Project:	Synthesis of systematic resources
Project acronym:	SYNTHESYS3
Grant Agreement number:	312253
Deliverable number:	3.3
Deliverable title:	Optimal digitisation pilot study
Deliverable author(s):	Sarah Phillips, Laura Green and Marie-Hélène Weech, Royal Botanic Gardens Kew
Date:	December 2014

Review of Current Digitisation Workflows and Equipment

Authors: Sarah Phillips[1], Laura Green[1] and Marie-Hélène Weech[1]

[1] RBG Kew, Surrey

Introduction

Recent developments in digitisation technologies and equipment have enabled advances in the rate of natural history specimen digitisation. However Europe's collections are home to over one billion specimens and currently only a small fraction of these specimens have been digitally catalogued with fewer still imaged. Globally, natural history collection data has been estimated to be between 1.2 to 2.1 gigaunits (specimens, lots and collections) of which a mere 3% is available on the Global Biodiversity Information Facility GBIF (http://www.gbif.org) (Ariño 2010). It is clear that institutions still face huge challenges when digitising the vast number of specimens in their collections. The aim of this study was to gather information from all SYNTHESYS3 partners on their current digitisation facilities, equipment and workflows and try to identify what were the biggest challenges they faced in their digitisation programmes and how they were prioritising the digitisation of their collections. A questionnaire was sent to all 18 partners which covered topics including equipment, workflows, data management and quality assurance, and challenges and future development. This review summarises some of the main components and underlying issues with respect to digitisation workflows as evidenced from the questionnaire responses, as well as providing some key recommendations based on these findings.

Questionnaire Results

COLLECTIONS

Nearly all institutions house more than one type of collection, with the number of different collections per institution averaging between 5 and 6. The size of each collection varies greatly from 15,000 specimens up to collections of 80 million but the average collection size was approximately 6 million. The prioritisation for collection digitisation has so far been driven by four main factors, digitisation of type specimens, discrete project-based funding to digitise a specific subset of the collections, research needs of the institution and loan or image requests. Other factors reported include prioritising the digitisation of new accessions into the collection, selecting a small sized collection to test out digitisation methodologies before rolling out the methodology to other collections, digitising specimens for public exhibitions and selecting specimens based on the overall condition of the collection.

A survey containing questions regarding general digitisation protocols, types of collections, quality control, data management, data storage and access and future development was sent to 18 SYNTHESYS3 partners (see Appendix 2.). Data was received from 14 partners with information on their collections and digitisation workflows.

Three institutions reported sending out a questionnaire to gather information and proposals for collections to be digitised, basing the resulting decision on scores for the factors

prioritised for that institution, e.g. research aims, collection management, external funding potential and public exhibition function.

Collections can be broadly broken down into the following groups:

Botanical Collections

All institutions housing botanical specimens (approx 70%) have digitised at least some parts of their collections. Collection sizes range from just over 60 thousand to over 6.5 million. The proportion of specimens that have been digitised varies greatly between institutions, however with the exception of Naturalis Biodiversity Centre, Leiden, which outsourced their digitisation process, institutions have a larger proportion of specimens databased than they have imaged with the number of specimens imaged being less than 1% to 10% on average. The average number of specimens digitised in-house per year ranged from around 1,500 to 75,000. Naturalis has imaged more than 90% of its specimens and captured metadata from approximately 66%, specimens are imaged first and the data captured later from the images. The Muséum national d'Histoire naturelle of Paris (MNHN) have also completed a large-scale digitisation program of herbarium specimens (P, PC) with nearly 6 million images available online. Only a minimum number of data fields were captured including scientific name, catalogue number and continent of origin. Optical character recognition and a crowdsourcing site have since been used to enrich the data from the images captured (https://www.idigbio.org/content/spnhc-2014-progress-digitization-massive-digitisation-paris-herbarium-nation-wide-program).

The time taken to digitise (database and image) a specimen was reported as being between 10 and 25 minutes. Longer times quoted were 60 minutes for cryptogam specimens and wood specimens. The Royal Museum for Central Africa, Tervuren, found wood samples require more complex imaging, 2D pictures of transversal sections are taken at magnifications of 2.5x and 4x, and repeated three times per specimen. This complex imaging reduced the number of specimens that can be imaged per day to 42 and sometimes new, thin sections needed to be taken also reducing the digitisation rate.

The quoted costs of digitisation per specimen varies, probably reflecting the difficulties in estimating costs and ensuring estimates are standardised by including the same factors. Three institutions report a cost of around 10 euros (20 euros for cryptogams) whereas other estimates vary from 0.6 euros to 1.64 euros.

Zoological Collections

Just over half of the institutions have zoological collections and all have digitised at least part of their collections. Collection sizes ranged from 650,000 to 27.5 million specimens and the percentage of collections digitised varied widely, probably partly due to the many different types of collections. However, many institutions reported that a large proportion of their type collections were completed, at least for specimen metadata. Although unclear from the answers given it seems that a much smaller percentage of specimens have been imaged. The average number of specimens digitised per year is much lower than that of botanical specimens between the ranges of 1,000 to 10,000 specimens per year.

The average time taken to digitise a specimen also appears to be longer than botanical specimens ranging from 15 to 45 minutes for 2D images and up to 3 hours for 3D specimens. One institution quotes much longer times up to 3-4 weeks however this is probably due to the fact that specimen identification has been included in this workflow. Zoological images captured per specimen vary between different institutions, many capturing more than one image of the specimens, from different angles. This appears to be reflected in the varying cost of digitisation which was reported to be greater than botanical specimens, ranging from 2 euros to 16 euros for 2D imaging and up to 28 euros for 3D imaging.

Entomological Collections

The 7 institutions with entomological collections all have large collections varying from 2.5 to 33 million specimens. Collection digitisation is mainly prioritised by the type collections and research needs. The percentage digitised ranges from less than 1% to 20%. For those institutions actively digitising their collections, the numbers digitised per year are between 3,000 to 10,000 specimens. Most of the partners appear to have similar numbers imaged as databased, this may be because for pinned insects the labels need to be removed from the pin for both complete databasing and imaging (Mantle et al., 2012) meaning both tasks may be completed simultaneously. Minimal information was provided but the time taken to digitise a specimen ranged from 10 to 60 minutes, the latter figure involved taking images of the front and back of the specimen. Costs, where provided, ranged between 1.63 to 5 euros per specimen. Two institutions reported rates per specimen drawer rather than individual specimens: drawers were digitised at a rate between 5-45 minutes at a cost 15 euros per drawer. The number of drawers digitised per year ranged between 2,000-5000. Each drawer or tray contains up to several hundred specimens, most drawers and specimens were given a 2D barcode (QR code or DataMatrix).

Paleontological Collections

Of all the institutions that house paleontological specimens (6 of 14) all but two contain collections between 1 million and 5 million specimens. However a relatively small proportion of the collections have been imaged when compared to other collections c. 2-4%. One exception to this is the Swedish Museum of Natural History that has completed 15.7 % of its 1,150,000 specimen paleontological collection. Parts of the collection were prioritised and a large concentration of these specimens have been digitised while other areas of the collection have had limited digitisation. The main prioritisation for many of the institutions appears to focus more around current research aims rather than type specimens. The average number of specimens digitised per year ranges between 3,000 and 10,000. The average digitisation time varies greatly between 5 to 60 minutes, depending on the complexity of the imaging which is completed: paleontological objects tend to be 3D scanned and imaged using X-rays. Costs per specimen given also vary widely between 0.5 to 21 euros, again this is likely to be dependent on the imaging stage.

Geological Collections

Five partners reported having geological and/or mineralogical collections and the sizes of these collections were mainly below 1 million. Digitisation of these collections appears to be more concentrated on capturing specimen metadata, with less than 1% of any collection

imaged. The average number of specimens digitised per year was lower than other collections ranging between 350 to 4,000 specimens. Very little information was given but the time required to digitise a specimen ranged from 10 to 30 minutes and the costs varied between 0.4 to 50 euros. The highest time and cost reported is probably due to the incorporation of specimen cleaning within the workflow.

