



## SYNTHESYS Final Publishable Summary Report

Grant Agreement number: 312253

Project acronym: SYNTHESYS3

Project title: Synthesis of systematic resources

1st September 2013 – 31st August 2017

### Contents

Executive Summary.....	2
Summary description of the project context and objectives .....	3
Description of the main science and technology results.....	7
JRA: Moving from physical to digital collections.....	7
NA2: Improving Collections Management and Enhancing Accessibility .....	13
NA3: Innovation, Impact and Sustainability .....	22
The potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and exploitation of results.....	32
Impact of the Joint Research Activity.....	32
Impact of the Networking Activities.....	35
Impact of Transnational Access .....	38
Dissemination activities and exploitation of results .....	39
Glossary:.....	42
The address of the project public website.....	42

## Executive Summary

SYNTHESYS3 has created an accessible and integrated European resource for researchers in the natural sciences in Europe and globally, providing access to leading European natural history (NH) collections held in museums and herbaria. These collections are world-class in terms of their geographic and taxonomic coverage, representing a unique infrastructure in Europe as a model of the natural world. Collections provide baseline data with which the research community addresses global challenges such as biodiversity loss, food security and sustainable supply of raw materials.

Present and future access to European NH collections relies on specimens and their accompanying data being well-preserved and widely accessible both now and in the long-term. As technologies have developed, new collections are growing within NH institutions in two forms: (1) *virtual collections* (e.g. digital, scanning electron microscope and video images, metadata, CT scans, chemical analytical data, DNA barcodes, genomic information) and (2) *new physical collections* (e.g. frozen tissue, DNA, RNA). These new collections increase demand from the User community, necessitating improved management and access.

Over four years SYNTHESYS3 has supported 1,110 research projects of physical users of the collections via its **Transnational Access** programme. SYNTHESYS3 Users are addressing global research issues including environmental and climate change and climate modelling, soil biodiversity, marine, freshwater and terrestrial ecosystem functioning, mineral deposit distribution and food security. Largely through taxonomic and systematic studies, Users delivered new and improved baseline bio- and geo-diversity data, and built new collaborations during the project to generate and add value to existing collections. 11,030 days of Access have been provided to 17 consortium institutions in 11 European countries. Users have recorded >1,400 research outputs which have already led to delivery of 460 peer reviewed papers (currently accepted, in press or published), 18 books/monographs and support to > 60 PhD student theses.

The SYNTHESYS3 Joint Research Activity (JRA) has focused on **extracting and enhancing data from digitised collections**. It has (i) developed and delivered open source software: *Inselect* that can recognise, process and annotate images that contain multiple specimens (e.g. whole drawer scans of pinned insects or slide arrays); (ii) developed 3D techniques to digitise NH objects to give researchers tools to undertake their research without the need of the physical object; (iii) reviewed and utilised innovative methods of data capture via technology and via crowdsourcing, and (iv) provided open access to major research datasets. These innovations have greatly increased access to digitised collections, increasing the democratisation of research by giving all researchers equal opportunity to do high quality research with digital objects regardless of their location.

Networking Activity 2 has focused on **improving collections management and accessibility of new physical and virtual collections**, producing protocols, best practice guides and policies that are openly available. This enables NH institutions to maximise the use of their existing resources whilst avoiding duplication of effort. NA2 has (i) developed policies for virtual and new physical collections management; (ii) produced best practice and standards for 3D imaging of NH specimens; (iii) developed strategic priorities for barcoding and DNA library creation of NH collections; (iv) developed high throughput protocols for data collection through DNA extraction; (v) strengthened networks of DNA and tissue banks; and (vi) developed the collections self-assessment tool providing benchmarks for management of digital and molecular collections.

Networking Activity 3 has utilised the consortium network to achieve **innovation, impact and sustainability**. NA3 has: (i) reviewed optimal digitisation technologies and procedures, (ii) tested crowdsourcing feasibility for digitisation (iii) promoted and disseminated SYNTHESYS3 outputs, (iv) fostered relationships with industry, and (v) collaborated to train users in new technological advances in digitisation via workshops and openly available, multilingual training videos. Reviewing a series of evidenced use cases has demonstrated how NH collections, expertise and services can be used to find solutions to a range of societal challenges. This constituted the first element for building a roadmap for the future access policy of European NH institutions and has been a critical step in the development of the new pan-European NH collections research infrastructure DiSSCo (Distributed System of Scientific Collections, <http://dissco.eu>), an ESFRI initiative which has been built on a foundation of SYNTHESYS activities.

## Summary description of the project context and objectives

The overall scientific objective of SYNTHESYS3 was to improve significantly the quality and flexibility of access to natural history (NH) collections for the widest possible range of European researchers. Much has been achieved in the management and accessibility of traditional NH collections. As technologies have developed, new collections are growing within NH institutions in two forms: (1) virtual collections (e.g. digital, scanning electron microscope and video images, metadata, DNA barcodes, CT scans and chemical analysis data) and (2) new physical collections (e.g. housing frozen tissue, DNA, RNA). Their management and accessibility have historically been fragmented. SYNTHESYS3 worked to integrate these new collections in ways that will have a lasting effect on the landscape of the infrastructures, broaden the spectrum of research that can be carried out, and create new information sources available over the web.

### Transnational Access

Access to collections, their associated supporting expertise and specialised equipment is vital in the field of NH research. A central element for SYNTHESYS3 was to deliver demand-led physical access (via Transnational Access, TA) to over 390 million specimens housed by the 17 collection-holding institutions within the consortium, organised into 11 Taxonomic Access Facilities (TAFs). SYNTHESYS3 aimed to provide a minimum of 10,795 User days through four annual open calls for proposals.

