated version of the workflow included in [80]

8. Quality Management for 3D Digitisation of Natural History Collections

Most 3D digitisation stations are custom-made, and camera based. Specimen positioning and imaging is mostly a manual process, however there are examples of automated digitisation stations [25, 26, 36, 58]. Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. Digitisation in the case of 3D imaging can mean manual operation of equipment. There are some components, such as turntables, camera arms and focus stacking rails from various vendors, which can produce semiautomated stations with different levels of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen under the digitisation equipment. The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. The variety of workflows and equipment covered by the Synthesys3 report shows considerable variations in the time per specimen, in range of 10 minutes to more than six hours specimens per year, as well as the costs per specimen in the range from 1.00 € to 500.00 € per specimen (at 2016 prices [12, 13]). This is because some workflows included costly equipment and software (microscopes and micro CT). The most flexible options for the widest variety of specimens were those based on photogrammetry and structured light scanning.

Three of the main digitisation workflows covered by the ICEDIG project [46]: Skins and Vertebrate Material (section 4), Liquid Preserved Specimens (section 5), and Pinned Insects (section 6) contain specimens which can be digitised using 3D digitisation methods. Additionally, collections and specimens not covered by the workflows described in sections 3 to 7 are also targets of 3D digitisation, for instance some herbarium specimens (live plants, fruits and seeds), molluscs, fossils (human, zoological, and botanical), and minerals (meteorites, ores, sediments, crystals).

The variety of collections and preservation media means that it is not possible to provide a general standardised method such as those developed for collections such as microscopy slides (section 3.1) or herbarium sheets (section 7.1). Consequently, the guidelines presented here are based on existing literature and examples from digitisation projects that have digitised parts of their collections using 3D digitisation methods.

8.1 3D Digitisation Standards

Natural history collections contain specimens which are real life 3D objects, and therefore limiting digitisation to 2D would result in a loss of potentially valuable information [10]. The diversity in the types of collections and specimens in natural history museum collections prevents the use of a single 3D digitisation technique and, as a result, a range of 3D digitisation techniques have been tried for different types of specimens. However, many technologies are still expensive or under development. The need for 3D digitisation has spurred the development of different custom solutions for capturing and publishing 3D specimen models online. The current consensus is to perform 3D digitisation at limited scales for specimens identified as having specific characteristics (e.g. scientific value, preservation state, uniqueness) that justify the cost of 3D digitisation. The quality criteria for 3D digitisation aligns with the intended use of the digitised specimens. The greatest potential of 3D digitisation is to support scientific research, especially when the features of the specimen that need to be illustrated cannot be faithfully reproduced in 2D images [9].



The Natural History Museum London (NHM), Naturalis Biodiversity Centre (Naturalis), and the Berlin Natural History Museum (Berlin Museum für Naturkunde Berlin MfN) have published different types of 3D views of some specimens from their collections [38, 58, 68, 69, 120]. Figure 60, Figure 61, and Figure 62 show examples of 3D digital specimens: a 3D model (textured wireframe), 360° movie (2D+) and 360° view (from 2D perspectives). The images in Figure 60 are screenshots of a 3D model of a 25.5-meter long Blue Whale skeleton on display at NHM. The published model is in blender format and consists of 247,170 triangles and 123,633 vertices³⁵. The smallest part of the model is the mid-section of the skull (Figure 60b) which consists of 3,936 triangles and 1,970 vertices.

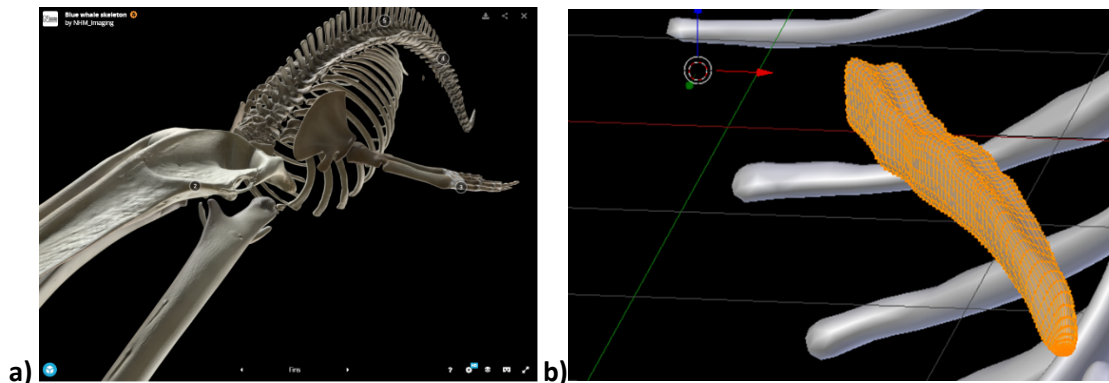


Figure 60 Views of 3D Model of a Blue Whale Published by the Natural History Museum³⁶. Image 'a' shows the left view of full skeleton. Image 'b' shows the mid-section of the skull highlighted as a mesh (the smallest independent component of the model) which consists of 3,936 triangles and 1,970 vertices³⁷

The images in Figure 61 are part of a 360° video of a pigeon (MPEG4). The video is composed of 35 frames, at a 10 frame per second rate (3.5 seconds playtime). Each frame has a width of 464 pixels and a height of 728 pixels. The nominal resolution of each frame is 300 DPI. Using this nominal resolution of 300 DPI as a reference, the size of the actual specimen is calculated as 2.55cm width by 3.40 cm height (calculated from Figure 61b). In this case, the 360° video is optimised for online viewing as published.

³⁵ NHM model available at: <https://sketchfab.com/models/8502dbef80ed4aa688c13c90cb14de73>

³⁶ Produced by [68]. url: <http://www.nhm.ac.uk/discover/blue-whale-skeleton-3d.html>,
<https://sketchfab.com/models/8502dbef80ed4aa688c13c90cb14de73>

³⁷ NHM model available at: <https://sketchfab.com/models/8502dbef80ed4aa688c13c90cb14de73>

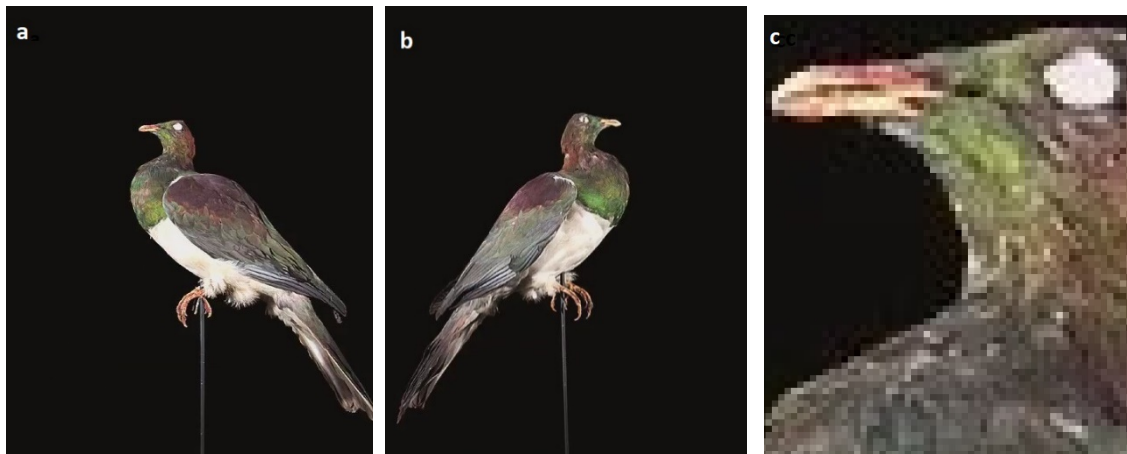


Figure 61 Images 'a' and 'b' show two frames of the 360° video of the Extinct Norfolk Pigeon (or Norfolk Island Pigeon – *Hemiphaga novaeseelandiae spadicea*), Published by Naturalis³⁸. Image 'c' shows a 8X magnification of the specimen's head (from 'a').

The images shown in Figure 62a to Figure 62i show part of the images from a 360° sequence of a beetle, consisting of 120 focus stacked images. The images are (nominally) stored at 300 PPI, however, using the scale bar on ZooSphere website as a reference, the resolution of the image is estimated to be 3,856PPI, or 152 pixels per mm. At this pixel resolution the actual magnification is close to 40X. The close-up image shown as Figure 62j is a 3mm section with a magnification of 20X (half the maximum possible from of the original image).

