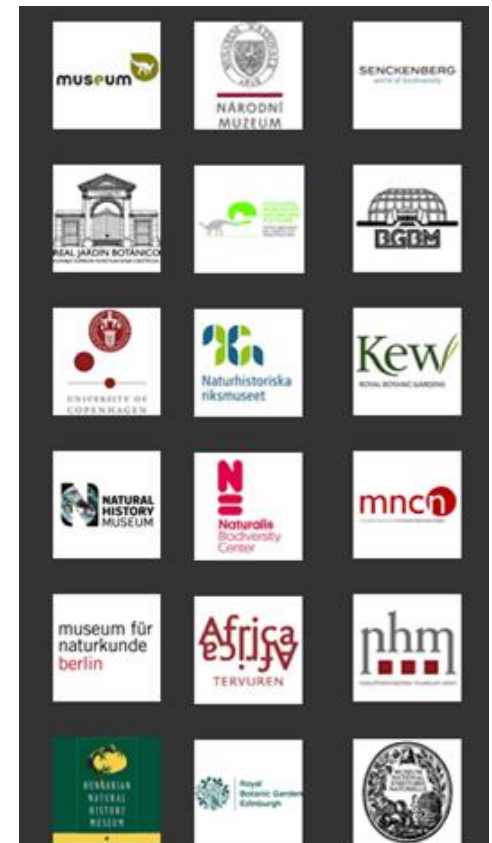


# SYNTHEsys

SYNTHEsys aims to produce an accessible, integrated European resource for research users in the natural sciences. SYNTHEsys will create a shared, high quality approach to the management, preservation, and access to leading European natural history collections.

- 3 projects (2004 – 2017 >33M€). EC funded
- 21 partners (18 with NH collections)
- Synthesys 3 (2014 – 2017)
- Estimated total of c.390 million specimens
- 3 core strands of work:



# SYNTHESYS

Synthesis of systematic resources



## Access

SYNTHESYS Project funding is available to provide scientists based in European Member and Associated States to undertake short visits to utilize the collections, staff expertise and analytical facilities at one of the 18 partner institutions for the purposes of their research.



## Joint Research Activities

The JRA will improve access to data stored digitally within natural history collections by developing mechanisms that enable institutions to enrich digital media with metadata to increase their usability to users.



## Network Activities

The SYNTHESYS Network Activities (NAs) are an integrated package that support and develop the natural history infrastructure in Europe. The greater integration that will result from these activities will enhance the experience of Access users and other visitors.

# NA 3 – Innovation, Impact and sustainability

## *Objective 1: Innovation*

### *Task 1.3 Pilot study into optimal digitisation technologies and equipment - (2014)*

The pilot will involve gathering data from participants on their current digitisation facilities and needs, plus a literature review and questionnaire on how their equipment has been used and their successes and challenges.

Outcome of the pilot study will be a digitisation work flow that can support users/collection managers when requesting/accessing surrogate collections for their research needs. Institutions can also benefit from the results when considering which new equipment to purchase for their infrastructure.

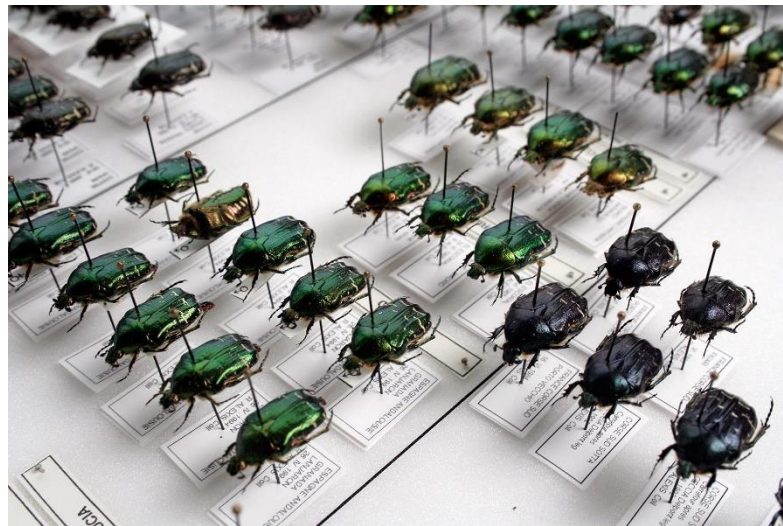
# Questionnaire by section

- General
- Collections
- Quality Control
- Data Management
- Data Storage and Access
- Success, challenges and future developments

Sent to 18 partners. Responses from 13 partners

# Prioritisation for digitisation

- Type specimens
- Project based funding to digitise discrete sections of collections
- Research needs of institution
- Loan or image requests
- Other: New accessions, methodology testing, public exhibitions, curatorial condition of the collections