## Joint Research Activity (JRA): Moving from physical to digital collections

The overall aim of the JRA was to address the new “virtual collections” of Natural History (NH) collections in a similar way that the physical collections are maintained and curated. This is a crucial given the speed at which new “born digital” collections, such as digital images are being added to the collection; given the large number of forms in which collections are created (for example sound recordings, videos, images) and given the volume of images that are being created during the digitisation process of the physical collections. All these collections, physical or virtual, “born digital” or as a result of digitisation, have data and metadata associated. To be useful to the end-user all this has to be accessible and of high quality. Hence the aim of the JRA was to focus on improving the quality of the digital collections and data within NH institutions and increasing access to the digital collections and the data.

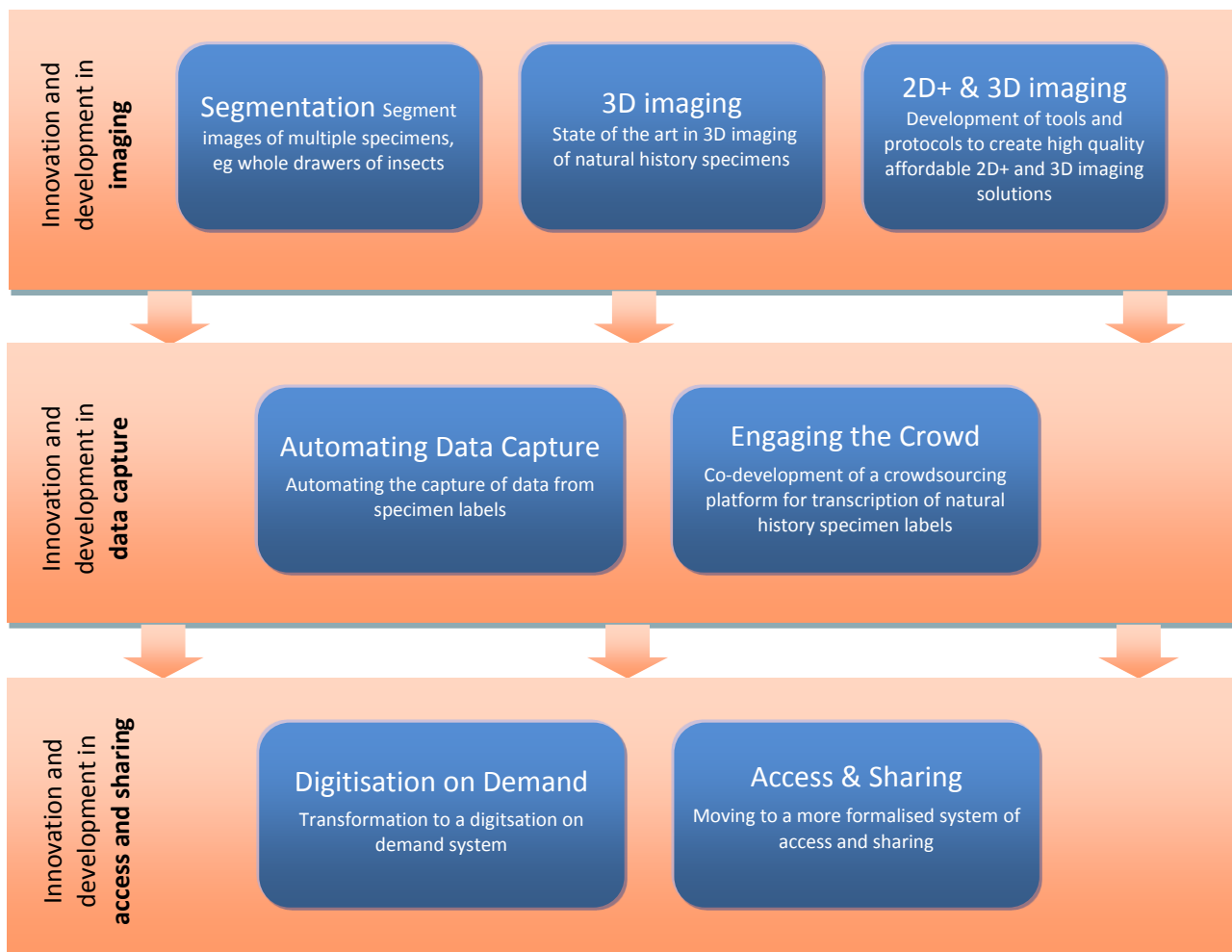
The JRA had four key objectives:

Objective 1: Automated data collection from digital images

Objective 2: New methods for 3D digitisation of NH collections

Objective 3: Crowdsourcing metadata enrichment of digital images

Objective 4: Access and management of an integrated European digital collection (with NA2)



## Networking Activity 2 (NA2): Improving collections management and enhancing accessibility

With increased importance of new technologies and world-spanning networks, NH institutions are faced with an ever expanding amount of digital data which need to be dealt with properly. These digital (or virtual) collections add immense value to the physical collections and are often of interest for new user groups, thereby significantly contributing to novel fields of application not anticipated before. Like these digital collections, new physical (molecular and tissue) collections are areas of rapid development and provide key material for biological research. To make the most of their potential and to maximise their reasonable use, digital collections and new physical collections need to be easily accessible, well cross-linked, and publicly visible, with quality controls and common standards.