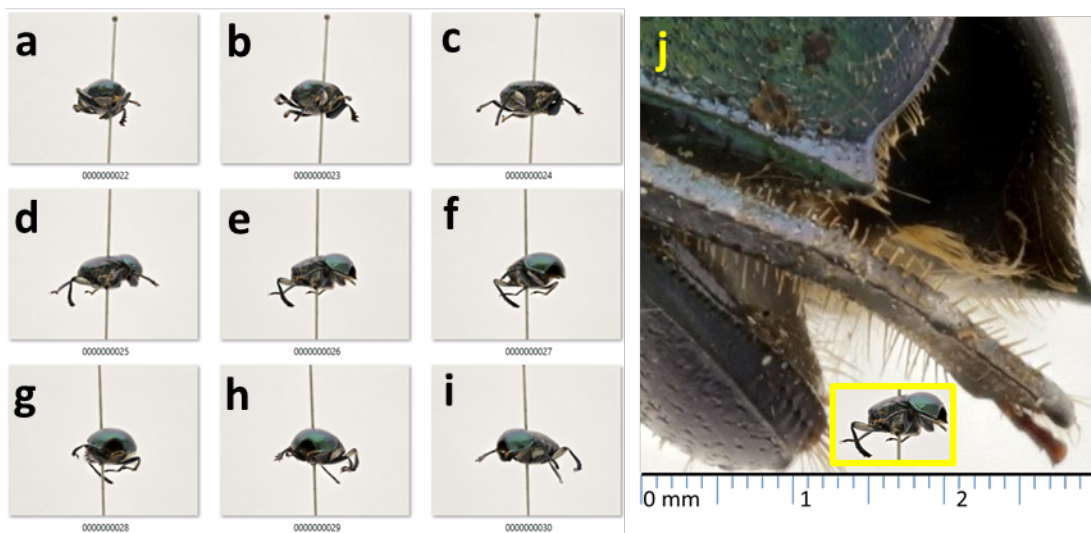


Figure 62 Images 'a' to 'i' show part of the 120 image sequence of a 360° Specimen View of of a Beetle (*Gymnopleurus virens*), Published by Berlin MfN³⁹. Image 'j' is a 20 X magnification of a 3mm section from 'e' on the right. The inset image above the scale bar shows the specimen as it would look at 1 to 1 size.

The remainder of this section focuses on the quality management for 3D digitisation methods. The examples included in Figure 61, and Figure 62 are examples 2D digitisation (movies as the projection of 3D models in a plane). For these type of images the relevant quality criteria are covered in sections 3, 4, 5, 6, and 7. The workflows that produce the 360° video (Figure 61) and 360° view (Figure 62)

³⁸ Naturalis BioPortal [67], URL: <http://data.biodiversitydata.nl/naturalis/specimen/RMNH.AVES.110112>

³⁹ ZooSphere MfN [120], URL: <http://www.zoosphere.net/sequence/143/Gymnopleurus/virens>

can be reused and integrated as part of a 3D digitisation workflow and are further used as examples in Section 8.2.

8.1.1 3D Model Quality Management Standards

In a similar way as for 2D images, the quality criteria for 3D models can be grouped according to the intended uses of those models. For instance, the criteria for a model to be published and inspected on the web will be different from the criteria for a model that is intended for 3D printing or research. However, where in the case of 2D the image is generally taken at the quality level needed by research and then downscaled for web viewing if necessary, 3D imaging is often done for public display purposes. This means that 3D imaging workflows and the quality that is needed vary greatly and need to be considered at the start. Further, the vast amount of data that is captured for a mid-quality 3D model (let alone research quality) requires down sampling of the data, especially for web (and AR and VR) viewing.

Different methods for assessing the quality of 3D models have been proposed [1, 2, 3, 4, 5, 16, 22, 27, 30, 33, 40, 59, 60, 65, 75, 76, 117]. These 3D model quality management methods have been developed to assess the amount of distortion after some processing (simplification, watermarking, transmission, among others). Accordingly, these methods can be classified according to the availability of a reference 3D model as: full-reference (the reference 3D model content is fully available), reduced-reference (part of the reference 3D model is available), and no-reference (no information available on the reference 3D model) [1, 2, 3, 4, 5, 27]. Full-reference and reduced-reference methods are interesting for validation of derived models, which result from watermarking, simplification, scaling, compression, fragmentation, integration, digital reconstruction and noise reduction [1, 5, 16, 22, 27, 30, 33, 40, 59, 60, 65, 75, 117]. No-reference methods are useful for assessing models for which the reference 3D model does not exist or it is not available [2, 3, 4, 76]. These methods could be useful for assessing new models as they are produced. The methods reviewed up to this point are mostly based on verifying the geometric structure of the 3D model, i.e. they assess the quality of the mesh [1, 2, 3, 4, 5, 27, 33, 59, 60, 75, 76, 117], point-cloud [22, 30] or surface (polygons) [16, 65] that form the 3D model, with little or no attention to the colour (or textures) of the model (except for [40]).

As shown above, several institutions have started to create and publish 3D models from their collections [38, 58, 68, 69, 120]. However, due to their exploratory nature, 3D digitisation has not required or produced standards for 3D digitisation comparable to those for 2D imaging. As an alternative, different guidelines and recommendations generated by industry can be adopted/adapted to suit the needs of the Natural History collections' digitisation [15, 45, 99]. For instance, CheckMate provides guidelines for assessing models to be published and used in gaming and display [15], StemCell provides recommendations on the characteristics to achieve model portability among different formats [99], and i-materialise provides recommendations for preparing models for printing with different formats [45].

8.1.2 Model Elements

Model elements refer to physical elements which are captured next to the specimen and which are intended to help in its identification, processing, and quality control. For research-quality 3D models, a list of six elements was derived from the recommendations: (1) Barcode, (2) Labels, (3) Colour Chart, (4) Scale, (5) Specimen and (6) Background [42]. The list has been verified by looking at



specimen examples from Naturalis BioPortal, the NHM London, MNHN Paris and the Smithsonian Institution [69]. The elements are not mandatory and can vary depending on the specimen being digitised.

The **Colour Chart** and **Scale Bar** are recommended for helping with quality control and post-processing [106].

Colour charts can help in verifying the lighting, white balance and colour accuracy of the image [10, 31, 54, 55, 61, 79]. For some specimens, colour information needs to be captured to provide the texture for rendering the model. Texture files are normally 2D images linked to different sets of polygons in the final model. Storing and linking colour charts to the texture sets can help in producing more realistic renderings.

Scale Bars are recommended to enable the calculation of the dimensions of the specimen and to identify the collection holding the specimen [10, 79]. The reference that the scale bars provide can support the building of the model.

There are different types of colour charts, and many of them have been used by different institutions in their digitisation pipelines. However, modern targets such as those from Image Science Associates are preferred over legacy targets (Colour Control Patches from Kodak) because they were developed for digital image creation and are made to very tight tolerances [18, 53]. Object level targets of this type include a ruler and can be used for verification of colour, sharpness and scale.

Labels are commonly placed next to the specimen. Clear capture of labels is important for further processing and documentation of the specimens [10, 79].

Barcode are identifiers used for cataloguing specimens which are also useful for linking them to digital specimens. Synthesys3 recommends the use of barcodes as internal identifiers which are important for further documentation and linking of the physical and digital specimens [10, 79]. There are different types of barcodes available. Line Barcodes (one dimensional), like the one on Figure 21, have been used in some digitisation projects, however, there were some misidentification problems (because of line features in the specimens). Consequently, the use of two-dimensional barcodes is recommended, and they are being adopted for new projects [118].

Specimen: 3D objects requires the acquisition of image sets which capture the specimen details from different perspectives. Relevant anatomical traits which need to be exposed need to be considered when defining the capture angles.

For 3D models of specimens, it is not necessary for all images to include all the elements. However, background, scale, and colour chart should be selected and included in images along with the specimen, to enable colour postprocessing and estimation of scale. Other elements such as the specimen labels and barcodes can be included in a separate photo [10, 70] and linked to the specimen through metadata.

8.2 3D Digitisation Workflows

The Synthesys3 project documented 12 different digitisation workflows from pilots performed at four European NHC intuitions (HCMR, MfN, RBINS, and RMCA) [12, 13, 79]. 3D digitisation workflows are slower and more complex than other digitisation workflows [79]. 3D digitisation workflows



throughput is affected directly by the 3D data acquisition process, the varied type of specimens, the number of collections which can contain materials suitable for 3D modelling (herbarium, mammalogy, ornithology, herpetology, to name a few), and the level of documentation of the collection. Additionally, some of the 3D digitisation workflows hardware and software are developed and integrated in-house, with resources available at each institution (especially in the case of photogrammetry) [12, 13, 50, 70, 79].