Anthropological and Archaeological Collections

There are only 4 institutions with anthropological collections: One partner has databased 70% of the inventory books for the collection; another has imaged around 15%, prioritising human ancestor specimens and death masks. The annual specimen digitisation rate reported ranged from 700-1500 specimens and the time to digitise a specimen ranged from 30 minutes to 3 hours. There are only two institutions with archaeological collections: one databased 90% of all prehistoric collections (no images) into Microsoft Access.

Other Collections

Other collections reported by partners include mycology, economic botany, book and journal collections and microscope slides. Specimens are generally digitised through specific projects and in response to requests.

DIGITISATION STAFF

The numbers of individuals working within the digitisation teams varied greatly between organisations from 2 to more than 30, spread over full-time and part-time members of staff and volunteers. However, it is clear that different institutions answered the questionnaires differently, so where some included all individuals working on the projects (either full- or part-time) others only included full-time members working on the collections.

Of the 14 respondents, two said their staff worked mainly independently while engaged in digitisation workflows, whereas most other institutions said their staff worked both independently and in teams. Either because there was a team of people each performing the same tasks, and thus members of the team were working on their own, or because team members had different roles and thus did not share much of the workflow. Overall, it seems most staff working on digitisation across institutions were working individually for at least one step of the workflow, within a coordinated team.

PRE-DIGITISATION CURATION

Approximately two thirds of the institutions performed at least some minimal curation, specimen conservation or collections management steps prior to digitisation. The most common pre-digitisation step was to check the identification of the specimens to be digitised or check the type status. Many institutions took the opportunity to complete a conservation

assessment of the specimens to be digitised, this included checking for pests and pest damage, changing old covers, specimen repairs and remounting of damaged botanical specimens. Specimen cleaning was regularly part of the workflow for geological collections, for entomological collections specimen drawers were often renewed or replaced and for some paleontological collections, boxes were replaced with acid-free boxes. One institution reported reorganising their botanical specimens according to geographical states for the part of the collection being digitised for a specific project. Only one institution mentioned the need to prepare collections to be handled due to health and safety aspects, as many historical collections have been treated with chemicals that are now considered unsafe: vertebrate skins have been treated with arsenic, herbarium collections with mercury, geology collections store asbestos-containing material.

One institution reported that in order to meet large quantity-driven digitisation targets, they chose to select only those collections that did not require any additional curation or collections management. However, they did give curators advanced notice to allow them to address some of the issues, e.g. label quality, before digitisation began.

DIGITISATION EQUIPMENT

There is a wide variety of imaging equipment being used in and between institutions, reflecting the wide variety of collections in the different institutions. The complete list of equipment can be seen in Appendix 1.

Most of the partners have their own digitisation facilities, usually located within the same building as the collections. However there are some exceptions, such as where the partner has outsourced the digitisation of their material, in which case the collection material has to be transported off site to the equipment location. Four institutions reported instances of locating the equipment directly within the collections area (National Museum Prague, Royal Botanic Gardens Kew, London, Natural History Museum, London and Museum für Naturkunde, Berlin).

Cameras

The survey shows that there are a vast number of different cameras of varying specifications being used by the institutions, however a large majority of these are DSLRs (Digital Single-Lens Reflex cameras) manufactured by Canon (http://www.canon.com) or Nikon (http://www.nikon.com). Only a few models are used by more than one institution; these include Canon EOS 600D, Canon EOS 5D Mark II, Canon EOS 5D and Nikon D-200. These are all high quality semi professional models from 2005 to 2011 and (with the exception of the Nikon D200) are at the higher end of the megapixel specifications of the cameras used. In the USA Canon DSLR was found to be the most common brand of camera used to digitise natural history collections although Nikon DSLR cameras were also in use (Nelson, 2012).

Camera equipment providing images of the highest resolution (36MP and above) are owned by institutions that house mainly botanical collections and illustrations. The LeafAptus-II 10 (http://www.mamiyaleaf.com) and the Phase one iXR (http://www.phaseone.com) are both used by two different institutions. It is possible that institutions that house other collections

e.g. zoology, palaeontology or archaeology collections, need to branch out into other more appropriate types of digitisation equipment when wanting to work at a resolution higher than ~24MP, such as 3D imaging or X-ray equipment etc. which would allow bone topography, morphologies, structures etc. to be captured with more useful information. Botanical specimens (with the exception of carpological material) tend to be mounted flat onto a sheet, which means useful information and measurements can be captured from 2D images. Institutions housing entomological collections tend to store them in large numbers within trays, and after capturing an overall image of the tray, concentrate on each specimen individually, which tend to be of a small size, allowing a higher PPI image to be captured with the use of a macro lens if required. Some institutions have also favoured the scanning camera approach for such collections. Scanning cameras utilise a line CCD sensor that is moved mechanically behind the lens and exposed line by line, thus allowing higher resolution images to be captured, such as the Pentacon series.

Only one institution listed the use of a film camera in addition to digital equipment, with all others listed only digital equipment. This institution uses a great variety of equipment to digitise their collections, including several models of film and slide scanners.

Institutions that do not house botanical specimens tend to use entry level to semi-professional level cameras and did not list camera equipment at the highest level, however these were also the institutions that most of the 3D scanner and X-ray equipment belonged to, which supports the above suggestion that alternative imaging equipment to digital cameras are used for 3D specimens when the quality of the required image exceeds a specified threshold.

Three institutions use SatScan in order to image entomological collections, one institution also uses it for imaging slides. SatScan comprises of a camera that is moved in two dimensions along precision rails positioned above the object being imaged. A number of images are taken and then stitched together to create a larger image. This approach aims to minimize distortion of images whilst capturing large areas but also imaging small specimens at very high resolutions (Blagoderov et al., 2010). The SatScan has been successfully used to image whole drawers containing entomological specimens (Mantle et al., 2012; Blagoderov et al., 2012). Alternative technology that can be used include the GigaPan system) and the DScan. The GigaPan system (www.gigapan.com) consists of a robot that can be fitted with a digital camera and mounted on tripods threads. The robot positions the camera to frame individual specimens and a remote release can be used to engage the camera, which captures overlapping tiles, Gigapan software is then used to stitch the image into one large panorama (Bertone and Deans, 2010; Bertone et al., 2012). The DScan consists of a consumer digital SLR camera with a photographic macro lens attached to the linear units used by computer numerical control positioning machines and controlled with proprietary software (Schmidt et al. 2012). However, no institution in the questionnaire reported using these systems. A review of current progress in whole-drawer imaging of insect collections, and the advantages and disadvantages of different methods used was recently published by Holovachov et al. 2014.

Scanners

Scanners were used mainly for 2D material such as standard herbarium specimen sheets, labels, documents (including archives), maps, illustrations, wood trunk sections and books.

There appears to be more duplication of models between institutions for scanners than for other imaging equipment. Three institutions use the Pentacon Scan (http://scanner.pentacon.de/en/home.html) scanning cameras (one using Pentacon Scan 5000, another using the 6000 model and the third using both). These scanning cameras are used by institutions mostly for the same function as a flatbed scanner, imaging 2D material at high resolutions. Of the flatbed scanners used, only two models, both Epson, are used by more than one institution. The Epson Perfection V750 PRO is an A4 scanner used by two institutions in order to image herbaria material and other 2D material. The other model, the Epson Expression 10,000XL, is the piece of equipment amongst all the imaging solutions listed from the questionnaire that is used by the most number of institutions and was listed by 6 different partners. All of these partners house botanical collections and it is likely that the scanners are used in conjunction with the HerbScan imaging solution – a system specifically designed to enable the high resolution digitisation of herbarium material without having to invert the specimens. The HerbScan system is the imaging solution recommended by JSTOR (http://www.jstor.org/) for use by Global Plant Initiative (GPI, http://gpi.myspecies.info/) partners when imaging their Herbarium type material. GPI specifications require that specimens be scanned at 600 PPI resolution, beyond the capacity of most DSLR cameras when used for whole sheet images of herbarium specimens (http://about.jstor.org/sites/default/files/misc/plants_hndbk_eng_2011.pdf). Four of the institutions using this equipment are GPI partners and so this could account for the popularity of this scanning system.