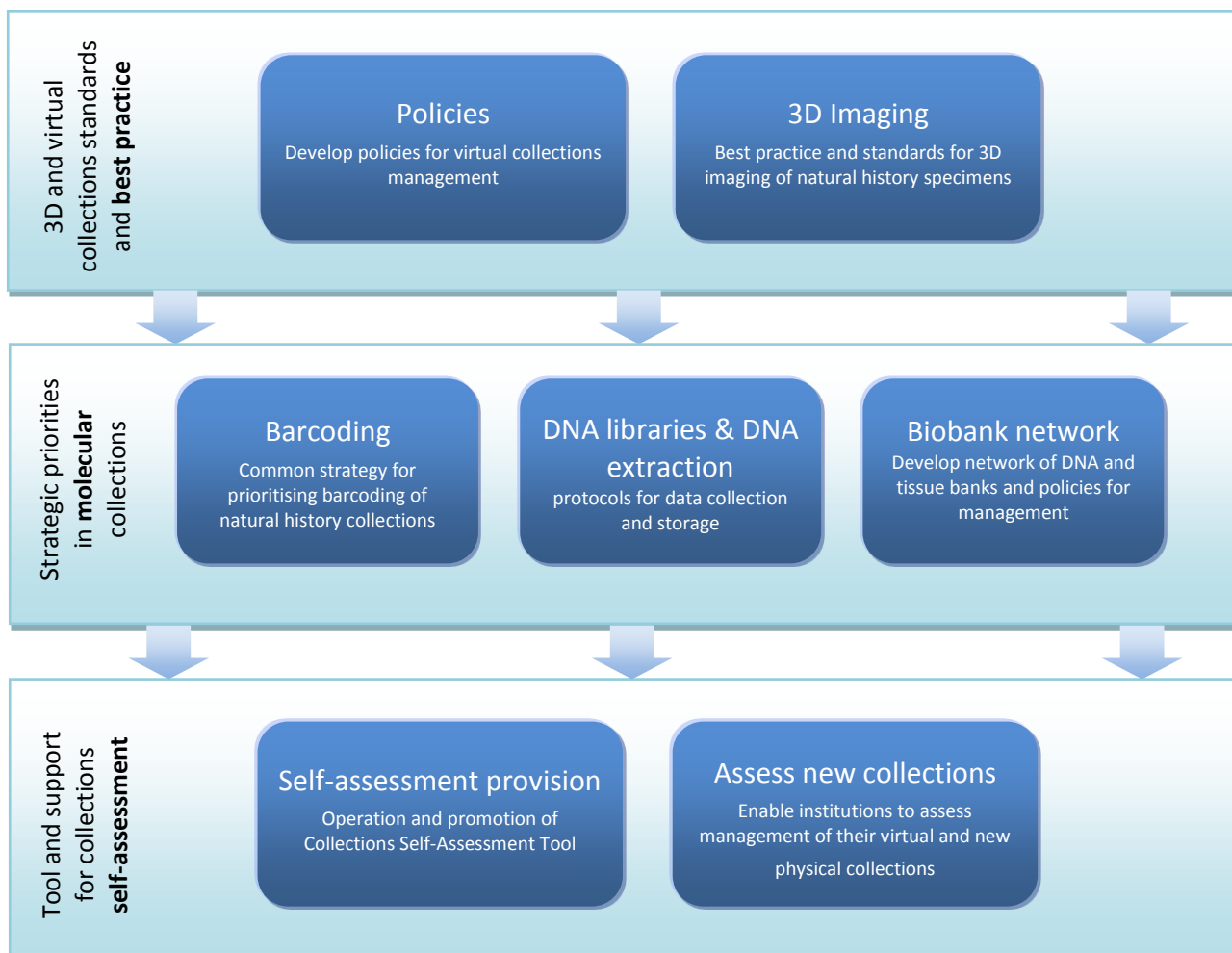
To react appropriately to these developments and meet the expanding needs and increased demand of users, the main concept of SYNTHESYS3 **NA2** was to provide a coherent, integrated management approach for NH collections, providing policies, best practices, protocols, strategic priorities and benchmarked standards in close collaboration with other relevant European and non-European initiatives. This built on successful improvement of traditional NH collections management in both preceding SYNTHESYS initiatives (project numbers 506117 and 226506) to focus on management of digital and new physical collections as well as on open and flexible access to these.

NA2 had three key objectives:

Objective 1: Managing new (virtual and physical) collections

Objective 2: Developing strategic priorities for molecular related NH collections

Objective 3: Develop Collections Self-Assessment Tool (CSAT) for new collections



## Networking Activity 3 (NA3): Innovation, Impact and Sustainability

The primary aims of the **NA3** were to ensure that SYNTHESYS3 produced sustainable deliverables, reaching the broadest possible audience and fulfilled its potential impact on the European Research Area. The focus included development and reinforcement of partnerships between NH institutions and industrial partners, plus ensuring the project remained innovative in light of future technological advances.

NA3 aimed to assist European NH institutions to operate in a more sustainable manner by producing a roadmap for future priorities to meet the needs of the research community. In order to build a sustainable future for digitisation of NH collections it is critical to collaborate with other large-scale programmes on training people in new technological advances in digitisation. NA3 aimed to foster relationships with industry and link with the US-based iDigBio programme (Integrated Digitised Biocollections, [www.idigbio.org](http://www.idigbio.org)) to prioritise and provide training in these new technologies. To complement the innovations within the JRA, NA3 evaluated current use of state-of-the-art digitisation equipment as well as crowdsourcing for large-scale data gathering to establish and publish guidelines and recommendations for technicians and users of NH collections.

Wide dissemination of the project's results would be essential to ensuring maximum impact. Dissemination both within and outside Europe, combined with making all data outputs freely available and open access, would ensure global impact on improving collections management thus enabling greater access to global NH collections by researchers.