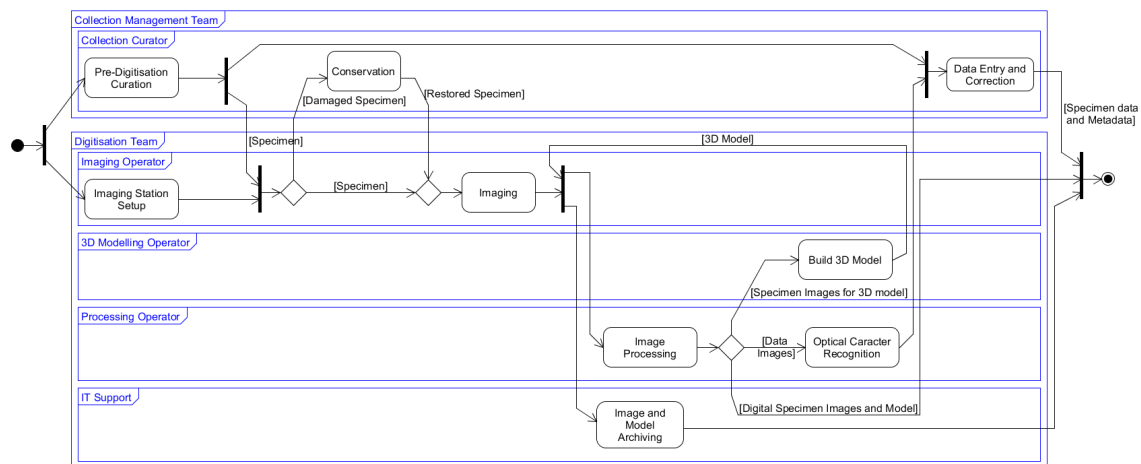


Figure 63 Generic 3D digitisation workflow

Figure 25 shows an example of a generic 3D digitisation workflow, derived from the digitisation activities recommendations found in iDigBio, and Synthesys3 [12, 13, 50, 70, 79]. Most of the tasks in the diagram coincide with those defined by iDigBio [50, 70], while the sequencing and alternative paths are derived from the Synthesys3 examples [12, 13, 79]. The diagram includes the activities which can occur in a 3D digitisation workflow, however, as discussed previously (see section 2.2), the order of activities can vary, and some activities can be excluded. For instance, the current efforts of MfN perform only “imaging” and “processing” (acquiring and composing stacked images) without building the 3D models from them [58]. Similarly, it is likely that data capture has already been done in a 2D digitisation workflow, so that the “data entry and correction” task can be omitted [79]. The number of images per specimen also influences the shape of the workflow instances. Table 17 provides a brief description of the workflow tasks and their influence in the production of images.

Table 17 Description of 3D workflow tasks, derived from iDigBio recommendation [50].

Workflow Task	Description/Subtasks	Influence Model/Image Quality
Pre-digitisation Curation	Selection of specimens to digitise; retrieval from storage; identification of specimens (barcoding); identification of Specimens requiring high-definition digitisation; defining safeguards for identifying and handling specimens treated with hazardous materials (e.g. mercury); transfer to digitisation station; and creation of skeletal metadata record.	Specimens should be selected and prioritised for digitisation by the collection curators. .
Imaging Station Setup	Digitisation equipment selection, acquisition, and set up; equipment testing/calibration; and training of digitisation technicians.	Equipment should be calibrated to minimise image post-processing.
Conservation	Required conservation/restoration of specimens selected for digitisation.	Some specimens may be damaged, fragile, or require to be re-mounted to display all relevant traits.

Table 17 Description of 3D workflow tasks, derived from iDigBio recommendation [50].

Workflow Task	Description/Subtasks	Influence Model/Image Quality
Imaging	Data capture; identification (barcoding); mounting for imaging (may require different mounts for different poses, darkening); imaging of specimen; creation of master file (raw or DNG); temporary storage of master file; and unmounting and return of specimen.	Identification, digitisation and [meta] data capture, so that images are correctly linked to the corresponding specimen records.
Image Processing (photogrammetry)	Creation of master files for archiving (raw, DNG, TIFF, JPEG2000); Image adjustment (using colour charts); creation of stacked images; creation of derivatives for OCR and data transcription; verification of naming and linking of files to digital specimen identifier (can be based or derived from barcode ID).	Verification of master image resolution and verification that derivatives adhere to quality standards.
Build 3D Model (photogrammetry)	Use of specimen images in the construction of the 3D Model.	3D model detail will vary according to the type of images and building process used. Images can be used for creating textures for the final model.
Build 3D Model (photogrammetry)	Other methods may not produce intermediate images to be processed, instead delivering complex models in proprietary formats.	The acquisition method influences the quality of the models produced
Image and Model Archiving	Transfer of master and derived files to archive servers, image servers, and public servers in preparation for publishing.	Verify that master and derived files are not corrupted in transfer to storage.
Optical Character Recognition	Automated extraction of data from images, for populating/complementing specimen record.	Verify suitability of derived image for OCR.
Data entry and correction	Manual extraction of data from images, for populating/complementing specimen record. Data capture and correction of digital specimen record before publishing.	Verification against reference image and recorded data before publishing.

The main variations on the characteristics of digitisation workflows depend on data and metadata required (or already present) as well as on the type of digitisation stations to be used.

8.2.1 Metadata Required for 3D Digitisation

The minimum metadata required for imaged specimens is the metadata for identifying the specimen and linking it to a record on the collection management system. Additionally, specific data which is contained with the specimen such as scientific name, field notes and collection name should also be included. In skin and vertebrate materials digitisation, this is normally contained as labels which are attached to the specimen or to its containing/storage medium. This is relevant for imaging because the image should capture the relevant metadata clearly for its later transcription and verification. As described in the section on model elements (8.1.2), this can be done in a separate image. The extraction of data from these images can entail automated or manual activities which take place before or after the 3D model is built. Alternatively, some institutions may already have portions of the metadata stored in a collection management system, in which case the most important information is the identification data. There is no naming convention, for 3D models, however, many collections use the barcode to name the 3D specimen files (for instance [58]).

Acquisition metadata for various 3D scanning techniques need to be recorded during acquisition. These settings are not saved (such as spacing for laser scanners and quality settings for some types of structured light scanners). The difficulty is that multiple settings may be used for a single object capture. Even if this capture metadata is stored in the raw data, this is not transferred to the output (mesh). This is not a problem for natural history to solve: there needs to be a wider discussion about this with manufacturers, software makers, industry and cultural heritage users.



Nevertheless, to assess the quality and origin of the models, in addition to the specimen and collection data it is important to also store metadata about capture and processing. Capture metadata should include the equipment and software used. Processing metadata should include the software and detailed processing operations performed (such as topology correction, remeshing, extent of hole filling, detection and deletion of duplicate or overlapping points, etc.). This is a current research topic for which many alternatives are being proposed. At the very least creating a provenance chain would help in assessing the level of processing required for producing the different versions of a model. This would promote higher use of 3D models, as researchers are aware of the processes involved and their relation to the model accuracy.

8.2.2 Digitisation equipment

As stated from the beginning, the majority of 3D digitisation stations are custom-made and camera-based in which specimen positioning and imaging is mostly a manual process (with some recent examples of automated digitisation stations [25, 26, 36, 58]). Manual stations have dedicated digitisation technicians who position, digitise and return each specimen. Digitisation in the case of 3D imaging can mean manual operation of equipment. There are some components, such as turntables, camera arms and focus stacking rails from various vendors, which can produce semiautomated stations with different levels of support for the digitisation technician. Some will refocus and adjust the camera or scanner; others will position the specimen under the digitisation equipment. The speeds of these digitisation lines and their costs are the main factors to consider when selecting digitisation equipment. In the case of 3D, following FAIR principles becomes more challenging because of the abundance of proprietary software and formats (e.g. laser and structured light scanners). Additional considerations include the possibility of building scripts for processing, and flexibility of equipment (scale and colour of objects, upgradeable components such as lenses, ease of use, cost, transportability). The variety of workflows and equipment covered by the Synthesys3 report shows considerable variations in the time in a range from 10 minutes to more than six hours per specimen. Similarly the costs per specimen in the range from 1.00 € to 500.00 € per specimen (based on equipment and prices from 2016 [12, 13]). This is because some workflows included costly equipment and software (microscopes and micro CT). The most flexible options for the widest variety of specimens were those based on photogrammetry and structured light scanning.

8.3 3D Digitisation Quality Management Methods

In line with the 3D digitisation standards (8.1), the quality management methods focus on the quality of the model and the presence of the recommended elements. The current literature on the Quality Management of 3D models is more focused on the model structure (point cloud, mesh, polygons) than on the colours and illumination of the specimen. Nevertheless, 2D quality management methods can be applied to the source 2D images to ensure that the derived model textures are accurate. More importantly, in the case of photogrammetry, the 2D quality management methods are needed to ensure the accuracy of the photogrammetric 3D model.

Quality management verifications should be performed periodically to minimise rework. The points of the workflow when they are realised should be those that follow the initial digitisation of the specimen and after the production of each derivative from it (including the 3D model).



8.4 Examples

The three examples selected to illustrate the implementation of 3D digitisation methods include ZooSphere [12, 13, 88, 120], CSIRO-ALA [71, 72] and Smithsonian 3D Program [9, 92, 93]. ZooSphere and CSIRO-ALA are included because they target some of the most difficult collections to handle due to their volume (number of specimens) and the special characteristics of those specimens (fragility and size). The Smithsonian example is included because of its strategic approach to produce a framework which can be adapted to any acquisition technique. Additionally, ZooSphere is the only documented ongoing 3D scanning programme for a natural history collection.

8.4.1 ZooSphere Workflow

The ZooSphere project is an ongoing digitisation program targeting entomology collections hosted at the MfN Berlin. The project includes a strategy for specimen selection, development of a custom digitisation station, development of digitisation software, selection of processing software, and publishing of imaged specimens [120]. Figure 64 presents the ZooSphere workflow and digitisation station.