Book scanners

All book scanners that are used by institutions are used primarily for imaging books, illustrations and in some cases herbarium material, where the material has been bound or is suitably flat. This could be due to the nature of many book scanners, which often have a glass plate to hold the book in place or a book cradle to support the spine of a book, rendering them unsuitable for any material that is not 2D, bound and/or flat. Of the book scanners used, there is one model that is used by two separate institutions, Minolta PS7000 (http://www.konicaminolta.com/). This scanner can image up to 600 PPI resolution and is used for imaging books and/or maps and photos. Two institutions use the Atiz bookdrive (but different models) scanner to image their books and bound herbarium specimens (http://www.atiz.com/). This system consists of an imaging setup using two DSLR's to image the opposite pages of a book at an angle in order to get a straight shot of the page whilst using a v-shaped book cradle, which in turn helps reduce stress on the spine of the book. The system allows for the cameras to be upgraded as the technology progresses.

3D scanners

A limited number of the partners have 3D scanning equipment which are used for bulkier collections. Museum für Naturkunde (MfN, Berlin) uses the Breuckmann Smartscan Duo system in order to scan their Paleontological collections (http://www.aicon3d.de)This solution uses two high resolution cameras with CCD image sensors capturing images of white light projections on the objects (http://species-id.net/o/media/b/ba/Manual_Smart_Scan_Duo_3D.pdf). The Royal Belgian Institute of Natural Sciences (RBINS, Brussels) reports using the low cost NextEngine laser scanner, http://www.nextengine.com/, MechScan (http://www.mechscan.co.uk/) and High Definition Imaging (HDI) equipment (http://www.lmi3d.com/products/hdi/). The Royal Museum of

Central Africa (RMCA, Tervuren) also reports using a HDI system. These two institutions are part of The Agora 3D consortium (http://agora3d.africamuseum.be/) composed of four Belgian federal scientific institutions which also includes the Royal Museums of Art and History and The Royal Institute for Cultural Heritage. This project aims to evaluate different 3D applications available on the market with special interest for Open Source technologies; from CT (Computed Tomography), to photogrammetry, surface scanning and MRI (Magnetic Resonance Imaging).

Slizewski A., et al.,(2010) tested three surface scanning systems: the low cost NextEngine laser scanner, the white light fringe projection Breuckmann Smartscan and the white light Fringe Projection Steinbichler COMET V 4M to evaluate the potential of such systems for digitising anthropological specimens comparing it with a “nominal” 3D model derived from µCT or CT data. In their tests, Breuckmann Smartscan produced the best models with the lowest deviation compared to the nominal µCTmodel. The Steinbichler was the fastest system but the quality of the resulting models was slightly lower. NextEngine was clearly lower quality than the tested high end systems but the ratio between the cost and the result was extremely favourable.

Electron microscopes

There were no models that were used by more than one institution. The complete list can be seen in Appendix 1.

Other equipment

Of the other equipment listed by the institutions, X-ray equipment accounted for approximately half of it. Three institutions use X-ray equipment in order to image the morphology of Zoological collections along with Paleontological and Archaeological collections. All institutions use different X-ray equipment, possibly due to the variety of needs, such as imaging the stomach contents of snakes (Hellenic Center for Marine Research HCMR) to non-invasive 3D scanning of Paleontological specimens (Swedish Museum of Natural History). Other equipment listed by the partners were microscope cameras.

Equipment Research

The recourse taken by institutions before purchasing imaging equipment included market research (5 institutions), advice from imaging and photographic experts (2 institutions, both imaging herbarium specimens) as well as from other institutions. On site testing and demonstrations were carried out by 5 institutions. Some projects, such as the Agora 3D project, permitted institutions to research and identify appropriate technologies, as part of the project scope.

Use of Imaging Equipment

Only three out of the fourteen institutions imposed no restrictions on which staff could have the use of their imaging equipment. Only one of these mentioned commercial and academic customers being permitted to use their scanner. In most other institutions, only specially trained staff are allowed to operate the equipment.

IMAGING

Imaging Standards

Most institutions do not follow any official imaging standards and it was clear from the responses given that overall there was a general lack of awareness about these. Of the options listed, 2 institutions use the Metamorfoze preservation imaging guidelines (http://www.metamorfoze.nl/english/digitization) but none of the institutions uses the Federal Agencies Digitization Guidelines Initiative (FADGI http://www.digitizationguidelines.gov/) Most institutions defined their own standards for imaging by looking at levels of detail and resolution etc., while some use other standards such as those listed by the International Association of Wood Anatomists (IAWA) and those required for the Global Plants Initiative.(http://about.jstor.org/sites/default/files/misc/plants_hndbk_eng_2011.pdf). There was a lack of automated software check of images to meet predefined standards.

Standards for mandatory Image components in images seem to vary across collections.

Botanical Images

All institutions who responded include a scale bar in the images, nearly all include a colour chart and most include their institution logo. Imaging the contents of capsules, paper packets found on Herbarium sheets, was only reported as being mandatory by 3 institutions and one of these reported that it was only mandatory for specific projects. Only 5 out of 9 institutions say that label information is mandatory to be captured within their images. 6 out of 9 institutions require a barcode in every image. One partner requires an image stamp in the image.

Nearly all partners with botanical collections don't crop their images down to individual specimens but image the whole sheet with all specimens clearly visible. Around half of these institutions keep just one image and produce several database entries and the other half save an image for each of the specimens. Royal Botanic Gardens Edinburgh reported sometimes cropping images of bryophytes, lichen or fungi to create separate images showing a single specimen and Royal Botanic Gardens cropped images of their Economic botany collection but not for standard herbarium specimen sheets.

Zoological Images

With the exception of Naturalis that required a larger number of components, all institutions required scale bars as obligatory components. All but one required only the label data as the other necessary component, but the one partner that didn't list label data did require a colour chart. Naturalis did state that the component criteria were dependent upon the specific requirements associated with a particular specimen type.

Paleontological Components

All institutions recorded that a scale bar and the label data needed to be recorded in the images (one partner stated label data was only sometimes needed). Approximately half of the partners with paleontological collections also require the institution logo in their images

including Naturalis who also require a colour chart and a barcode. One of these partners states that the logo is only needed if the image is intended for external use.

Geological Components

Not much information was provided on the geological collections but it appears that, like in other collections, the scale bar is important, as is label data. One institution also requires a colour chart and greyscale.

Entomological Components

Label data appears to have replaced scale bars as the most important component for entomological specimens. This may be due to the nature in which most of the entomological specimens are imaged: in trays and then separately in a lot of cases, so scale bars may not fit in all images. However, label data may be the easiest way to tell the specimens apart, when the images are not of high enough resolution for identification. Barcodes also are frequently recorded as necessary components (5 out of 8).

Image File Naming Conventions

The majority of image files created during digitisation follow a naming convention that uses the acquisition/ accession number or a catalogue number - this enables the use of a unique identifier that can link the image to a record. Some collections have barcodes applied to the individual specimens or to the drawers that the specimens are housed in and any files resulting from imaging are named using the corresponding barcode number. Some of the more complex imaging techniques, such as X-ray and 3D scanning, require a more detailed naming methodology, using collection numbers or barcodes to identify the specimen, followed by suffixes of scan number, version number etc. where several different scans and image reconstructions are created.

Botanical Filename Formats and Resolution

The majority of files are saved in TIFF format (approximately 200-250 MB) and also in smaller sized JPEG format. Some partners use DNG format and bitmap images. The majority of institutions followed the resolution requirements and predetermined standards set for the Global Plants Initiative project, imaging all herbarium specimens at approximately 600 pixels per inch (PPI). However, it is known those institutions that have outsourced imaging of their herbarium sheets have implemented a resolution of 300 PPI. Lower resolutions have been reported for the imaging of cryptogam specimens 72 or 240 PPI, this may be because higher resolution images are not so useful for identification purposes. Wood specimens have also been imaged at the lower resolution of 300 PPI, in these instances a number of images need to be taken and the decision was made on a balance between resolutions and file size for image storage which still meets the needs of the end users. There are a couple of different naming protocols that are followed but the most common appears to be naming after the catalogue or acquisition number. One similar approach to this is using a barcode which is entered into the database as a unique identifier. One partner uses the collection number followed by a number for each scan.