NA3 had three key objectives:

Objective 1: Innovation

Objective 2: Impact

Objective 3: Sustainability





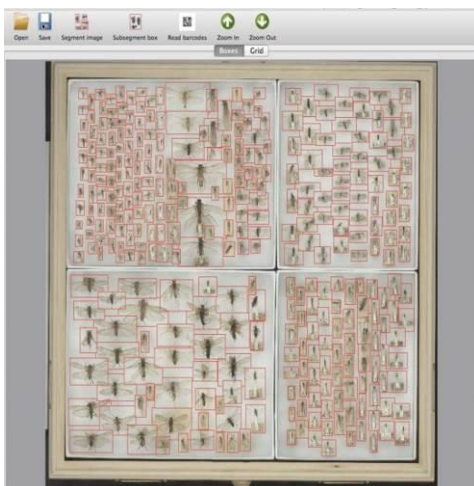
# Description of the main science and technology results

## JRA: Moving from physical to digital collections

### Innovation & Development in Imaging

#### Insect image segmentation software

Natural history (NH) collections are vast and varied and present many substantial challenges to digitisation. At the Natural History Museum, London (NHM), for example, there are an estimated 33 million insect specimens housed in 130 thousand drawers:

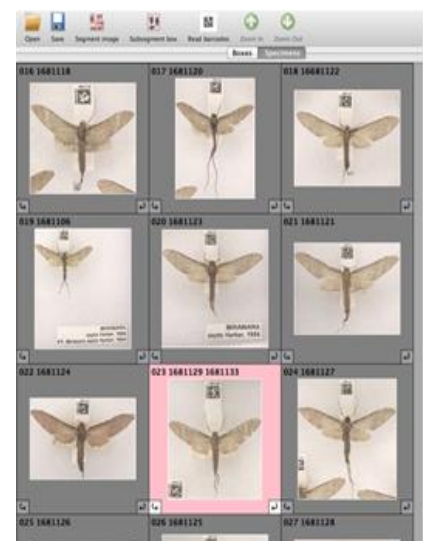


It is a lot easier and quicker to image 130 thousand drawers rather than 33 million individual insects. However, by themselves drawer-level images are not very useful. Manually cropping each image takes too much time and without unique identifiers the individual images are of questionable value. The challenge is to efficiently get a single image of each object along with its associated metadata.

[Inselect](#) is a desktop application developed in SYNTHESSYS3 that automates the cropping of individual images of specimens from whole-drawer scans and

similar images that are generated by digitisation of museum collections. It combines image processing, barcode reading, validation of user-defined metadata and batch processing to offer a high level of automation. Inselect runs on Windows and Mac OS X and is open-source: <https://naturalhistory.museum.github.io/inselect/>.

Inselect is now being used or evaluated by more than ten NH organisations across at least six countries to assist the digitisation of microscope slides, pinned insect specimens, malaise trap samples and palaeontological specimens.



## Zoosphere 3D imaging and website

3D imaging of NH objects is as diverse as the different types of objects in collections. The main goal of developing 3D techniques to digitise NH objects is to give researchers a tool to undertake their research without the need of the physical object. Access to the digitised collections will be greatly increased, allowing for the democratisation of research, giving researchers or museums with a low budget the opportunity to do high quality research with the digital objects.

ZooSphere was developed at the Museum für Naturkunde (MfN) to create digital 3D “spherical” representations of pinned insects. This tool enables automatic image capture from up to 1,000 different perspectives around the pinned insect. The image capture, the positioning of the object and the camera motion for the stacking process is achieved by a piece of software developed as a part of the project.

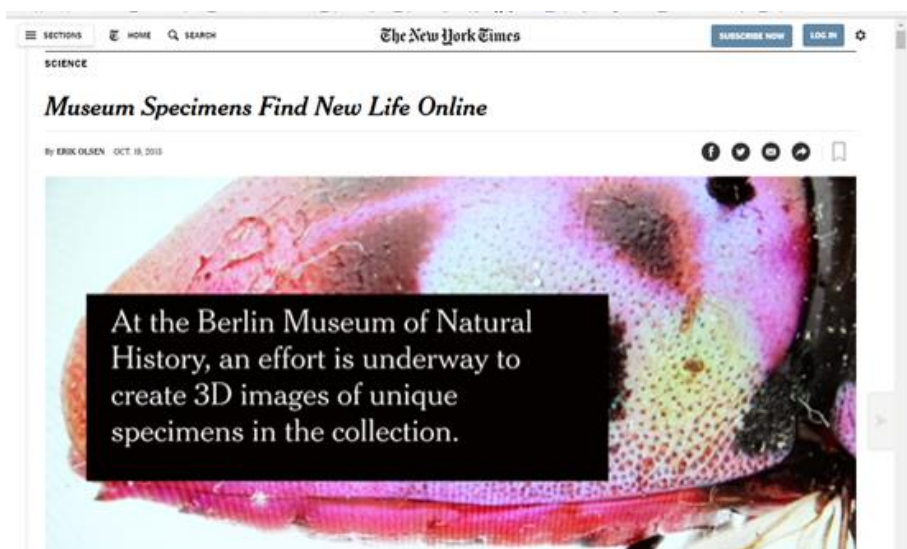


All ZooSphere objects are available on the webpage <http://www.zoosphere.net/>. The interactive viewer presents a preview of the image sequence for every object. To obtain a full resolution image a java software component developed within the project is freely available for download.

To date more than 110 specimens have been digitised using the ZooSphere and placed online, including both pinned and amber preserved insect specimens. Some of these image sequences have been requested by scientists from other Institutions for their research. Google Analytics logged 10,847 users and 17,731 page views from 20.09.2015 to 20.09.2016 with 25% returning visitors.

ZooSphere has received international attention through articles published by the New York Times and the VICE magazine (print magazine and website focused on arts, culture, and news topics).

The Department of Agriculture and Food in South Perth, Australia, has stated interest in adapting the ZooSphere technology for pest control services.





## Affordable 2D+ & 3D imaging solutions

The high resolution multimedia recording of small specimens is a real challenge for NH museums who are working on mass digitisation programs. The quality of the resulting image, the cost of the equipment, the human work and the learning curve are important parameters in order to define a general digitisation strategy. SYNTHESYS3 evaluated several techniques in order to allow the digitisation of NH specimens for a real scientific purpose. The goal is to produce images or virtual models which provide at least the same level of information than the direct observation of the specimen.