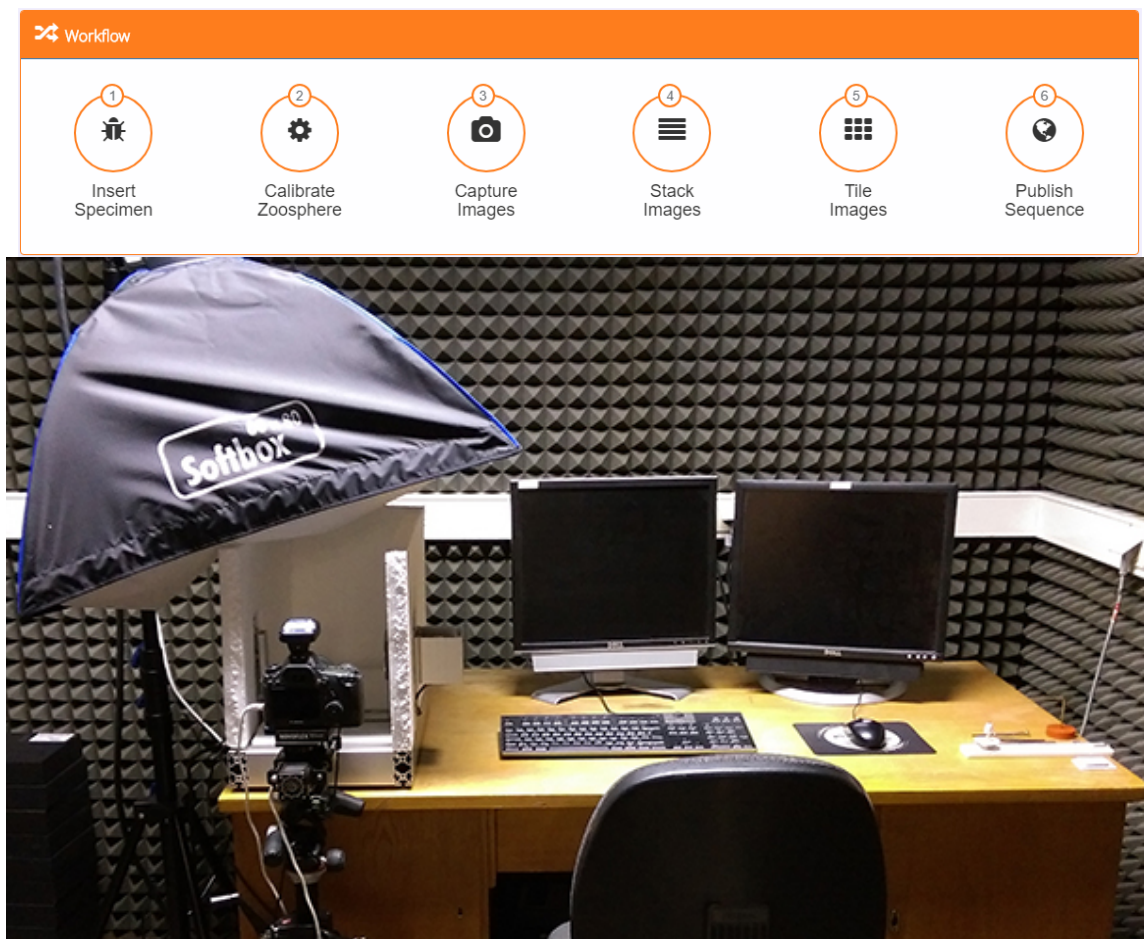


Figure 64 Schematic overview of the ZooSphere Workflow (top) and the custom-built digitisation station (bottom) [120]

The ZooSphere approach presents a viable solution for digitisation of large collections. In this case, instead of aiming to digitise the entire collection, the digitisation prioritises type specimens and digitisation on demand. The digitisation of type specimens is programmed and prioritised by the collection curators, they are responsible for selecting and retrieving the specimens to be processed, as well as for the validation of the metadata for published specimens. For digitisation on demand,

the digitisation team and the curators may select specimens from the collections in the museum which have been requested by scientists in other institutions. Additionally, the strategic 3D digitisation of type pinned insect specimens is not occurring in isolation, but as part of a wider strategy which also includes drawer level and individual level fast digitisation of the collections.

Although the current focus is on the acquisition of 2D image sequences for each specimen, the quality of the images makes possible the construction of 3D models from those sequences (as shown in Figure 65). Another advantage of this approach is that, since the sequences are preserved as stable sources, multiple 3D model building techniques can be applied and compared.

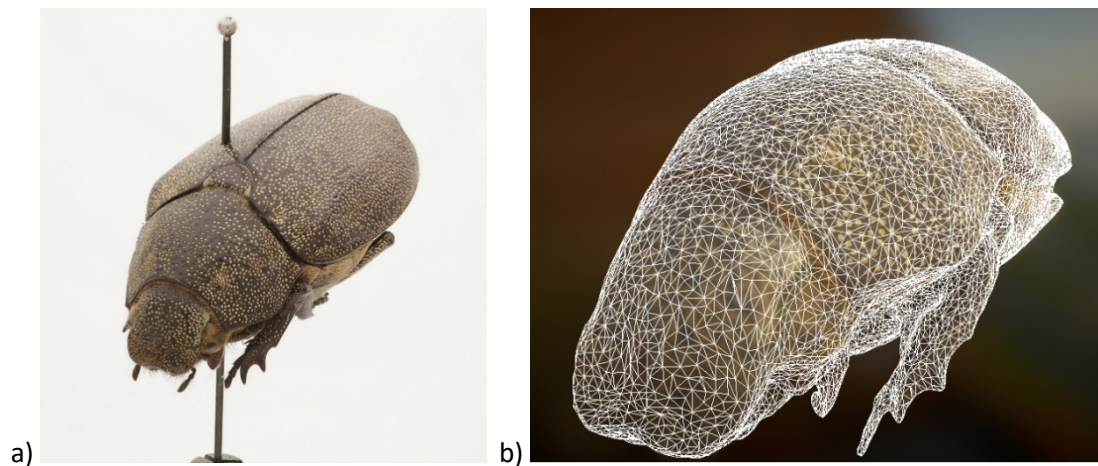


Figure 65 The image shows: (a) one of the photos in a sequence published by ZooSphere and (b) a 3D model created using images from the same sequence⁴⁰

8.4.2 CSIRO-ALA Workflow

The CSIRO-ALA project proposed a full 3D digitisation workflow. The project included a strategy for specimen selection, development of a custom digitisation station, development of digitisation software, selection of processing software, and publishing of imaged specimens [72, 73]. Figure 66 shows schematic diagrams of the configuration of the digitisation station and Figure 67 shows the three main stages of the workflow for 3D digitisation of pinned insects.

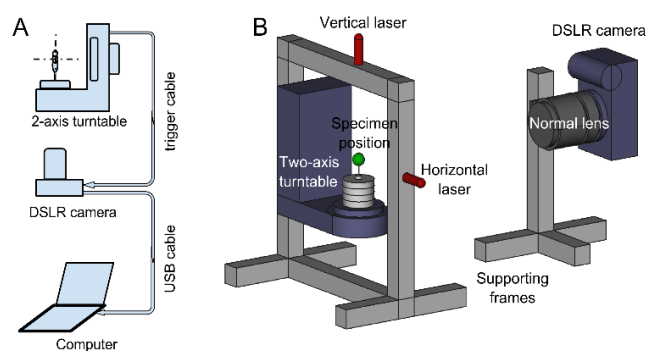


Figure 66 Diagrams of the configuration of the CSIRO-ALA digitisation station for normal-mode image acquisition (no macro photography). Diagram A: system connections. Diagram B relative positions of the camera and the turntable. The green sphere in B marks the centre of rotation and mounting location of specimens. The turntable triggers the camera after reaching predetermined pan and tilt angles [72].

⁴⁰ ZooSphere MfN [93], <http://www.zoosphere.net/sequence/76/Leucopholis/irrorata>

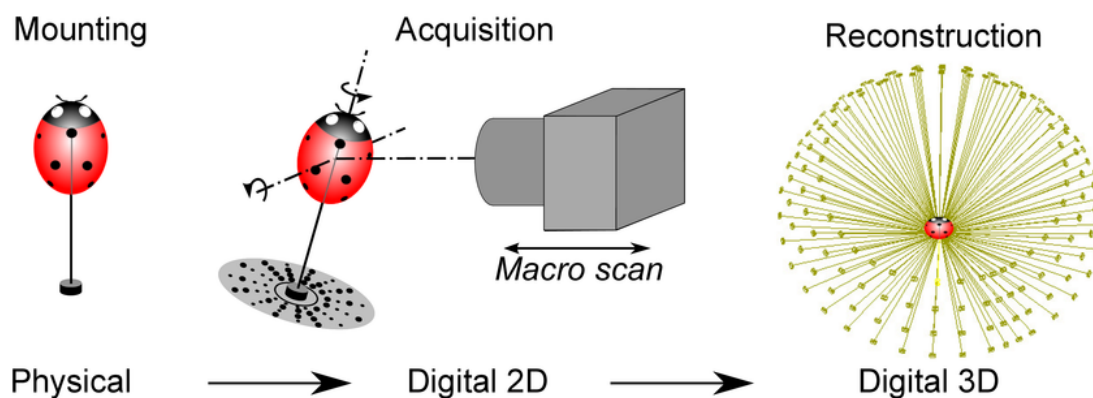


Figure 67 Main stages of the 3D digitisation of insects proposed by CSIRO-ALA Mounting, Acquisition and Reconstruction. Mounting includes selection of the appropriate calibration target, mounting platform, and mounting the specimen. Acquisition includes the acquisition of 2D images (which can contain macrophotography and photo-stacking if required). Reconstruction includes the actual building of the 3D models (point-cloud and texture files) [72].