# Pre-digitisation Tasks

Ca. two thirds of the institutions performed at least some minimal curation, specimen conservation or collections management steps prior to digitisation. This included:

- Checking the identification of the specimens
- Checking the type status
- Checking for pests and pest damage
- Specimen cleaning
- Specimen drawers renewed
- Specimen boxes replaced with acid-free boxes
- Repairing and remounting specimens
- Prepare collections for handling – due to Health and Safety reasons



# Equipment

- Wide variety of equipment used in and between institutions reflecting the diversity of collections.
- Most equipment and models only used at one institution. Large majority DSLR's Canon or Nikon. Many models available that regularly change.
- Flatbed scanners used to digitise 2D material – herbarium sheets, labels, documents, archives, maps, illustrations, wood trunk sections and books.
- 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.



# Equipment

- SatScan used by two institutions - drawers of entomological specimens, molluscs and microscope slides.
- 3D scanning equipment owned by 4 partners (MechScan, SmartScan Duo, NextEngine, HDI Advance R3X, MicroScribe 6G2LX/MicroScan, HDI LMI). Used for digitising Molluscs, paleontological specimens, types and producing 3D exhibition models.
- Book Scanners used to image books, illustrations and bound Herbarium specimens.
- X-ray equipment – for fish morphology, sex determination and stomach content for snakes, pathogenic research on bones, Zoological, Paleontological and Archaeological collections.
- Microscope Cameras.



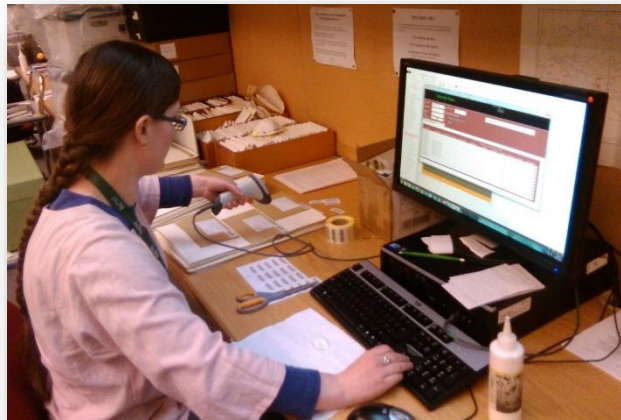
# Imaging Standards

- Most institutions did not appear to follow any official imaging standard and it appeared that awareness of the standards was under developed. Institutions defined their own standards or followed project driven standards for imaging.
- 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.



# Metadata

- For the majority of collections full specimen metadata was captured prior to the imaging step. Two institutions reported capturing partial metadata only.
- Only one institution reported imaging specimens prior to any specimen metadata capture.
- 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. Only one institution captured partial data first, then imaged then full data capture.

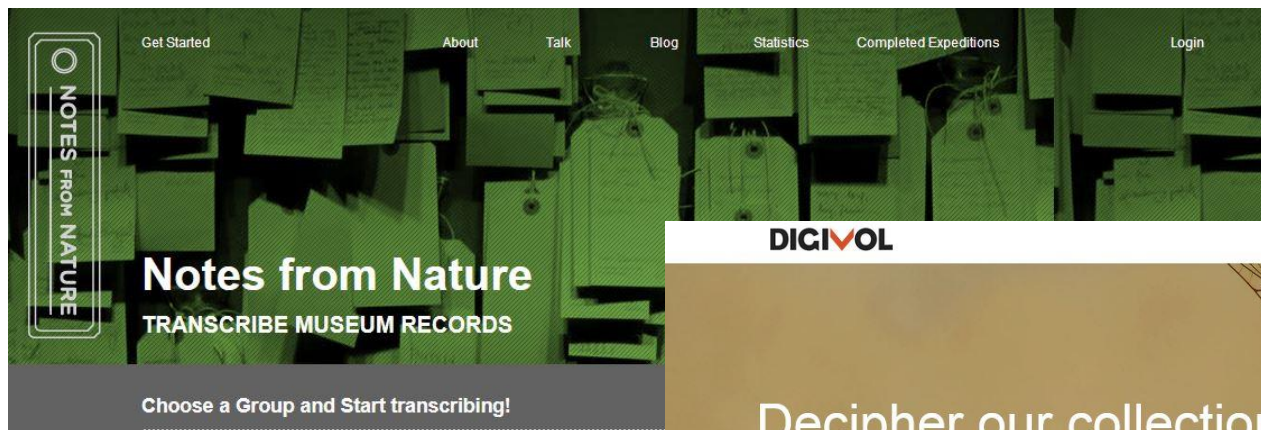


# Outsourcing

- Implemented in one institution only for Botanical sheets and periodicals. Considered by four institutions.
- Challenges of outsourcing:
  - Fragility of specimens and lack of collections handling experience.
  - Private handling of public world heritage is questionable and subjected to uncertainty.
  - Risk of transporting specimens.
  - Lack of expertise in collection data transcription.
- Advantages of outsourcing:
  - Reduced cost per specimen.
  - Higher digitisation rates.
  - No management of digitisation staff.
  - More time for research.