Zoological Filenames and Formats

Zoological images are almost exclusively stored in TIFF and JPEG format. Some institutions store their images in other file formats – one converting TIFF to PNG for storage reasons and another also keeps the images in other formats such as OBJ, PLY, STL etc. after the JPEGs have been processed in order to create 3D images. Unlike the botany collections, one of the most common file naming approaches is to begin with the species name followed by the catalogue number and acquisition number. The resolution of the images is very variable, possibly due to the range in sizes of specimens unlike botanical collections which mainly consist of standard sized herbarium sheets, however it seems to be of lower pixel density than for images of botanical specimens. Two Institutions reported imaging at the highest resolution that is achievable by the equipment.

Paleontological Filenames and Formats

The image format used is mainly TIFF and JPEG with images saved under the catalogue/ acquisition number. Institutions imaged at different resolutions, 400 PPI, 300 PPI and 180 PPI. One partner imaged at 300 PPI but stated that this is a limitation of the equipment and if specimens could be scanned instead then they could achieve 2400 PPI but that a pixel density of at least 600 PPI would be desirable. One institution is imaging at 300 PPI only if the image is to be published, otherwise 180 PPI is used.

Geological Filenames and Formats

The geology specimen images are usually saved in JPEG format although one institution saves in TIFF and another saves the images in the RAW scanner filename format and then processes the images into other file formats when creating 3D images. There is very little information reported on the resolution of the images created although National Museum Prague report imaging at 3000 PPI.

Entomological Filenames and Formats

Specimen images are mostly saved in JPEG and TIFF formats. A variety of equipment is used, even within the same institution – possibly due to the different ways in which the specimens are mounted. The images tend to be named following the species names, although some institutions add a unique ID to the end of the filename. Resolutions were given in a mix of PPI, MP (mega pixels) and pixel dimensions. Images are produced at a maximum resolution of 300 PPI but this is probably because they are imaged with cameras rather than scanners so resolution is possibly reduced.

Anthropological Filenames and Formats

Specimen images are saved in TIFF and JPEG formats. Raw images are created but are deleted after stacking. There was not much data available on size, one institution reports 12 MP but stacks images so the images may be of higher resolution (although interpolated) by the end of the process. National Museum Prague state a resolution of 2000 PPI. File names used are the acquisition or identifying number, one partner uses species name for human remains specimens.

Other Collection Filenames and Formats

Less information was reported for other collections, Archaeological scientific drawings were scanned at 1200 PPI and Mycological specimens were imaged at 300 PPI and 600 PPI.

WORKFLOWS

From the results of the questionnaire, we were able to see that the most common order in which tasks are performed (were all the tasks to be performed) across collections are as follows: selection; transfer of material from one area to another; application of barcodes and "other" tasks; full (or partial) data capture, imaging; records management; returning material; and QA (Quality Assurance).

The majority of institutions are still capturing specimen metadata prior to the imaging step in their main digitisation workflows. Only one institution reported the imaging of specimens prior to any specimen metadata capture. Another institution reported capturing partial specimen metadata prior to imaging but then capturing full specimen metadata in the last step of the digitisation process. We believe this latter workflow is more prevalent than suggested in the questionnaire results as we are aware of trials of this workflow in other institutions, including within Royal Botanic Gardens Kew although it is currently not the main digitisation workflow within that institution.

The vast majority of institutions included a full data capture step within their digitisation workflows with only two institutions reporting the inclusion of only a partial data capture step. Only four institutions reported a two step data capture process of partial data capture followed by full data capture at a later point within the workflow.

The most frequent collection type to appear in this survey was botanical. Of all the collection types, this was the one with the most similarity between institutions in terms of the steps carried out and the order in which these steps were performed. This likeness between workflows could be explained by the similarity in equipment used to digitise the specimens and the layout of the collections, but also 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.

It is important to note that other collection types were present in fewer institutions than were botanical collections and this may account for the fact that other collection types appeared to share fewer of the steps carried out between institutions holding that same collection type, as well as in the ordering of the workflow steps. However, the two anthropological collections showed remarkable similarity between their workflows, differing only on the data capture step, one institution capturing full and the other, only partial data at this step. This could be explained by their use of large equipment and long imaging times, which may require the workflow to work around these factors, limiting the number of permutations in the steps.

Data Entry and Management

Metadata Collected at Each Step

The amount of metadata captured ranges greatly between the institutions and their collections but the majority of institutions differentiated between the data recorded at the point of image capture and the data taken from the specimen labels. At the point of image

capture, the data that is collected generally includes information regarding the types of equipment, resolution, filename and other information regarding the date and user. Where the equipment being used produces large numbers of images and complex files (e.g. X-ray) a large amount of other data is recorded such as stitching versions of images, different angles, measurements etc. In some cases the information is recorded within the filename, but often additional documents are used to record this data. Specimen label data varies between collections but core information captured includes taxonomic name, country, collector and collector number.

Methods to Collate Collection Data

All institutions were entering some if not all of the specimen metadata in-house. Two institutions reported using trained volunteers in-house, one reported using undergraduate students and one institution reported using overseas collaborators to capture metadata. Only three institutions were currently investigating using crowdsourcing options for microscopic slides and herbarium specimen collections. Two institutions included Optical Character Recognition (OCR) in their workflow; one of these uses it for digitisation of literature only.

Only Royal Botanic Gardens Edinburgh reported using OCR in the digitisation of natural history collections for Herbarium specimens. Drinkwater R. et al., 2014 reported that when compared to an unsorted, random set of specimens, those which were sorted based on data added from the OCR were quicker to digitise. Of the methods tested, the most successful in terms of efficiency used a protocol which required entering data into a limited set of fields and where the records were filtered by Collector and Country. The survey and subsequent discussions with the digitisation staff highlighted their preference for working with sorted specimens, in which label layout, locations and handwriting are likely to be similar, and so a familiarity with the Collector or Country is rapidly established. The use of OCR for specimen metadata capture appears to be more common in institutions in the USA (Nelson, 2012).

Outsourcing

Less than a third of the institutions have even considered outsourcing the digitisation of their collections, with only one institution having implemented this process. One partner stated that outsourcing had not been considered as digitisation was used to answer specific research questions and the aim was not to mass digitise all of the institution’s specimens.

The advantages in outsourcing that were cited included probable faster rates of digitisation and a reduced cost per specimen, alleviating the issue of limited space for digitisation stations within the organisation and making use of industrial process experience and project management knowledge held by contractors. However the largest reasons for choosing not to outsource the digitisation of collections was the high risk of damage to fragile specimens through transport, pests and inexperienced non specialised staff handling the specimens. This was closely followed by concerns about the quality of the data that would result from non-specialist curators and staff undertaking the digitisation work who would be inexperienced in interpreting specimen labels that are often handwritten and difficult to read requiring specialist domain knowledge. Any savings due to reduced digitisation costs could be negated due to the amount of supervision, quality checking and error correction that might be needed leading to a duplication of effort. Other disadvantages mentioned were the logistics in moving huge numbers of specimens to a different location including the unknown

legal issues that might arise or even prevent specimens being transported to the outsourcing suppliers and that current levels of pre-digitisation conservation work would be impossible to maintain under very high rates of digitisation.

Geographic Information Systems (GIS)

GIS is carried out in approximately 50% of the digitisation workflows. However, in general only a proportion of the specimens are routinely georeferenced depending on priority. In some cases coordinates are only recorded if they are found on the specimen labels. Two institutions report developing software to include a georeferencing step, one indicating that this step is highly automated and has achieved good results.

Quality Assurance (QA)

Data:

Quality Assurance (QA) on data is performed by approx. 50% of the institutions. Where QA is performed, the majority of the time it is done by a separate individual to the one that captured the data, either by dedicated QA officers, curators of the collections or a second person who checked and augmented the records. One institution reported that the curators of the collections directly supervise the data collection process in order to improve quality.