- 2D+
  - Focus Stacking (UV + White Light)
  - Reflectance Transformation Imaging (RTI)
  - MiniDome
- 3D:
  - Photogrammetry
  - Extended Depth of Field Photogrammetry
  - Laser Scanning
  - Structured Light Scanning
  - $\mu$ CT scanning



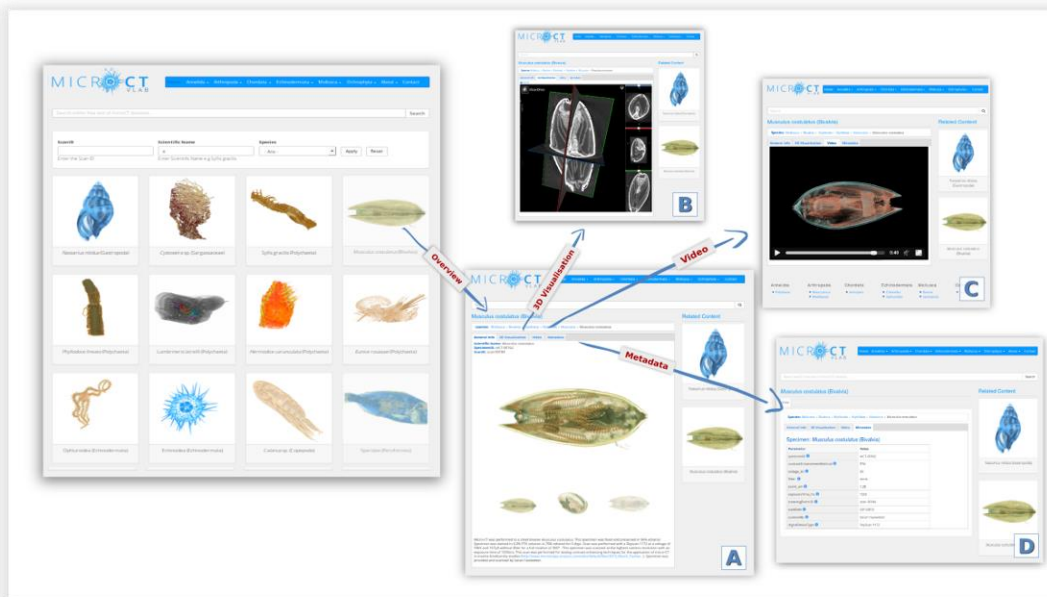
All the above techniques were thoroughly tested and adapted so they would be suitable for the digitisation of a particular part of the NH collections. For each technique a workflow was written and shared online through the Wiki-page of the Best Practice Handbook for 3D digitisation of a Museum Collection (<http://biowikifarm.net/v-mfn/3d-handbook/>).

Following the testing, one approach, focus stacking, was selected and commercial top-end set-ups were compared with a semi-automatic low budget approach. The tests found a possible solution for mass digitisation of specimens. The largest benefit is the price of the set-up which is approximately €3,000.

## Micro-CT imaging solutions

Micro-computed tomography (micro-CT) is a non-destructive imaging technique which allows the rapid creation of high-resolution three-dimensional data (images). Based on x-ray imaging, it allows a full virtual representation of both internal and external features of the scanned object. The resulting 3D models can be interactively manipulated on screen (rotation, zoom, virtual dissection, isolation of features or organs of interest), while 3D measurements can also be performed – from simple length and volume measurements to density, porosity, thickness and other material-related parameters.

The Hellenic Centre for Marine Research (HCMR) has been the leading partner for this SYNTHESYS3 task, which contains two major fields of research: (a) developing protocols and workflows for the rapid digitisation of collections (sample preparation, scanning parameters, model creation), and (b) creation of web-based tools that allow the user to display and manipulate the 3D tomograms. HCMR created a comprehensive [Micro-CT Handbook](#) with best practice protocols for NH museum specimens in collaboration with experts in micro-CT imaging. Demonstration datasets used while developing the protocols can be displayed and manipulated in a web based platform created by HCMR in collaboration with the LifeWatchGreece infrastructure (<https://microct.portal.lifewatchgreece.eu/>).



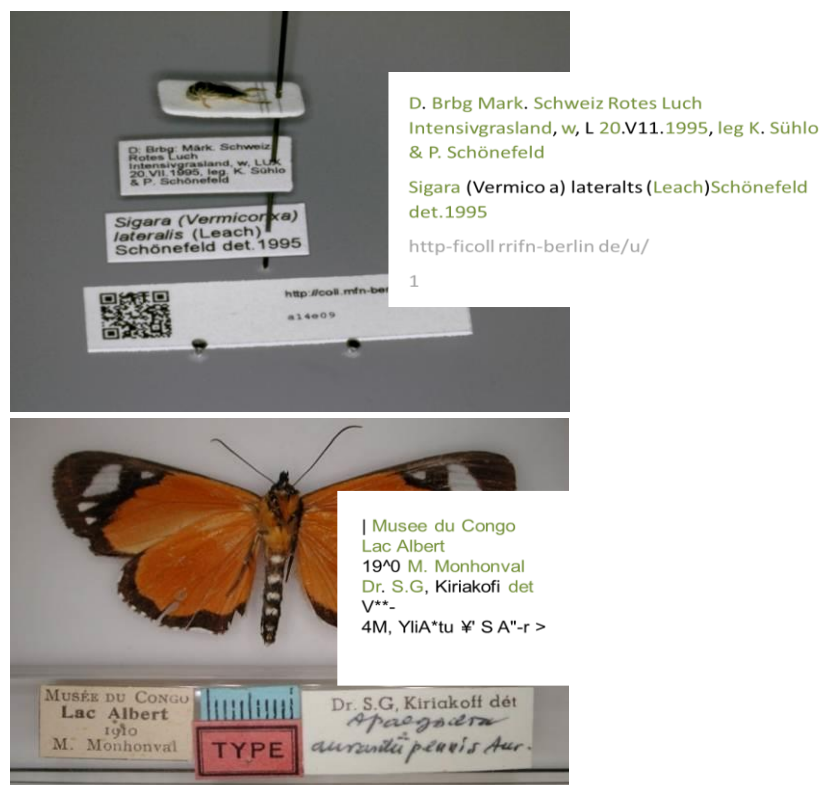
In collaboration with the University of Vienna, a number of protocols to remove the stains (in particular iodine stains and phosphotungstic acid stains) caused by chemicals during preparation of specimens for scanning were developed and tested. A total of 600 samples have been scanned for the development of protocols. The results and respective observations have been added in the [Micro-CT Handbook](#).