The CSIRO-ALA approach presents a flexible solution for 3D digitisation of collections containing specimens of variant sizes, by including an alternative for using macrophotography and photo-stacking if needed. In addition to the development of the system, the research group has also studied in detail the challenges and pending issues for achieving higher throughput, decreasing costs and improving precision of the models [73].

The results from this proposal have been published as a demonstrator⁴¹, however there is no specific plan to incorporate this into the current digitisation projects of Australian institution as part of the wider ALA strategy.

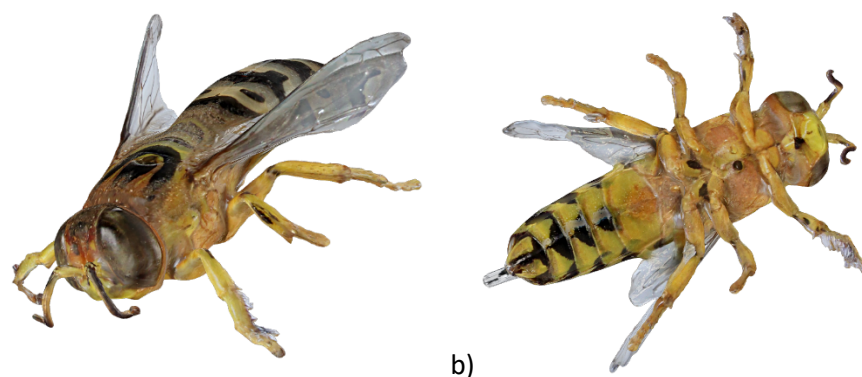


Figure 68 The image shows two views of a bee, published to demonstrate the CSIRO-ALA proposal⁴²

8.4.3 Smithsonian 3D Digitisation

The Smithsonian Institution has implemented a 3D digitisation program which covers a wide range of museum collections, including natural history. In addition to this, they have designed, tested, and applied multiple 3D digitisation techniques (Laser, Photogrammetry, Structured Light, etc.) [9, 91, 92, 93]. In this context, the appropriate digitisation method is selected according to the characteristics of the specimen (as described in [9] - Item Driven Image Fidelity (IDIF)). The 3D digitisation strategy encompasses specimen selection, development of a custom digitisation station, development of digitisation software, selection of processing software, publishing of imaged

⁴¹ <https://www2.ala.org.au/chuong/>

⁴² <https://www2.ala.org.au/chuong/Bee-from-X3D.html>

specimens, and encouraging use and exploitation of 3D digital resources [93]. Figure 69 shows the Smithsonian Institution metadata model. The metadata model attempts to encompass all the metadata needed to fully document a 3D capture event, and initially focused on photogrammetry capture. Photogrammetry was chosen as the starting point because it's an accessible, non-proprietary technology and the 'raw' data, image files are also non-proprietary and have existing best practices for preservation. Additionally, there is a high amount of complexity around how a photogrammetry project can be executed compared to other 3D capture methods. The plan is that, by addressing photogrammetry first, the model will account for many "edge cases" in capture techniques and will be straightforward to extend to other capture types (laser, micro CT, structured light scanning).

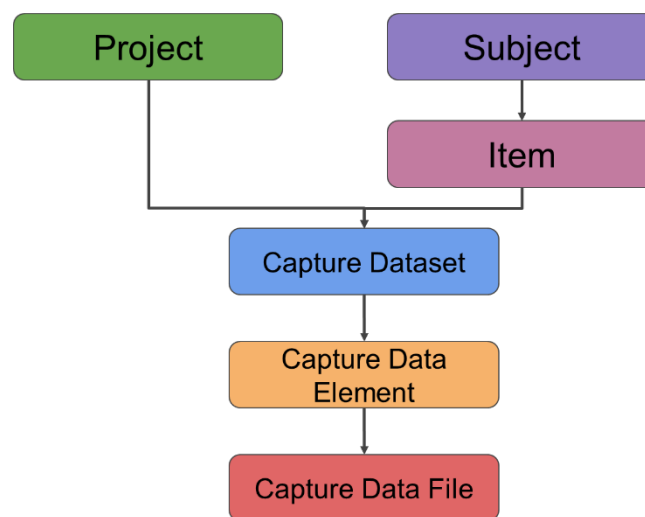


Figure 69 The image shows the common metadata model for 3D assets [92]. This model is adjusted to the definition of a generic common workflow for 3D digitisation which can be adapted to the requirements of any item in any collection.

The Smithsonian 3D digitisation strategy presents an adaptable plan for 3D digitisation of collections containing specimens of variant sizes, by considering a generic workflow and data model which can be adapted to different capture technologies. Moreover, the plan actively seeks to promote the use of the digital specimens in different ways (as shown in Figure 70).

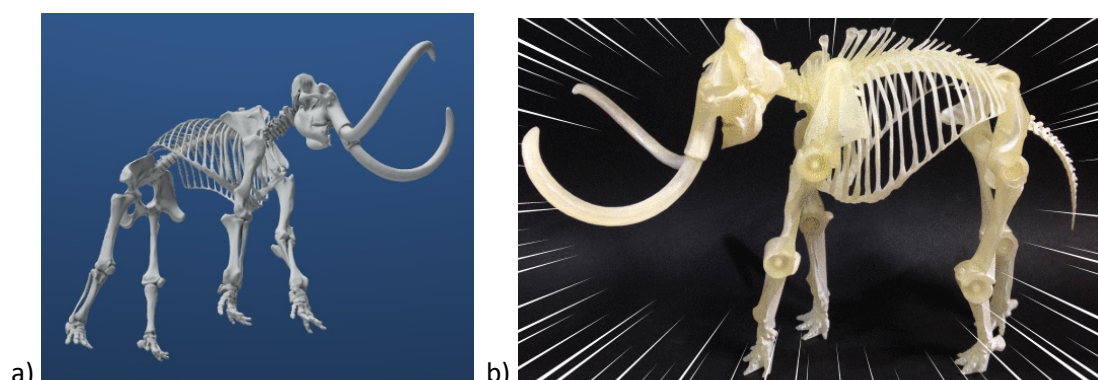


Figure 70 The Smithsonian Institution digitisation project actively encourages the use of digitisation products. For instance, the images above show a published 3D specimen (a) which has been reused to produce a printable articulated model (b) [39]

9. Conclusions

This report provides an overview of the practices and criteria applicable to the workflows and products of NHC digitisation projects.

This work provides a set of guidelines describing:

- the concrete quality criteria which may be applicable for each type of specimen,
- the types of digital products produced following each set of criteria,
- the specific times at which QC/QA verifications should be introduced in existing digitisation workflows, and
- the mechanisms to verify the quality and provenance of a given digital specimen instance.

These guidelines can influence and inform areas of research and development for larger digitisation projects, such as DiSSCo.

Achieving this will require to work with the different teams researching the digitisation workflows (Tasks 3.1 and 3.2) within ICEDIG to validate the basic expectations about the requirements and prospects of the application of quality management in digitisation workflows, and propose updates and extensions to the criteria presented here.



Glossary

Acutance: is the measure of how well the photographic medium handles edge contrast. High acutance gives crisp, clean edges with high edge contrast so they stand out clearly. Low acutance gives fuzzy edges that are less distinct. Film acutance is changed by development; there are high-acutance developers available. In the digital domain sharpening increases acutance [81].

Camera rig: is a modular piece of equipment used to extend the usefulness of a camera, whether through accommodating additional shooting styles, allowing for additional gear to be mounted safely, or for smoothing out the motion of the shot [102].

Colour chart (colour reference card): it is a flat, physical object that has many different colour samples present. They can be available as a one-page chart, or in the form of swatch-books or colour-matching fans [106].

Colour reference charts is a type of colour chart intended for colour comparisons and measurements. Typical tasks for such charts are checking the colour reproduction of an imaging system, aiding in colour management or visually determining the hue of colour. Examples are the IT8 and ColorChecker charts [106].

Colour selection chart is a type of colour chart presented as a palette of available colours to aid the selection of spot colours, process colours, paints, pens, crayons, and so on – usually the colours are from a manufacturer's product range. Examples are the Pantone and RAL systems [106].

Colour Depth or colour depth (see spelling differences), also known as bit depth, is either the number of bits used to indicate the colour of a single pixel, in a bitmapped image or video frame buffer, or the number of bits used for each colour component of a single pixel [107].

Delta E (ΔE , δE , dE) is a metric for understanding how the human eye perceives colour difference. The term delta comes from mathematics, meaning change in a variable or function. The suffix E references the German word *Empfindung*, which broadly means sensation [87].

Dots per inch (DPI, or dpi) is a measure of spatial printing or video or image scanner dot density, in particular the number of individual dots that can be placed in a line within the span of 1 inch (2.54 cm) [108]. Monitors do not have dots, they have pixels; the closely related concept for monitors and images is pixels per inch (or PPI). Many resources use the terms DPI and PPI interchangeably.