Images:

Eleven out of the thirteen institutions said they performed some level of quality assurance on their images and checked at least one of the following aspects: presence of all necessary components; completeness of specimen represented (whether additional images of the same specimen were required); level of visible details; quality of stitching; legibility of QR codes; focus; cropping; filename and metadata. However, the level of QA performed varied greatly between institutions, with only two institutions specifying that they checked all the images produced and others checking 10% of images, following the Global Plants Initiative standards, checking images for focus and artefacts. Image QA is mostly carried out by people other than the staff who had imaged the specimen.

Almost all QA is carried out by staff performing a visual check on the images produced with none or very little automatic software checking. Royal Botanic Gardens Kew convert images taken on the HerbScan system to GIF format where pixilation or other image artifacts can be more easily spotted visually in the black and white image. One institution reported the use of 3D inspection and mesh processing software GOM Inspect (http://www.gom.com/3d-software/gom-inspect.html) for checking their 3D images.

Data Storage and Access

Nearly all of the institutions hold their data on internal in-house servers. Only two institutions hold their images on hard drives, one of which is set to move to a server soon. Five institutions using a server also back up their images onto a hard drive or keep copies on computers.

All participants, except one, are making their images available on the internet, the majority through their institutional website. Around half of these are also making their images available through other portals such as The Global Biodiversity Information Facility (GBIF) (3), Europeana (3), BRAHMS online (3) and JSTOR Global Plants (2).

Main Users of Digital Collections and How These are Utilized

As expected, the main users stated were internal and external researchers, including students and collection managers. Other users included the general public, and digital collections being used for exhibitions, publications and some artist requests.

Licensing

The majority of partners implement licensing through one of the following creative commons licences: CC-BY-NC-SA, CC-BY-SA, CC-BY-NC or CC-BY. Only one institution uses CC-0 for data and one, open access for images. Two institutions implement their own in-house policies and agreements. Within the OpenUp project (http://open-up.eu/) a report was written on IPR (Intellectual Property Rights) problems and solutions for the domain of natural history which included a survey of the licensing agreements under which partners routinely shared data and content with third parties (http://open-up.eu/content/deliverables-and-components-pu)

Digital Collections Curation

All institutions responded to say that they curate their digital collections, however from the details given it seems that the curation undertaken may be limited. Cross-checking, quality checking and data cleaning tasks were reported as well as backup of data and images. For botanical specimens some institutions rescanned specimens where new annotations had been added and/or updated the specimen record. It was unclear for some institutions what was meant by digital collections curation and the question should have been further defined in the questionnaire.

Successes, Challenges and Future Developments

Limitations in Digitisation Rates

The factor most often stated as a limitation for digitisation rates was lack of human resources (13 institutions) closely followed by funding (11 institutions). These two factors are obviously closely linked as generally obtaining more funding allows the recruitment of more digitisation staff. Of those partners that ranked the limitation factors, 58% ranked human resources highest and 33% funding.

Other limitation factors quoted included lack of an adequate data storage solution (6), technology (6), equipment (5), physical workspace (4) and collection handling (2).

One institution stated that the use of an industrial book scanner (or equivalent) would speed up imaging of vascular plants. This institution was using a HerbScan which can produce extremely high resolution images but has slow scanning speeds. Other institutions imaging botanical collections have converted to using high resolution cameras to increase digitisation rates.

Two additional limitations were given that were not listed in the questionnaire. These were: keeping the physical and digital collections in synchronisation and difficulties managing a highly diverse ecosystem of data management and digitisation workflows. Some institutions reported not having a central digitisation unit with many different departments developing their own solutions and workflows. This would make it difficult to develop a coherent digitisation strategy.

Strategies to Increase Rates

The strategies reported as used to increase rates were quite varied. However, a common strategy was to put in efforts to acquire more staff, either through short term funded projects or new full-time posts. Another trend was the limitation in digital infrastructure with lack of efficient data storage and content management solutions. Many institutions reported the need to continuously improve their databases and workflows. Other specific examples include addition of an OCR step to aid data capture, changes from scanners to digital cameras to increase imaging rates, specimen data capture by collaborators in other institutions who have knowledge of a relevant geographical area and standardising processes to define rules about which data and images are captured. Also mentioned was increasing the skills and knowledge of staff through training. Only one institution mentioned that rates might be improved by investigation of digitisation technologies.

Summary

Responses to the questionnaire indicated the following main findings:

- Regarding collections, nearly all institutions house more than one type of collection, with the average number varying between five and six per institution. The size of each collection varied greatly from 15,000 specimens up to 80 million, with an average collection size of around 6 million. The prioritisation for collection digitisation was driven by four main factors: digitisation of type specimens; discrete project-based funding to digitise a specific subset of the collections; research needs of the institution; and loan or image requests.

- In response to questions on pre-digitisation curation, approximately two thirds of the institutions performed at least some minimal curation, specimen conservation or collections management steps prior to digitisation. The most common pre-digitisation step was to check the identification of the specimens to be digitised or to check the type status.

- There is a vast range of digitisation equipment in use, the complete list can found in Appendix 1, reflecting the wide variety of collections in the different institutions. Most

equipment and models were only used at one institution; however the large majority of cameras were DSLRs manufactured by Canon or Nikon.

- Cameras used at more than one institution included, the Canon EOS 600D, Canon EOS 5D Mark II, Canon EOS 5D, Nikon D-200, LeafAptus-II series, Phase one iXR and the SatScan System.

- There appears to be more duplication of models between institutions for scanners than for other imaging equipment. Those Scanners used by more than one institution include the Pentacon Scan, Epson Perfection V750 PRO and the Epson Expression 10,000XL.

- The Epson Expression 10,000XL is the piece of equipment amongst all the imaging solutions listed from the questionnaire that is used by the most number of institutions and was listed by five different partners. All of these partners house botanical collections and it is likely that the scanners are used in conjunction with the HerbScan imaging solution recommended by JSTOR for use by Global Plants Initiative (GPI) partners when imaging their Herbarium type material. GPI specifications require that specimens be scanned at 600 PPI resolution, beyond the capacity of most DSLR cameras when used for whole sheet images of herbarium specimens.

- Book scanners used by more than one institution include the Minolta PS7000 and the Atiz bookdrive.

- Most institutions did not appear to follow any official imaging standard and it appeared that their awareness of such standards was under developed. Institutions defined their own standards or followed project driven standards for imaging e.g. those developed for the GPI project. There was a lack of automatic software checking of images the majority of QA was undertaken visually checking the images for at least one of the following aspects: presence of all necessary components; completeness of specimen represented (whether additional images of the same specimen were required); level of visible details; quality of stitching; legibility of QR codes; focus and cropping.

- The majority of institutions are still capturing full specimen metadata prior to the imaging step in their main digitisation workflows. Only one institution reported imaging specimens prior to any specimen metadata capture and only two institutions reported capturing partial metadata only. Four institutions reported a two step data capture process of partial data capture followed by full data capture at a later point within the workflow.

- The main reported factor limiting the rates of digitisation was not related to equipment but to the challenge in securing funding for digitisation. Currently there is a lack of available human resource within many institutions with natural history collections to progress with digitisation programmes.

- Suitable digitisation infrastructure was also seen as a greater impediment to digitisation rates than the lack of suitable digitisation equipment. It is clear from the questionnaire that Institutions with natural history collections require knowledge and assistance not

only with the digitisation workflows themselves but with the management of the data and images that are created. As digitisation rates increase managing the larger volume of digital data created becomes more complex.

- Most institutions still perform all digitisation in-house and there is a reluctance from many institutions to consider outsourcing due to the increased risks of damage to specimens by staff who are untrained and lacking curation expertise. There is also a widespread assumption that the quality of data delivered will be low and that any cost saving will be outweighed by an increase in quality assurance and data cleaning tasks that will be needed.

- Nearly all of the institutions hold their data on internal in-house servers, and make majority images available on the internet. The majority of partners implement licensing through one of the following Creative Commons licences: CC-BY-NC-SA, CC-BY-SA, CC-BY-NC or CC-BY.

Key Recommendations

This report acknowledges the wide variety of approaches, requirements and equipment needs for natural history specimen collection digitisation, recognising that "a one size fits all" approach is not a viable solution. However, a number of common themes were present throughout all of the responses to the questionnaire, which has helped formulate the following key recommendations.