## Innovation & Development in Data Capture

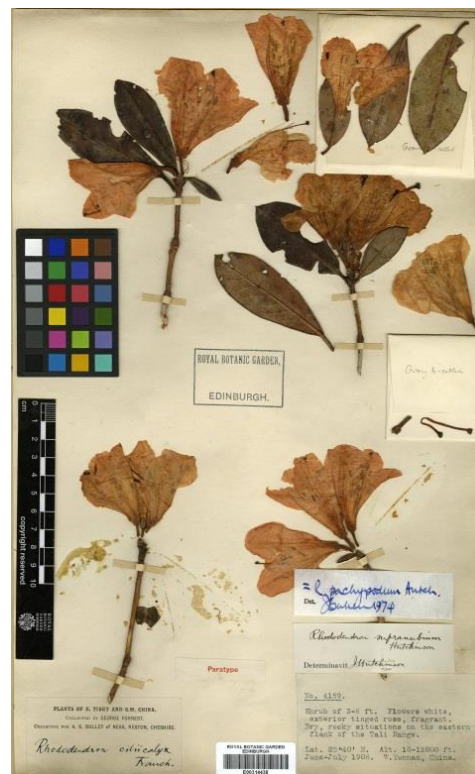
### Using Optical Character Recognition (OCR) to capture data

A range of OCR options currently available were reviewed and trialled by six SYNTHESYS3 institutes. The trials involved images representing a range of material including plants, insects, molluscs and fossils. The results emphasised the usefulness of using OCR technology in the digitisation workflow, and discovered three options which provided the best results. One is server-based (ABBYY Recognition Server v3), one runs on a PC (ABBYY FineReader v12 Professional) and one is an online service (Onlineocr.net).

ABBYY Recognition Server v3 was installed at the Royal Botanic Garden Edinburgh (RBGE) and has now been used to process over 400,000 specimens. The OCR output has been used to find all the specimens from a single collector or a single country. These pieces of data have then been added as a batch process and been presented to staff, volunteers and online crowdsourcing platforms for additional data entry. Over 150,000 specimens have been transcribed through this process at RBGE.



EDINBURGH  
 Determinavit  
 Paratype  
 Ilo. 4159  
 PLANTS OF E. TIBET AND S.W. CHINA.  
 Collected by GEORGE FORREST.  
 Collector for A. K. BULLEY of NESS, NESTON, CHESHIRE.  
 CMCvi  
 ROYAL BOTANIC GARDEN EDINBURGH  
 E00314438  
 Shrub of 3-6 ft. Flowers white., exterior tinged rose, fragrant. Dry, rocky  
 situations on the eastern flank of the Tali Range.  
 Lat. 25°40' IT. June-July 1906.  
 Alt. 10-11000 ft. W.Yunnan, China.  
 E00314438



## Engaging the crowd to capture data

A specimen image only tells a small part of the story. Vital contextual information is stored on labels, and turning these into useful digital data presents a challenge. Despite improvements in handwriting recognition software, label data can be difficult to capture automatically with accuracy. Crowdsourcing – using citizen scientists and volunteers to perform tasks which collect data and integrate into research workflows – is a potentially very effective way of addressing this problem. SYNTHESYS3 partnered with Notes from Nature ([www.notesfromnature.org/](http://www.notesfromnature.org/)), part of the Zooniverse crowdsourcing platform, to launch [crowdsourcing projects](#) centred around the capture of handwritten label data for digitised museum collections. Projects included:

**Miniature Lives Magnified:** The pilot project, launched in August 2016, concerned a collection of chalcids – a group of parasitoid wasps – from the NHM London. Chalcids are of high scientific importance due to their role as a biological control agent which can protect crops and reduce invasive species. Around 100,000 microscope slides containing chalcids were imaged and their labels transcribed by 1,000+ ‘digital volunteers’ over a 10 month period.

**Miniature Fossils Magnified:** The second ‘Magnified’ project focused on digitised slides containing foraminifera - ocean sediment-dwelling organisms – from shallow tropical seas. These tiny specimens are hugely important as they provide information on how the environment, climate and ocean have changed over the past 500 million years. So far 2,071 slides have been transcribed with a third and final batch being released in autumn 2017.

 NATURAL HISTORY MUSEUM

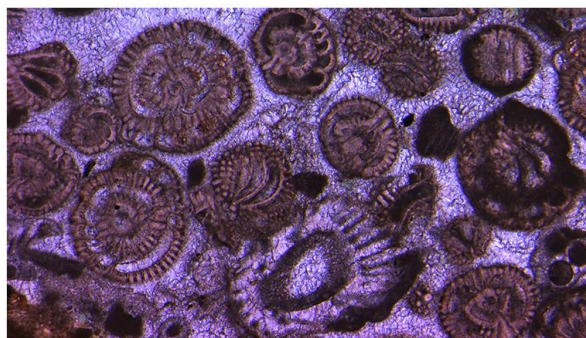
Visit Discover Take part Join and support Shop Schools Our science Search

 Hours and admission

 Become a Member

Home / Take part / Citizen science / Miniature Fossils Magnified

## Miniature Fossils Magnified



### At a glance

Transcribe microscope slide labels.