Image resolution is the detail an image holds. The term applies to raster digital images, film images, and other types of images. Higher resolution means more image detail [109].

Quality assurance (QA) is a way of preventing mistakes and defects in manufactured products and avoiding problems when delivering solutions or services to customers. ISO 9000 defines QA as "part of quality management focused on providing confidence that quality requirements will be fulfilled"[115].



Quality control (QC) is a process by which entities review the quality of all factors involved in production. ISO 9000 defines QC as "A part of quality management focused on fulfilling quality requirements"[116].

Pixels per inch (PPI, or ppi) is a measurement of the pixel density (resolution) of an electronic image device, such as a computer monitor or television display, or image digitizing device such as a camera or image scanner. PPI can also describe the resolution, in pixels, of an image file [114].

Radiometric resolution determines how finely a system can represent or distinguish differences of intensity and is usually expressed as several levels or a few bits, for example 8 bits or 256 levels that is typical of computer image files. The higher the radiometric resolution, the better subtle differences of intensity or reflectivity can be represented, at least in theory. In practice, the effective radiometric resolution is typically limited by the noise level, rather than by the number of bits of representation [110].

Scale refers to apparatuses or systems used for measuring: the graduated marks on a line or rule used to measure distances and ascertain relative dimensions; the equally divided grid-lines on the surface of a map, chart or plan that enable ratios of area and distance to be established; the ratio pertaining between a model and the reality it represents or projects [35].

Sharpness is a subjective term combining the ideas of resolution (amount of detail) and acutance. A sharp image is one that is clean, crisp and finely detailed. The main causes of an image not being sharp are camera shake, poor focus and possibly using too small a lens aperture (diffraction effect). The ideal strategy is to get as sharp an image as possible from the camera. Post-camera sharpening then needs to be applied. With film, this may involve choice of developer and good darkroom practice with critical focus on the enlarger and flatness of the negative and paper. All digital images and scans need sharpening in stages: first for capture, possibly locally for creative effects and final sharpening for a specific output medium. Sharpening works by increasing local contrast along edge detail [81].

Spatial Resolution The measure of how closely lines can be resolved in an image. Spatial resolution depends on properties of the system creating the image, not just the pixel resolution in pixels per inch (PPI). For practical purposes, the clarity of the image is decided by its spatial resolution, not the number of pixels in an image. In effect, spatial resolution refers to the number of independent pixel values per unit length [111].

Spectral resolution is the ability to resolve spectral features and bands into their separate components. Colour images distinguish light of different spectra. Multispectral images can resolve even finer differences of spectrum or wavelength by measuring and storing more than the traditional three of common RGB colour images [112].

Temporal resolution refers to the precision of a measurement with respect to time. Movie cameras and high-speed cameras can resolve events at different points in time. The time resolution used for movies is usually 24 to 48 frames per second (frames/s), whereas high-speed cameras may resolve 50 to 300 frames/s, or even more. Many cameras and displays offset the colour components relative to each other or mix up temporal with spatial resolution [113].



References

1. Abouelaziz, et. al. (2015) Reduced Reference 3D Mesh Quality Assessment Based on Statistical Models 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems DOI: 10.1109/SITIS.2015.129
2. Abouelaziz, I., El Hassouni, M., Cherifi, H. (2016) No-Reference 3D Mesh Quality Assessment Based on Dihedral Angles Model and Support Vector Regression A. Mansouri et al. (Eds.): ICISP 2016, LNCS 9680, pp. 369–377, 2016. DOI: 10.1007/978-3-319-33618-3_37
3. Abouelaziz, I., El Hassouni, M., Cherifi, H. (2017) A convolutional neural network framework for blind mesh visual quality assessment 2017 IEEE International Conference on Image Processing (ICIP), 17-20 Sept. 2017, Beijing, China DOI:10.1109/ICIP.2017.8296382
4. Abouelaziz, I., El Hassouni, M., Cherifi, H. (2018) Blind 3D mesh visual quality assessment using support vector regression Springer Multimedia Tools (2018) 77:24365–24386 <https://doi.org/10.1007/s11042-018-5706-1>
5. Alexiou, E., Ebrahimi, T. (2018) Point cloud quality assessment metric based on angular similarity IEEE 2018 978-1-5386-1737-3/18 IEEE
6. Allan, L., Hardy H., Livermore, L., Nieva, A., Owen D., Rosin, P., Smith, V., Sun, X., Woodburn, M., (2018-04-10) Work meeting at Natural History Museum London to discuss collaboration on WP3 tasks.
7. Ameisen, D. et.al. (2013) Stack or trash? Quality assessment of virtual slides. Pathology, vol. 8, no. 1, pp. 1746–1596, 2013. 3
8. American Museum of Natural History (2018) Fluid Preserved Specimens. <https://www.amnh.org/our-research/natural-science-collections-conservation/general-conservation/health-safety/fluid-preserved-specimens>.
9. Anderson N. I. (2017) Item Driven Image Fidelity (IDIF) or “HITTING the Digital Capture Sweet Spot”, Smithsonian Digitization.
10. Ariño, A. H., Galicia, D. (2005). Taxonomic Grade Images. 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.
11. Atlas of Living Australia website at <http://www.ala.org.au>. Accessed November 2018. Digital specimens/pages cited where used
12. Brecko, J., Mathys, A., et.al. (2016). "Handbook of best practice and standards for 3D imaging of Natural History specimens," SYNTHESYS3 project report, 2016. <https://drive.google.com/open?id=0B2yIFo9B44xfWk1WSHgXVU51WVk>
13. Brecko J., Mathys A., et.al. (2016). Handbook of Best Practice and Standards for 3D imaging of Natural History specimens. SYNTHESYS3 Project Deliverable. Retrieved April 10, 2018: http://biowikifarm.net/v-mfn/3d-handbook/3d_Imaging_Handbook:Main_Page
14. Brilmyer G (2013) Ant Databasing Project Protocols, Moreau Lab. From: https://www.idigbio.org/wiki/images/a/aa/Ant_Databasing_Protocols_August.pdf
15. CheckMate Advisory Board (2018) CheckMate 3D Modeling Standard. <https://www.turbosquid.com/3d-modeling/checkmate/checkmate-specifications/checkmate-pro-specification/>
16. Cignoni, P. Rocchini, C., Scopigno R. (1988) Metro: Measuring Error on Simplified Surfaces Computer Graphics Forum Volume 17 (1988) pp. 167-174
17. Collins, C. (2014). Standards in the Care of Wet Collections. Conservation Centre at the Natural History Museum. Mon, 2014-07-21 14:14. Retrieved: 2018/10/29. From: <http://conservation.myspecies.info/node/33>



18. Cultural Heritage (2018) Numerically Evaluating Image Quality, Digital Transitions Division of Cultural Heritage, Retrieved on 2018-08-21, From: <https://dtdch.com/digitization-program-planning/numerically-evaluating-image-quality/>
19. De Smedt, S., Bogaerts, A., Groom, Q., & Engledow, H. (2018). Botanicalcollections. be: The new virtual herbarium of Meise Botanic Garden (BR). *Biodiversity Information Science and Standards*, 37. doi: 10.3897/biss.2.26140
20. Digitarium (2017). Digitarium digitises collections of the Natural History Museum of Norway. From: http://www.uef.fi/arkisto/uef-en/2013/Digitarium_digitises_collections_of_the_Natural_History_Museum_of_Norway.html
21. Drob, C. (2013) Overview About Project Quality Management. *Studies and Scientific Researches. Economics Edition*, No 18, 2013. <http://dx.doi.org/10.29358/sceco.v0i18.212>
22. Dumić, E., Duarte, C. R., da Silva Cruz, L.A. (2018) Subjective evaluation and objective measures for point clouds — State of the art 2018 First International Colloquium on Smart Grid Metrology (SmaGriMet) DOI: 10.23919/SMAGRIMET.2018.8369848
23. Durrant, J. (2018) Semi-supervised semantic and instance segmentation, Labelling Training Data. Retrieved on 2018-02-13, from: <https://github.com/NaturalHistoryMuseum/semantic-segmentation/wiki/Labelling-training-data>
24. EDC (2018) Resolution for Remote Sensing. Environmental Data Centre, University of Rhode Island. Retrieved on: 21/08/2018 from: www.edc.uri.edu/nrs/classes/NRS409/RS/Lectures/HowRemoteSensWork.pdf
25. EDICO SK (2016) Witikon - Mass 3D digitisation at a national scale, fully automated 3D digitization of cultural heritage artefacts, Retrieved April 17, 2018: <http://witikon.tichevody.sk/wp-content/uploads/2016/06/Witikon-nationwide-mass-digitization-23062016.pdf>
26. EDICO SK (2018) Witikon Photogrammetry System, Retrieved April 17, 2018: <http://witikon.eu/>
27. Elloumi, N. et. al. (2013) Perceptual Metrics Quality: Comparative Study for 3D Static Meshes 2013 IEEE International Conference on Systems, Man, and Cybernetics DOI 10.1109/SMC.2013.557
28. Engledow, H., De Smedt, S., Groom, Q., Bogaerts, A., Stoffelen, P., Sosef, M., & Van Wambeke, P. (2018). Managing a Mass Digitization Project at Meise Botanic Garden: From Start to Finish. *Biodiversity Information Science and Standards*. doi: 10.3897/biss.2.25912
29. ENVRI plus (2018). The ENVRI Reference Model. From: <http://envri.eu/rm>
30. Evangelos, E., Touradj, E. (2018) Benchmarking of objective quality metrics for colorless point clouds IEEE. 2018 33rd Picture Coding Workshop (PCS 2018), 24-27 June 2018 in San Francisco, United States. DOI: 10.1109/PCS.2018.8456252
31. Federal Agencies Digital Guidelines Initiative (2016) Technical Guidelines for Digitizing Cultural Heritage Materials. From: http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf
32. Federal Agencies Digital Guidelines Initiative (2018) OpenDICE and AutoSFR. From: <http://www.digitizationguidelines.gov/guidelines/digitize-OpenDice.html>
33. Feng, X., et.al. (2013) A perceptual quality metric for 3D triangle meshes based on spatial pooling 2013 IEEE International Conference on Systems, Man, and Cybernetics
34. Finnish Biodiversity Information Facility (FinBIF) (2018). Laji.fi The portal of the. Finnish Biodiversity Information Facility, <https://laji.fi/>. Digital specimens/pages cited where used.
35. Fisher, A. (2012). Photographic Scale. *Philosophy of Photography*. Volume 3 Number 2. doi: 10.1386/pop.3.2.310_1
36. Fraunhofer IGD (2017) 3D Scanning Technologies, Fraunhofer Institute for Computer Graphics research IGD. Brochure. https://www.cultlab3d.de/wp-content/uploads/2018/05/2017-02-27_CHD-Brosch%C3%BCre_EN_web.pdf