- Three institutions were currently investigating using crowdsourcing options using microscopic slides and herbarium specimen collections.
- Two institutions included Optical Character Recognition (OCR) in their workflows, one of these used it for digitisation of literature only.



# Georeferencing

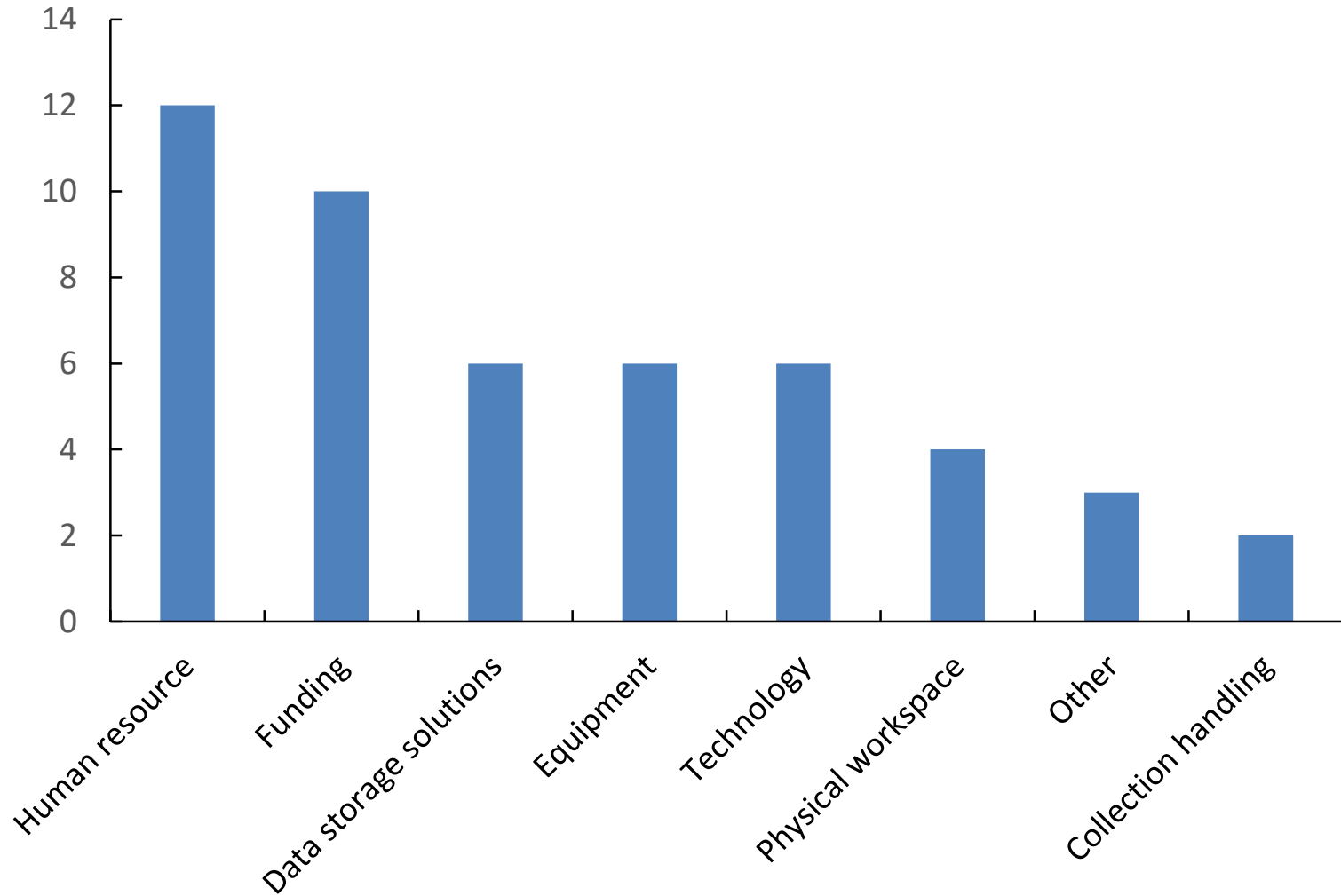
- Georeferencing is carried out in approximately 50% of the digitisation workflows. However, in general only a proportion of the specimens are routinely georeferenced depending on priority. In some cases coordinates are only recorded if they are found on the specimen labels. Two institutions report developing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