- Grant Resources – In order to assist institutions there should be a central repository listing possible funding bodies which would consider funding the digitisation of museum collections. Additionally, strategies for successful dissemination of project results (e.g. papers, publications, conferences, posters, technical briefs, stake-holder engagement) may help towards raising awareness of the importance of digitisation and therefore may have a beneficial effect upon the availability of securing additional funded work.

- Digitisation Resources Repository – as indicated by the questionnaire there are many different types of equipment and workflows that are currently in place within the various institutions. In the USA, the Integrated Digitized Biocollections (iDigBio), is coordinating the National digitisation effort through the Resource for Advancing Digitization of Biodiversity Collections (ADBC) program funded by the National Science Foundation. As part of their activities, iDigBio (www.idigbio.org) is collating together a wide range of resources for digitisation including example digitisation protocols, imaging documents and resources, imaging station equipment and specifications and database resources and tools (https://www.idigbio.org/wiki/index.php/Digitization_Resources). We recommend that institutions in the EU also share their information and workflows in a shared repository so that lessons learned are shared throughout the EU. Staff of iDigBio carried out a more comprehensive study of 28 digitisation programs in 10 museums and academic institutions, an initial questionnaire was followed up by on-site visits to observe the workflows. This enabled more workflow specific and detailed recommendations to be put forward (Nelson, 2012).

- Most institutions did not appear to follow any official imaging standard and it appeared that there awareness of such standards was under developed. It is therefore a recommendation from the current survey that a set of harmonised guidelines and standards be agreed upon by the relevant experts involved in digitisation, in order to set a bench-mark for quality assurance of digitisation procedures. The majority of QA checks were undertaken visually and there may be potential to implement image checks against image standards using automated software to speed up the process of QA.

- The majority of institutions are still following an in-house digitisation workflow of capturing full specimen metadata first followed by imaging. This approach while effective is time consuming. Only a few institutions of those surveyed had implemented different alternative approaches e.g. Outsourcing, addition of an OCR step, capturing partial data first and full data later or crowdsourcing. We recommend that successful adaptations to workflows that increase digitisation rates are disseminated more widely so they are more likely to be implemented in other institutions.

References

Ariño AH (2010) Approaches to estimating the universe of natural History collections data. Biodiversity Informatics 7: 81-92.

Bertone M, Blinn R, Stanfield T, Dew K, Seltmann K, Deans A (2012) Results and insights from the NCSU Insect Museum GigaPan project. ZooKeys 209: 115-132. doi: 10.3897/zookeys.209.3083

Bertone M, Blinn R, Stanfield T, Dew K, Seltmann K, Deans A (2012) Results and insights from the NCSU Insect Museum GigaPan project. ZooKeys 209: 115-132. doi: 10.3897/zookeys.209.3083

Blagoderov, Vladimir , Kitching, Ian, Simonsen, Thomas, and Smith, Vincent. Report on trial of SatScan tray scanner system by SmartDrive Ltd.. Available from Nature Precedings <http://hdl.handle.net/10101/npre.2010.4486.1> (2010)

Blagoderov V, Kitching I, Livermore L, Simonsen T, Smith V (2012) No specimen left behind: industrial scale digitization of natural history collections. ZooKeys 209: 133-146. doi: 10.3897/zookeys.209.3178

Drinkwater R, Cubey R, Haston E (2014) The use of Optical Character Recognition (OCR) in the digitisation of herbarium specimen labels. PhytoKeys 38: 15-30. doi: 10.3897/phytokeys.38.7168

Holovachov, O, Zatushevsky, A and Shydlovsky, I 2014. Whole-Drawer Imaging of Entomological Collections: Benefits, Limitations and Alternative Applications. *Journal of Conservation and Museum Studies* 12(1):9, DOI: http://dx.doi.org/10.5334/jcms.1021218

Mantle B, LaSalle J, Fisher N (2012) Whole-drawer imaging for digital management and curation of a large entomological collection. ZooKeys 209: 147-163. doi: 10.3897/zookeys.209.3169

Nelson G, Paul D, Riccardi G, Mast A (2012) Five task clusters that enable efficient and effective digitization of biological collections. ZooKeys 209: 19-45. doi: 10.3897/zookeys.209.3135

Schmidt S, Balke M, Lafogler S (2012) DScan – a high-performance digital scanning system for entomological collections. ZooKeys 209: 183-191. doi: 10.3897/zookeys.209.3115

Slizewski A., Friess M. & Semal P. 2010. Surface scanning of anthropological specimens: nominal-actual comparison with low cost laser scanner and high end fringe light projection surface scanning systems. Quartär, 57: 179-187

Appendices

Appendix 1. Equipment List

Cameras

AxioCam MRc
Canon
Canon 1200D
Canon EOS 100D
Canon EOS 300D
Canon EOS 40D
Canon EOS 50D
Canon EOS 550D
Canon EOS 5D
Canon EOS 5D Mark II
Canon EOS 5D Mark3
Canon EOS 600D
Canon EOS 6D
Canon EOS 7D
Canon EOS-1 Ds Mark III
Canon EOS-1 Ds MarkX
Canon EOS-3 50D
Canon SX40 HS
Hasselblad 500C/M
Kodak DCS Pro SLR
Leaf Aptus-II 10
Leaf Aptus-II 12
Lenses: AF-S Nikkor 60mm (2 items), AF Micro Niccor 105 mm, Niccor zoom lens, Nexus macro lens, Nexus zoom 18-55mm, Canon 100mm, Canon 35mm macro. Filters: Polarisaton filters.
Lenses: Canon MP-E65/2.8 with 2xExtender EF 2xIII; Canon EF 100/2.8L
Nexus 7
Nikon
Nikon D1X
Nikon D200
Nikon D3
Nikon D300
Nikon D4
Nikon D50
Nikon D5200
Nikon D60
Nikon D70
Nikon D700
Nikon D800E
Nikon D90

Nikon digital sight DS-2Mv
Nikon/Canon
Olympus DP71
Olympus E-30
Olympus E-330
Panasonic Lumix
PhaseOne iXR
SmartDrive SatScan

Scanners
Pentacon Scan 5000
Pentacon Scan 6000
Canonscan 9000F
Epson Expression 10,000XL
Epson A4
Epson GT 10000
EPSON Perfection V750 PRO
Fujitsu fi6770 A3
Ricoh copiers A0 scanner
UMAX power look A·3
WideTEK 25-200

BookScanners
SMA2 Book scanner
ProServe ScannTech 602i-3
Fujitsu scansnap SV600
EPSON Perfection V700 Photo A4
Atiz Bookdrive Mini
Minolta PS7000
CopyBook I2S
Zeutschel Zeta
Atiz BookDrive Pro
Zeutschel
I2S

SlideScanners
Reflecta Digitdia 5000
Coolscan 4000 35mm
Canon Mikrofilmscanner MS 300
Epson
Konica Minolta Dimage 5400
Nikon 5ED
Nikon LS 5000 ED
Nikon LS2000 &Nikon SF-200
Nikon Super Coolscan 8000 ED
Nikon Supercoolscan 9000

3D Scanners
MECHSCAN
SmartScan Duo

Nextengine
HDI Advance R3X
MicroScribe 6G2LX/MicroScan
HDI LMI

Electron & X-ray Imaging Equipment
Hitachi S-3700N
1 FEI SEM
FEIINSPECT
FEI QUANTA 200
JEOL JSM-6480LV scanning electron microscope
HyperProbe Electron Probe Microanalyzer (EPMA): Jeol JXA-8530F
Scanning Electron Microscope (SEM): Jeol JSM 5000,
Scanning Electron Microscope (SEM): Jeol JSM 6400
Scanning Electron Microscope (SEM): Jeol JSM 6610-LV
TEM LEO 912 AB
LEO Supra 55VP scanning electron microscope
Leica DC150
Leica DFC 490
Leica EC3
SEM Hitachi S-4300 FE
Zeiss EVO 15LS SEM
LEO 1455 VP SEM
Metris X-Tek HMX ST Computed Tomography System
Carl Zeiss Ultra Plus Field Emission Scanning Electron Microscope (SEM)
Gatan X-ray Ultra Microscope for nano-CT
Hitachi H-7100 Transmission Electron Microscope
Leica EM-KMR-2 Knifemaker TEM
Micro-CT scanner x-ray to provide mathematical representation of a 3D object
X-Ray machine VisiX with high resolution digital x-ray detector Dereo WA2 and DereoHR1
SkyScan 1172 (X-ray)

Optical Microscopy Equipment
Confocal Espectral LEICA TCS SPE
Zeiss AxioScan
Zeiss AxioZoom
Zeiss AxioImager M2 microscope with motorised z control and DIC optics

Appendix 2. Digital Collections Questionnaire

SYNTHESYS3 WP 3

Questionnaire on existing digitisation workflows

Prepared by:	Royal Botanic Gardens, Kew	**Date:**	August 2014

Purpose

The purpose of the questionnaire is to gather data from participants on their current digitisation facilities and needs, finding out how equipment has been used and the successes and challenges faced. For this purpose we are defining digitisation to include capture of specimen data, images and media. Where information is unknown or inapplicable please enter “Unknown” or “N/A” as relevant and move onto the following question.