37. García Rojo, M., et.al. (2006) Critical Comparison of 31 Commercially Available Digital Slide Systems in Pathology. *International Journal of Surgical Pathology* Volume 14 Number 4 October 2006 285-305. doi: 10.1177/1066896906292274
38. Gassó Miracle, E. (2017) 15 things you didn't know about the Naturalis collection. Naturalis Biodiversity Centre. Published: 2017-07-04, Retrieved 2018-11-20, from: <https://science.naturalis.nl/en/about-us/news/collection/15-things-you-didnt-know-about-naturalis-collection/>
39. Gensyou, T. (2018) Articulated Woolly Mammoth Manga, 16/05/2018. <https://dpo.si.edu/blog/articulated-woolly-mammoth-manga>
40. Guo J. (2016) Contributions to objective and subjective visual quality assessment of 3d models Doctoral Thesis, University of Lyon
41. Haston, E., et al. (2012) Developing integrated workflows for the digitisation of herbarium specimens using a modular and scalable approach. *ZooKeys* 209: 93–102 (2012)
42. Häuser, C.L., Steiner, A., Holstein, J. & Scoble, M. J. (eds.) (2005): *Digital Imaging of Biological Type Specimens. A Manual of Best Practice. Results from a study of the European Network for Biodiversity Information.* Stuttgart. viii + 309 pp.
43. Heerlien, M. et.al. (2015) The Natural History Production Line: An Industrial Approach to the Digitization of Scientific Collections. *ACM J. Comput. Cult. Herit.* 8, 1, Article 3 (February 2015), 11 pages. doi: <http://dx.doi.org/10.1145/2644822>
44. Hudson, et. al. (2015) INSELECT: Automating the Digitization of Natural History Collections. Published: November 23, 2015 <https://doi.org/10.1371/journal.pone.0143402>
45. I-materialise (2018) Beginners Guide to 3D printing <https://i.materialise.com/en>
46. ICEDIG Project (2017). ICEDIG Description of Activities (DoA), 777483 ICEDIG – Part B
47. ICEDIG Project (2019). MS 17 Interim report on rapid 3D capture methods in biological collections and related fields. Work in progress, unpublished (Due 2019/09/30).
48. ICEDIG Project (2019). D3.7 Rapid 3D capture methods in biological collections and related fields. 2019/01/31.
49. ICEDIG Project (2019). D6.6 Provisional Data Management Plan for the DiSSCo infrastructure. Work in progress, unpublished (Due 2019/10/30).
50. iDigBio (2015) Workflow Modules and Task Lists, Recovered from: <https://www.idigbio.org/content/workflow-modules-and-task-lists>
51. InvertEBase (2018) InvertEBase Data Portal Homepage, digital specimens cited where used.
52. Image Access (2017) Frequently Asked Questions, Quality Controlled Scanning, ISO 19264, FADGI, Metamorfoze. Image Access GmbH, Germany. Accessed on 20/07/2018, From: https://support.imageaccess.de/downloads/product_manuals/FAQ/FAQ-Quality-Controlled-Scanning.pdf
53. Image Science Associates (2017) Object-Level Target Family, Accessed on 21/08/2018, From: http://www.imagescienceassociates.com/mm5/merchant.mvc?Screen=PROD&Store_Code=ISA001&Product_Code=OL1&Category_Code=GT
54. JSTOR (2018). JSTOR Global Plants: Guidelines for Scanning Specimens. From: https://guides.jstor.org/ld.php?content_id=31764146
55. JSTOR (2018). JSTOR Plants Handbook. From <http://www.snsb.info/SNSBInfoOpenWiki/attach/Attachments/JSTOR-Plants-Handbook.pdf>
56. Kingham, E. L. (2016). Project Pickle – Conserving our Specimens Preserved in Fluid. *Museums & Collections Blog*, University College London, on 1 August 2016. Retrieved: 2018/10/29. From: <http://blogs.ucl.ac.uk/museums/2016/08/01/project-pickle-conserving-our-specimens-preserved-in-fluid/>



57. Krishnan, A. Tamma, K. (2016). Divergent morphological and acoustic traits in sympatric communities of Asian barbets. *Royal Society Open Science*. 3: 160117.
<http://dx.doi.org/10.1098/rsos.160117>
58. Kroupa A, Pluta M, Schurian B, Glöcker F (2014) ZooSphere - A tool for automated spheric image capturing and interactive 3D visualization of biological collection objects. The Society for the Preservation of Natural History Collections annual conference, June 22-27, 2014, Cardiff, UK.
https://www.researchgate.net/publication/268779382_ZooSphere_-_A_tool_for_automated_spheric_image_capturing_and_interactive_3D_visualization_of_biological_collection_objects
59. Lavoué, G. (2011) A Multiscale Metric for 3D Mesh Visual Quality Assessment Eurographics Symposium on Geometry Processing 2011, Editors Mario Botsch and Scott Schaefer. Volume 30 (2011), Number 5
60. Lavoué, G., Cheng, I., Basu, A. (2018) Perceptual quality metrics for 3D meshes: Towards an optimal multi-attribute computational model. 2013 IEEE International Conference on Systems, Man, and Cybernetics. 978-1-5386-5195-7/18 06722310
61. Library of Congress (2018). Technical Standards for Digital Conversion of Text and Graphic Materials From: <https://memory.loc.gov/ammem/about/techStandards.pdf>
62. Livermore, L., et.al. (2017) Digitising Louse Slides. NERC / Natural History Museum pilot project.
<http://www.nhm.ac.uk/our-science/our-work/digital-museum/digital-collections-programme/digitising-slide-collections.html>
63. Mantle, B.L., La Salle, J., Fisher, N. (2012). Whole-drawer imaging for digital management and curation of a large entomological collection. In: Blagoderov V, Smith VS (Ed) No specimen left behind: mass digitization of natural history collections. *ZooKeys* 209: 147–163. doi: 10.3897/zookeys.209.3169
64. Meyer, R (2014). Imaging Insects with Digital Slide Scanners. Meyer Instruments. Presented during the webinar from the Fluid-preserved Arthropod and Microscopic Slide Imaging Interest Group on 17 June 2014. From: https://www.idigbio.org/wiki/images/9/91/Imaging_Insects_with_Scanners.pdf
65. Muzahid, A. A. M., Wan, W., Feng, X. (2018) Perceptual Quality Evaluation of 3D Triangle Mesh: A Technical Review ICALIP 2018: International Conference on Audio, Language and Image Processing, Shanghai, China dx.doi.org/10.1109/ICALIP.2018.8455210
66. National Archives (2016) Digitisation at The National Archives. From: <http://www.nationalarchives.gov.uk/documents/information-management/digitisation-at-the-national-archives.pdf>
67. Natural History Museum (2018) The Natural History Museum Data Portal, digital specimens cited where used.
68. Natural History Museum (2018) Imaging and Analysis Centre, <http://www.nhm.ac.uk/our-science/departments-and-staff/core-research-labs/imaging-and-analysis.html>, digital specimens cited where used.
69. Naturalis Biodiversity Centre (2018) BioPortal, the Data Portal of the Naturalis Biodiversity Centre, digital specimens cited where used.
70. Nelson, G., Paul, D., Riccardi, G., Mast, A. R. (2012) Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys* 209: 19-45. doi: 10.3897/zookeys.209.3135
71. Nelson, G., et. al. (2015) Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Applications in Plant Sciences* 2015 3(9): 1500065Ap
72. Nguyen, C. et. al. (2014). Capturing natural-colour 3D models of insects for species discovery and diagnostics. *PloS one*, 9(4), e94346
73. Nguyen, C. et. al. (2017) Towards high-throughput 3D insect capture for species discovery and diagnostics. *eScience 2017, BigDig 24 Oct - High Throughput Digitization for Natural History*