# Strategies used to increase rates

- Dedicated digitisation staff
- Get more funding for more human resources
- Collaborations - data capture by partners
- Investigate in development of new digitisation technologies
- Data storage facilities increase
- Improving our database and education of digitisation staff
- Continuous improvements of database and workflow; involving all staff (spreading knowledge and making digitisation more "fun")
- OCR to aid data capture
- Standardise processes; define what minimal object data needs to be registered; only make images when certain rules apply
- Move from scanners to digital camera

# Limitations on digitisation

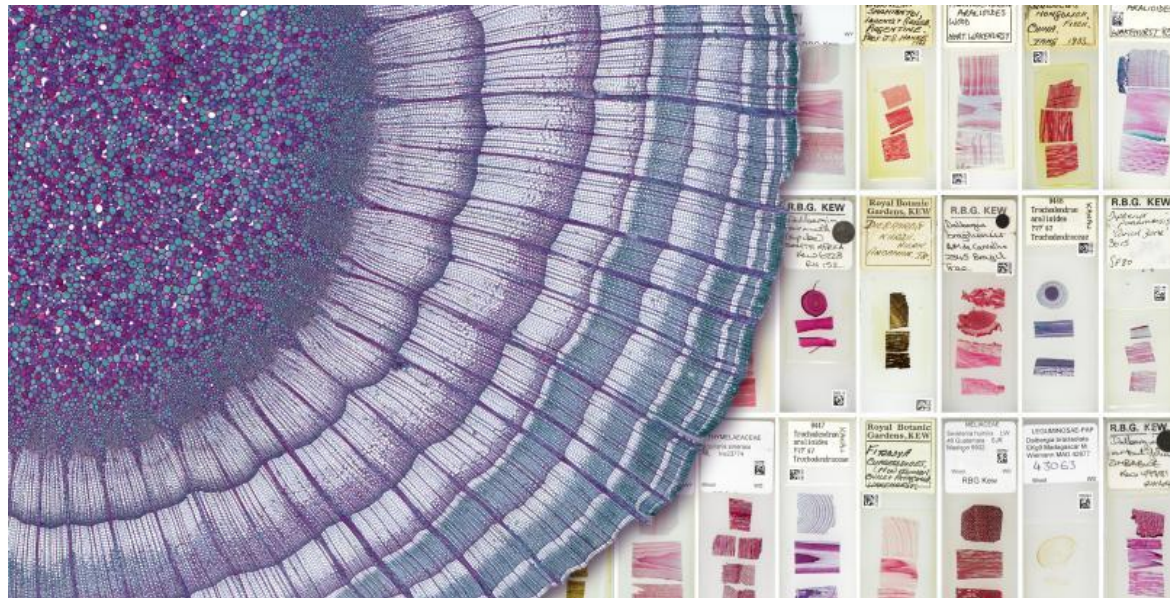


# Limitations Comments

- Highly diverse ecosystems of data management and digitisation workflows
- Lack of staff with dedicated work hours for digitisation
- We are currently lacking the digital infrastructure; we do not as of yet have a common, solid, collections management database system. The bottle neck is therefore clearly after the imaging, when images need to be filed
- Keeping the physical and digital collection in sync.
- The use of an industrial book scanner (or equivalent) would speed up the imaging of vascular plants



- A lack of suitable digitisation infrastructure was seen as an impediment to digitisation rates. It is clear from the questionnaire that institutions with natural history collections require knowledge and assistance, not only with the equipment and digitisation workflows, but also 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.



# Recommendations

- List of grants/funding bodies available for digitisation of collections
- Disseminate digitisation successes more widely. Demonstrate the importance of digitisation and develop and share use cases
- Repository to share workflows – There are many different types of equipment and digitisation workflows a central place to access information on them would be extremely valuable. Can we publish more of them?
- Lack of knowledge of official image standards. QA of images all done visually, should we investigate automatic checks of image quality to speed up some aspects of QA?

# Recommendations

- 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 and manual. Only a few institutions of those surveyed had implemented alternative approaches e.g. Outsourcing, addition of an OCR step, capturing partial data first and full data later, crowdsourcing and testing automated georeferencing. 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. Keep investigating automated processes.

# Lessons Learnt

- Test and Test again
- Difficult to get the balance right between creating a too large and complex survey and getting enough details you need to draw conclusions
- Perhaps better to narrow down to specific areas
- Follow up interviews to discuss issues raised in the questionnaire
- Only sent to partners - a wider sample size needed

Thanks to all synthesys participants that took part in the survey.

Full Report can be found here

<http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.3-Review-of-Digitisation-workflows-and-equipment.pdf>

Following on from this a digitisation workflow template that can support users/collection managers when requesting/accessing surrogate collections for their research needs has been developed by the NHM.

<http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.3-NHM-Digitisation-Protocol-Template.pdf>