This questionnaire is divided into the following sections:

Institution:		Date:	

1. General

A. i) What type/s of collections do you house in your institution and what is the approximate size of the collection/s? Please provide more information where necessary

	Collection	Size of Collections
☐	Botanical	
☐	Entomological	
☐	Zoological	
☐	Geological	
☐	Paleontological	
☐	Illustrations	
☐	Correspondence	
☐	Mycological	
☐	Other	

ii) How are your collections stored? e.g. compactors, spirit collection

B. i) What types of equipment are used to digitise your collections?

	Specifications (make and model)	Quantity
Cameras		
Flatbed Scanners		
Book Scanners		
Slide Scanners		
3D Scanners		
Electron Microscope		
Other		

ii) What software packages are used as part of the digitisation workflow and for what function?

Software	Function

C. What function does each section of imaging equipment have within the digitisation process? (e.g. Electron microscope used for imaging pollen etc.)

	Function
Cameras	
Flatbed Scanners	
Book Scanners	
Slide Scanners	
3D Scanners	
Electron Microscope	
Other	

D. What research if any was done into finding imaging solutions prior to purchase?

E. Is any curation, specimen conservation or collections management performed at the point of digitisation?

☐ **Yes** ☐ **No**

Please provide more details

F. **i) How many people work on digitising the collections?**

_____Full Time _____Part Time

ii) How many people have use of the imaging equipment do you have restrictions on who uses the imaging equipment.

iii) Do people work independently or as part of a team when digitising?

Please provide more details – indicating if different individuals have different roles within the workflow.

2. Collections

For each of the collections stated in part 1.A.i) please answer the following questions. N.B. Please use additional collection sheets at end of questionnaire if more than one collection stated.

A. Collection type

B. i) Has any of this collection been digitised?

☐ **Yes** ☐ **No**

ii) If yes, what proportion as a percentage of the collection has been completed? (Please provide more information where necessary, indicating what steps of the digitisation process i.e. data capture, imaging, georeferencing have been completed)

C. What is the current annual rate of digitization?

_____**Specimens per year**

D. How has the digitisation of the collection been prioritized so far?

E. Are there any pre-digitisation stages required in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

F. Are the digitisation facilities located close to the collection?

☐ **Yes** ☐ **No**

Please provide more details

G. What file format and size are the collection images?

☐ **Tif MB** ☐ **Jpeg MB** ☐ **Raw MB** ☐ **PNG MB**

☐ **Gif MB** ☐ **Jpeg 2000 MB** ☐ **Other MB**

Other

H. What naming conventions are used for the image files? e.g. acquisition number

I. i) At what resolution are the digital images produced.

_____ pixels per inch or _____ MP for collections ___" x __"

ii) How was this resolution decided upon and does it meet the needs of the users?

J. i) What, if any, components are mandatory within your images?

- ☐ **Scale bar**
- ☐ **Institution logo**
- ☐ **Colour chart**
- ☐ **Grayscale chart**
- ☐ **Capsule contents**
- ☐ **Barcode**
- ☐ **Label information**
- ☐ **Other (please give more detail)**

K. What is the imaging protocol followed where multiple specimens appear in one unit (e.g. more than one specimen on a herbarium sheet, more than one lichen on a rock, Insects in a tray etc.)

	please provide more details
Whole unit captured in one image, displaying all specimens	
Whole unit captured in several images, each image displaying all specimens but is named separately	
Whole unit captured in several images, each cropped to display one specimen	
Part of unit captured in one image, to display one specimen	
Other	

L. Please number the following workflow tasks in the order in which they are performed. Leave blank any that are not part of your workflow and add any additional tasks not listed under "Other"

	Selection of material
	Transfer of material between areas
	Barcoding
	Partial Data capture
	Full Data capture
	Imaging
	Record management
	Returning material
	Quality control
	Other (please detail)

M. What is the approximate time taken for the whole workflow from start to finish to image and capture the data from one specimen?

_____minutes per specimen

N. What is the approximate cost per specimen for digitisation from start to finish?

_____Euros per specimen (or specify other currency)

3. Quality Control

A. What, if any, imaging standards do you adhere to?
Please provide details of level etc where possible

- ☐ **Metamorfoze**
- ☐ **FADGI**
- ☐ **Other**
- ☐ **None**

B. Please provide details of any Quality Assurance performed on your images or data.

Images	
Data	
Other	

4. Data management

A. What metadata is captured at each point of the digitisation workflow?

B. What methods are used to collate collection data?

- ☐ **In-house data entry**
- ☐ **Crowdsourcing**
- ☐ **Optical Character Recognition (OCR)**
- ☐ **Other**

Please provide more details

C. Is any geo-referencing included in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

5. Data storage and Access

A. How do you store and access your images?

B. Do you make your images available? If Yes, through what medium?

☐ Yes ☐ No

C. How do you license your digital collections for the following?

Images for research	
Data for research	
Images for publication	

D. Is there ongoing curation on your digital collections?

☐ Yes ☐ No

Please provide more details

6. Successes, Challenges and Future developments

A. What, if any, are the fields that provide the greatest limitations in digitisation rates? (please rank stating 1 as the highest and add any additional points below)

☐	**Equipment**	
☐	**Human resource**	
☐	**Technology**	
☐	**Physical workspace**	
☐	**Funding**	
☐	**Collection handling**	
☐	**Data storage solutions**	
☐	**Other**	

Please provide more details where needed

B. Who are the main users of your digital collections and how are the collections currently utilized?

C. What strategies if any have been implemented to increase digitisation rates?
Please give detail of the two main strategies

D. i) Has the outsourcing of collection digitisation to an external company been implemented or considered?

- ☐ **Outsourcing considered**
- ☐ **Outsourcing implemented**
- ☐ **No**

ii) What advantages and challenges would be presented if outsourcing the digitisation of your collections?

Advantages	
Challenges	

Please enter any further information or comments below

Many thanks for taking the time to complete this questionnaire

Please use the following continuation sheets to add additional collections data

A. Collection type 2

B. i) Has any of this collection been digitised?

☐ **Yes** ☐ **No**

ii) If yes, what proportion as a percentage of the collection has been completed? (Please provide more information where necessary, indicating what steps of the digitisation process i.e. data capture, imaging, georeferencing have been completed)

C. What is the current annual rate of digitization?

_____Specimens per year

D. How has the digitisation of the collection been prioritized so far?

E. Are there any pre-digitisation stages required in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

F. Are the digitisation facilities located close to the collection?

☐ **Yes** ☐ **No**

Please provide more details

G. What file format and size are the collection images?

☐ **Tif MB** ☐ **Jpeg MB** ☐ **Raw MB** ☐ **PNG MB**

☐ **Gif MB** ☐ **Jpeg 2000 MB** ☐ **Other MB**

Other

H. What naming conventions are used for the image files? e.g. acquisition number

I. i) At what resolution are the digital images produced.

_____ pixels per inch or _____ MP for collections ___" x __"

ii) How was this resolution decided upon and does it meet the needs of the users?