Collections from: <https://press3.mcs.anl.gov//bigdig/files/2017/09/Nguyen-Shared-BigDig2017-Towards-high-throughput-3D-insect-capture.pdf>

74. Nieva de la Hidalga A, Owen D, Spasic I, Rosin P, Sun X (2019) Use of Semantic Segmentation for Increasing the Throughput of Digitisation Workflows for Natural History Collections. *Biodiversity Information Science and Standards* 3: e37161. <https://doi.org/10.3897/biss.3.37161>
75. Nouri, A., Charrier, C., L'ezoray, O. (2016) Full-reference saliency-based 3D mesh quality assessment index 2016 IEEE International Conference on Image Processing (ICIP)
76. Nouri, A., Charrier, C., L'ezoray, O. (2017) 3d blind mesh quality assessment index Society for Imaging Science and Technology. *Multimed Tools Appl* (2018) 77:24365–2438 doi.org/10.2352/ISSN.2470-1173.2017.20.3DIPM-002
77. O'Brien, M. F. (2016) DIY Vertebrate Specimen Photography, iDigBio, Strategies for Vertebrate Digitization Workshop. From: <https://www.idigbio.org/wiki/images/7/77/MFO-DIY-APR2016.pdf>
78. Patel AA, Gil-Bertson J, Dhir R, Becich MJ. (2003) Virtual quality assurance protocol: a quality assurance protocol for pathology slides from multiple institutes. *Advancing pathology informatics, imaging, and the Internet (APIII 2002)*. Scientific Session and E-Poster Session Abstracts. *Arch Pathol Lab Med*. 2003; 127:804-805. Available from: <http://www.archivesofpathology.org/doi/full/10.1043/1543-2165%282003%29127%3C789%3AAPIIAT%3E2.0.CO%3B2>
79. Phillips, S., Green, L., and Weech, M.H. (2014). Optimal digitisation pilot study, SYNTHESYS3, Deliverable 3.3, <http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.3-Review-of-Digitisation-workflows-and-equipment.pdf>
80. Picturae (2017) Biodiversity Digitizing Herbarium Sheets, From: <https://picturae.com/images/download/EN-Biodiversity-web.pdf>
81. Prakel D. (2010). *The Visual Dictionary of Photography*. AVA Publishing. pp. 19–. ISBN 978-2-940411-04-7.
82. Price, B. et.al. (2018) ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation. Preprint. Open Science Framework. doi:10.17605/OSF.IO/UVWRN
83. Rainbow, P. S. (2009) Marine biological collections in the 21st century. *Zoologica Scripta* 38 (2009): 33-40.
84. Ryder S. (2018). Dataset: Wallace and Banks drawers. Resource: Drawer-level images. Natural History Museum Data Portal (data.nhm.ac.uk). <https://doi.org/10.5519/0097454>, Retrieved: 2018/11/02 09:02 (GMT)
85. Sabaj Perez, M. H. (2009) Photographic atlas of fishes of the Guiana Shield. p. 53–93 In: Vari, R. P., C. J. Ferraris, Jr., A. Radosavljevic, and V. A. Funk, eds. Checklist of the freshwater fishes of the Guiana Shield. *Bulletin of the Biological Society of Washington*, no. 17
86. Samak, A. (2015) An Optimization Approach to Scanning Skin Direct Immunofluorescence Specimens, IEEE 978-1-4673-6795-0/15
87. Schuessler Z. (2016) Delta E 101. From: <http://zschuessler.github.io/DeltaE/learn/>
88. Schurian, B. (2018) ZooSphere workflow demonstration at Museum of Natural History Berlin (MfN) on the 22 of March 2018.
89. Seregin, A. P. (2017). Digital Herbarium of Moscow State University: The Largest Russian Biodiversity Database *Biology Bulletin*, 2017, Vol. 44, No. 6, pp. 584–590.
90. Shankar Paul, J., Tiwari, K.L., Jadhav, S.K. (2015). Long Term Preservation of Commercial Important Fungi in Glycerolat 4°C. *International Journal of Biological Chemistry* 9 (2): 79-85, 2015. ISSN 1819-155X / DOI: 10.3923/ijbc.2015.79.85
91. Smithsonian Institution (2018) Preserving specimens in liquid preservatives. Department of Botany. <https://botany.si.edu/projects/algae/collpres/liqpres.htm>
92. Smithsonian Institution (2018) Smithsonian 3D Metadata Model. Smithsonian Institution Digitization Program Office. <https://dpo.si.edu/blog/smithsonian-3d-metadata-model>



93. Smithsonian Institution (2018) Smithsonian 3D Programme. Smithsonian Institution Digitization Program Office. <https://3d.si.edu>
94. Sonka, M., Hlavac, V. Boyle, R. (2008) Image Processing Analysis and Machine Vision, Thompson 2D Digitisation of 2D fossils
95. Speers, L. (2005). E-Types – A New Resource for Taxonomic Research. 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.
96. Summerfield, R., et.al. (2016). Digitising Slide Collections. NERC / Natural History Museum pilot project. <http://www.nhm.ac.uk/our-science/our-work/digital-museum/digital-collections-programme/digitising-slide-collections1.html>
97. Thakur D. R. (2010) Process Quality Management. International Journal of Management Prudence. Volume 1 Issue 1 pp 74-80 <http://www.publishingindia.com/ijmp/19/process-quality-management/53/235/>
98. Tegelberg et. al. (2014) High-performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. TAXON 63 (6). December 2014: 1307–1313
99. TurboSquid (2017) StemCell <https://www.turbosquid.com>
100. University of Navarra (2005) Museum of Zoology - Natural History Collections. Digital specimens cited where used. <http://www.unav.es/unzyec/mzna/indexi.html>
101. University of Nebraska (2018) Basic Insect Morphology, University of Nebraska-Lincoln, Department of Entomology. Retrieved 2018/11/02 17:25, From: <https://entomology.unl.edu/scilit/basic-insect-morphology>
102. Videomaker (2013) Camera Rig Buyer's Guide, from: <https://www.videomaker.com/article/f22/17184-camera-rig-buyer's-guide>
103. Vollmar, A., Macklin, J.A., Ford, L. S. (2010) Specimen Digitization: Challenges and Concerns. Biodiversity Informatics, 7, 2010, pp. 93 – 112
104. Watson, K. (2014) Herbarium Specimen Imaging: Standards and Suggestions, iDigBio Small Herbarium Workshop – Botany 2014. The New York Botanical Garden. https://www.idigbio.org/wiki/images/8/8e/Watson_iDigBio-Botany-2014-Workshop.pdf
105. Wikipedia (2018) Anatomical Terms of Location, from: https://en.wikipedia.org/wiki/Anatomical_terms_of_location
106. Wikipedia (2018) Colour Chart, from: https://en.wikipedia.org/wiki/Color_chart
107. Wikipedia (2018) Colour Depth, from: https://en.wikipedia.org/wiki/Color_depth
108. Wikipedia (2018) Dots per inch, from: https://en.wikipedia.org/wiki/Dots_per_inch
109. Wikipedia (2018) Image Resolution, from: https://en.wikipedia.org/wiki/Image_resolution
110. Wikipedia (2018) Image Resolution Radiometric Resolution, from: https://en.wikipedia.org/wiki/Image_resolution#Radiometric_resolution
111. Wikipedia (2018) Image Resolution Spatial Resolution, from: https://en.wikipedia.org/wiki/Image_resolution#Spatial_resolution
112. Wikipedia (2018) Image Resolution Spectral Resolution, from: https://en.wikipedia.org/wiki/Image_resolution#Spectral_resolution
113. Wikipedia (2018) Image Resolution Temporal Resolution, from: https://en.wikipedia.org/wiki/Image_resolution#Temporal_resolution
114. Wikipedia (2018) Pixel density, from: https://en.wikipedia.org/wiki/Pixel_density
115. Wikipedia (2018) Quality Assurance, from: https://en.wikipedia.org/wiki/Quality_assurance
116. Wikipedia (2018) Quality Control, from: https://en.wikipedia.org/wiki/Quality_control
117. Yildiz, Z. C., Oztireli, A.C., Capin, T. (2018) A machine learning framework for full-reference 3D shape quality assessment. <https://doi.org/10.1007/s00371-018-1592-9>



118. Zaman, T. (2015) 2D vs 1D Barcodes for Digital Image Readout. July 8th, 2015. Retrieved: 2018-08-21, From:
https://docs.google.com/document/d/1CjK0yTWIWr0pAkYSJVYocZDZ2_NhZutFvLe81XnGJfc
119. Zhang, T. et. al. (2018) SlideNet: Fast and Accurate Slide Quality Assessment Based on Deep Neural Networks
120. ZooSphere Project (2018) ZooSphere Project Data Portal, from Berlin Natural History Museum, digital specimens cited where used.