**Type of activity:** Online

**Who can take part?** Adults and students (Key Stage 4+)

**When?** Any time

**How long will it take?** Two minutes per slide

Be a digital volunteer for the Museum. Extract research data from microscope slides of some of the world's smallest fossils.

### Project team

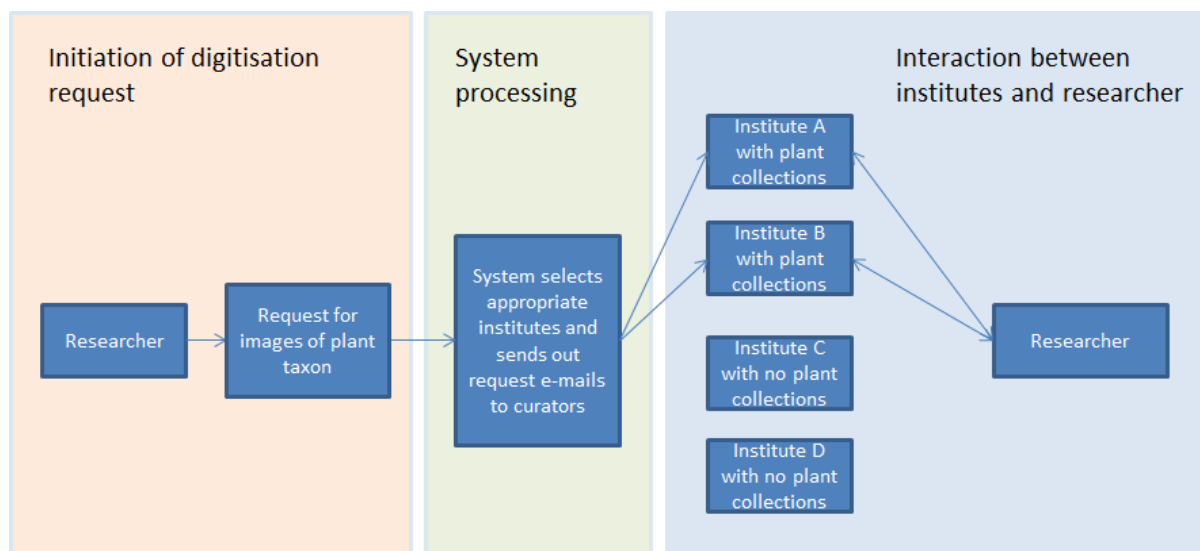
• Margaret Gold, Science Community Coordinator (SYNTHESYS)

## Innovation & Development in Access & Sharing

### Digitisation on Demand (DoD)

A comprehensive [report on DoD](#) was produced based on two case studies carried out at RBGE and the Swedish Museum of Natural History (NRM), a survey of current DoD infrastructure and capacity within European NH collections, and a study of potential funding models including the inclusion of crowdfunding.

The report included a recommendation for building and deploying a European networked DoD service.



Diagrammatic overview of a potential European DoD network of institutes holding NH collections.

The case study carried out at NRM brought together outcomes from across the SYNTHESYS3 project including imaging workflows, data transcription workflows and crowdsourcing and citizen science platforms. This project included the trial digitisation of fossil brachiopods and the Münchenberg herbarium. This report essentially paves the way for the implementation of a networked approach to DoD across Europe in collaboration with the Consortium of European Taxonomic Facilities (CETAF, <http://cetaf.org>).

### Access and sharing collections

Open access to the data is critically important for the current and the future users of the data. Successful open access to the collections' data and images will involve having suitable policies in place, having a clear strategy for selecting data portals and aggregators as well as ensuring that consistent data standards are being used to share data and images.

#### Access Policies

Work to review and develop access policies for institutes was carried out within the NA2 work package and was closely linked to this task within the JRA. The [Recommendations for a Management Policy on Digital Collections](#) (MPDC) will form the basis of European institutional virtual (or digital) data management policies and procedures that are more closely aligned in the future. The MPDC has been produced in the form of an online handbook which has now been adopted by the CETAF Digitisation Working Group. A short policy statement on digital data produced by guests or other external users of NH collections was also written and this is to be adopted by the CETAF Collections Working Group.

#### Standards for sharing data

The process of sharing data is reliant on standards. A review of existing standards and data schemas was carried out to ensure that the necessary standards were available for use for the wide range of collections held in NH institutes. The review covered the main standards and schemas used for NH collections and looked at workshops and discussions that have taken place in the past to harmonise the standards across



domains. As we move forwards we will need to ensure that appropriate standards are used and that any gaps in standards or schemas are identified and resolved by either extending existing standards or by developing new standards within the community.

### Aggregators for accessing and sharing collections

Finally, an assessment of existing data aggregators was carried out. A survey showed over 90% of the 27 institutes responding both use and send data to aggregators. A key portal for access to occurrence data relating to NH collections is the Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org)). This portal currently contains a total of 136,311,486 aggregated records of preserved specimens including fossils from institutes around the world. A comparison of different data portals was carried out for one institute (RBGE) which found some remarkable figures to demonstrate the level of use of NH collections data. In a single month in 2016, over 18 million RBGE specimen records were downloaded.