J. i) What, if any, components are mandatory within your images?

- ☐ **Scale bar**
- ☐ **Institution logo**
- ☐ **Colour chart**
- ☐ **Grayscale chart**
- ☐ **Capsule contents**
- ☐ **Barcode**
- ☐ **Label information**
- ☐ **Other (please give more detail)**

K. What is the imaging protocol followed where multiple specimens appear in one unit (e.g. more than one specimen on a herbarium sheet, more than one lichen on a rock, Insects in a tray etc.)

	please provide more details
Whole unit captured in one image, displaying all specimens	
Whole unit captured in several images, each image displaying all specimens but is named separately	
Whole unit captured in several images, each cropped to display one specimen	
Part of unit captured in one image, to display one specimen	
Other	

L. Please number the following workflow tasks in the order in which they are performed. Leave blank any that are not part of your workflow and add any additional tasks not listed under "Other"

	Selection of material
	Transfer of material between areas
	Barcoding
	Partial Data capture
	Full Data capture
	Imaging
	Record management
	Returning material
	Quality control
	Other (please detail)

M. What is the approximate time taken for the whole workflow from start to finish to image and capture the data from one specimen?

_____minutes per specimen

N. What is the approximate cost per specimen for digitisation from start to finish?

_____Euros per specimen (or specify other currency)

A. Collection type 3

B. i) Has any of this collection been digitised?

☐ **Yes** ☐ **No**

ii) If yes, what proportion as a percentage of the collection has been completed? (Please provide more information where necessary, indicating what steps of the digitisation process i.e. data capture, imaging, georeferencing have been completed)

C. What is the current annual rate of digitization?

_____**Specimens per year**

D. How has the digitisation of the collection been prioritized so far?

E. Are there any pre-digitisation stages required in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

F. Are the digitisation facilities located close to the collection?

☐ **Yes** ☐ **No**

Please provide more details

G. What file format and size are the collection images?

☐ **Tif MB** ☐ **Jpeg MB** ☐ **Raw MB** ☐ **PNG MB**

☐ **Gif MB** ☐ **Jpeg 2000 MB** ☐ **Other MB**

Other

H. What naming conventions are used for the image files? e.g. acquisition number

I. i) At what resolution are the digital images produced.

_____ pixels per inch or _____ MP for collections ___" x __"

ii) How was this resolution decided upon and does it meet the needs of the users?

J. i) What, if any, components are mandatory within your images?

- ☐ **Scale bar**
- ☐ **Institution logo**
- ☐ **Colour chart**
- ☐ **Grayscale chart**
- ☐ **Capsule contents**
- ☐ **Barcode**
- ☐ **Label information**
- ☐ **Other (please give more detail)**

K. What is the imaging protocol followed where multiple specimens appear in one unit (e.g. more than one specimen on a herbarium sheet, more than one lichen on a rock, Insects in a tray etc.)

	please provide more details
Whole unit captured in one image, displaying all specimens	
Whole unit captured in several images, each image displaying all specimens but is named separately	
Whole unit captured in several images, each cropped to display one specimen	
Part of unit captured in one image, to display one specimen	
Other	

L. Please number the following workflow tasks in the order in which they are performed. Leave blank any that are not part of your workflow and add any additional tasks not listed under "Other"

	Selection of material
	Transfer of material between areas
	Barcoding
	Partial Data capture
	Full Data capture
	Imaging
	Record management
	Returning material
	Quality control
	Other (please detail)

M. What is the approximate time taken for the whole workflow from start to finish to image and capture the data from one specimen?

_____minutes per specimen

N. What is the approximate cost per specimen for digitisation from start to finish?

_____Euros per specimen (or specify other currency)

A. Collection type 4

B. i) Has any of this collection been digitised?

☐ **Yes** ☐ **No**

ii) If yes, what proportion as a percentage of the collection has been completed? (Please provide more information where necessary, indicating what steps of the digitisation process i.e. data capture, imaging, georeferencing have been completed)

C. What is the current annual rate of digitization?

_____**Specimens per year**

D. How has the digitisation of the collection been prioritized so far?

E. Are there any pre-digitisation stages required in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

F. Are the digitisation facilities located close to the collection?

☐ **Yes** ☐ **No**

Please provide more details

G. What file format and size are the collection images?

☐ **Tif MB** ☐ **Jpeg MB** ☐ **Raw MB** ☐ **PNG MB**

☐ **Gif MB** ☐ **Jpeg 2000 MB** ☐ **Other MB**

Other

H. What naming conventions are used for the image files? e.g. acquisition number

I. i) At what resolution are the digital images produced.

_____ pixels per inch or _____ MP for collections ___" x __"

ii) How was this resolution decided upon and does it meet the needs of the users?

J. i) What, if any, components are mandatory within your images?

- ☐ **Scale bar**
- ☐ **Institution logo**
- ☐ **Colour chart**
- ☐ **Grayscale chart**
- ☐ **Capsule contents**
- ☐ **Barcode**
- ☐ **Label information**
- ☐ **Other (please give more detail)**

K. What is the imaging protocol followed where multiple specimens appear in one unit (e.g. more than one specimen on a herbarium sheet, more than one lichen on a rock, Insects in a tray etc.)

	please provide more details
Whole unit captured in one image, displaying all specimens	
Whole unit captured in several images, each image displaying all specimens but is named separately	
Whole unit captured in several images, each cropped to display one specimen	
Part of unit captured in one image, to display one specimen	
Other	

L. Please number the following workflow tasks in the order in which they are performed. Leave blank any that are not part of your workflow and add any additional tasks not listed under "Other"

	Selection of material
	Transfer of material between areas
	Barcoding
	Partial Data capture
	Full Data capture
	Imaging
	Record management
	Returning material
	Quality control
	Other (please detail)

M. What is the approximate time taken for the whole workflow from start to finish to image and capture the data from one specimen?

_____minutes per specimen

N. What is the approximate cost per specimen for digitisation from start to finish?

_____Euros per specimen (or specify other currency)

A. Collection type 5

B. i) Has any of this collection been digitised?

☐ **Yes** ☐ **No**

ii) If yes, what proportion as a percentage of the collection has been completed? (Please provide more information where necessary, indicating what steps of the digitisation process i.e. data capture, imaging, georeferencing have been completed)

C. What is the current annual rate of digitization?

_____**Specimens per year**

D. How has the digitisation of the collection been prioritized so far?

E. Are there any pre-digitisation stages required in the workflow?

☐ **Yes** ☐ **No**

Please provide more details

F. Are the digitisation facilities located close to the collection?

☐ **Yes** ☐ **No**

Please provide more details

G. What file format and size are the collection images?

☐ **Tif MB** ☐ **Jpeg MB** ☐ **Raw MB** ☐ **PNG MB**

☐ **Gif MB** ☐ **Jpeg 2000 MB** ☐ **Other MB**

Other

H. What naming conventions are used for the image files? e.g. acquisition number

I. i) At what resolution are the digital images produced.

_____ pixels per inch or _____ MP for collections ___" x __"

ii) How was this resolution decided upon and does it meet the needs of the users?

J. i) What, if any, components are mandatory within your images?

- ☐ **Scale bar**
- ☐ **Institution logo**
- ☐ **Colour chart**
- ☐ **Grayscale chart**
- ☐ **Capsule contents**
- ☐ **Barcode**
- ☐ **Label information**
- ☐ **Other (please give more detail)**

K. What is the imaging protocol followed where multiple specimens appear in one unit (e.g. more than one specimen on a herbarium sheet, more than one lichen on a rock, Insects in a tray etc.)

	please provide more details
Whole unit captured in one image, displaying all specimens	
Whole unit captured in several images, each image displaying all specimens but is named separately	
Whole unit captured in several images, each cropped to display one specimen	
Part of unit captured in one image, to display one specimen	
Other	

L. Please number the following workflow tasks in the order in which they are performed. Leave blank any that are not part of your workflow and add any additional tasks not listed under "Other"

	Selection of material
	Transfer of material between areas
	Barcoding
	Partial Data capture
	Full Data capture
	Imaging
	Record management
	Returning material
	Quality control
	Other (please detail)

M. What is the approximate time taken for the whole workflow from start to finish to image and capture the data from one specimen?

_____minutes per specimen

N. What is the approximate cost per specimen for digitisation from start to finish?

_____Euros per specimen (or specify other currency)