The following **JRA deliverables** were realised (all available on the JRA pages of the SYNTHESYS website: <http://www.synthesys.info/joint-research-activities/jra-1-synthesys-3/>):

Title	Explanation	Lead	Link
Edge detection technology	Refined method of edge detection technology for natural history collections	RBGE + NHM	<a href="http://www.synthesys.info/wp-content/uploads/2013/07/JRA-Del.-4.1-Edge-detection-technology-report-InSelect.pdf">http://www.synthesys.info/wp-content/uploads/2013/07/JRA-Del.-4.1-Edge-detection-technology-report-InSelect.pdf</a>
Optimal automated metadata capture	Report on optimal automated metadata capture for natural history collections	RBGE	<a href="https://www.dropbox.com/s/mgu036hx0ilyaia/Del%204.2%20SYNTHESYS3JRATask1.2FinalReport.compressedtwice.pdf?dl=0">https://www.dropbox.com/s/mgu036hx0ilyaia/Del%204.2%20SYNTHESYS3JRATask1.2FinalReport.compressedtwice.pdf?dl=0</a>
Crowdsourcing for metadata enrichment	Report on crowdsourcing as a tool for metadata enrichment	MfN	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del.4.3-Crowdsourcing-for-metadata-enrichment.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del.4.3-Crowdsourcing-for-metadata-enrichment.pdf</a>
Protocols for optimal MicroCT	Report on protocols for optimal MicroCT use for natural history collections	HCMR	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del4.4-HandbookofbestpracticeformicroCT.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del4.4-HandbookofbestpracticeformicroCT.pdf</a>
Digitisation on demand	Report on feasibility of a Digitisation on Demand (DOD) service	NHM	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del4.5-Digitisation-on-Demand.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del4.5-Digitisation-on-Demand.pdf</a>

## NA2: Improving Collections Management and Enhancing Accessibility

### Policies for virtual collections management

A SYNTHESYS3 survey and more general research has showed that policies on virtual (digital) collections management and related official procedures are, if present at all, fragmented and scattered both across the landscape of natural history (NH) institutions and due to diverse institutional regulations. Also, it was acknowledged that specific national and institutional regulations complicate, and in some cases overrule, the development of common definitive policies across European institutions with NH collections.

The main aim of this NA2 task therefore was to collate already existing European virtual (digital) collection policies and to use these as a basis for recommendations on common European digital data management policies and procedures. Areas covered by these recommendations were to include, for example, sustainable storage, open access, intellectual property rights, and ABS (Access and Benefit Sharing).

Results were mainly based on the feedback from a survey on the gathering of digital data produced by visiting researchers of NH collections and the presence of respective procedures and/or policies as well as on the results of a more general research on existing and already established collections management policies covering digital data.



The task's final output was split into two components and two documents of common interest for institutions with NH collections have been produced:

- 1) a short [policy statement](#) on digital data produced by guests or other external users of NH collections;
- 2) an online handbook and practical guideline with recommendations on Management Policies on Digital Collections (MPDCs): <http://synthesys3.biowikifarm.net/syn3/NA2/objective1/task1/cmp>

The **policy statement** "Common Principles on Digital Data produced by External Users of NH Collections" gives a set of common principles for institutions housing NH collections across Europe and world-wide. These principles concern the handling of digital data produced by visitors or other external guests of these collections. It gives a description of the legal framework the undersigning institution adheres to and its commitment regarding digital collection data. Also it declares principles to optimise the management of digital collections and of digital collection data produced by external users of NH collections.

The **online handbook** or **practical guideline** "Recommendations: Management Policy on Digital Collections" (<http://synthesys3.biowikifarm.net/syn3/NA2/objective1/task1/cmp>) was in close collaboration with SYNTHESYS3 members as well as CETAF via the Collections Policy Board (CPB) and CETAF Digital Working Group (DWG<sup>1</sup>). It serves as a practical tool which gives comprehensive information and recommendations to support NH institutions in their efforts to create their own, specific management policies on digital collections. Created on Biowikifarm, it is a continually evolving document, which can be regularly updated, adjusted and extended as necessary even now after completion of SYNTHESYS3. For this purpose, the handbook was handed over to and adopted by the DWG for further maintenance, development, and promotion in May 2017. Its text is freely available under the Creative Commons Attribution-ShareAlike License (CC BY-SA 3.0).

**Chapter A. About Recommendations: Management Policy on Digital Collections and Supplements**

- aim and rough content of the handbook
- SYNTHESYS background.

**Chapter B. Overview: Background and Aims of a Management Policy on Digital Collections**

- nature of a MPDC as well as its benefits and limitations
- measures to enhance the MPDC's acceptance
- premises for a lasting relevance and actual realisation of the specific MPDC.

**Chapter C. Contents and Suggested Structure of a MPDC**

- target groups of a specific MPDC
- clarification of authorities and responsibilities
- property rights and licensing
- digitisation approaches.

**Chapter D. Related Links and Documents**

- external websites on relevant topics
- websites with collections management policies covering digital data.

**Chapter E. Glossary**

**Chapter F. Attachment (Proposed Phrasings)**

- phrasings (sorted by topic) which may be used as templates and adapted by NH institutions to create their own MPDCs.

---

<sup>1</sup> Collections Policy Board (CPB) and Digital Working Group (DWG) are both CETAF bodies (<http://cetaf.org/about-us/cetaf-bodies>)

## NA2/objective1/task1/cmp

### Contents [hide]

#### Recommendations: Management Policy on Digital Collections

##### **A:** About *Recommendations: Management Policy on Digital Collections* and Supplements

###### A 1 Aims and Authors

##### **B:** Overview: Background and Aims of a *Management Policy on Digital Collections*

###### B 1 Reasons for Developing Management Policies on Digital Collection Data

###### B 2 Acceptance and Revision Processes of a MPDC

##### **C:** Contents and Suggested Structure of a MPDC

###### C 1 Introduction

###### C 1.1 Statement and Definition of Digital Data

###### C 1.2 Challenges (Regarding Handling of Digital Data)

###### C 1.3 Justification for Data Sampling, Digitization, and Data Storage

###### C 1.4 Justification for Provision, Facilitation, and Enhancement of Data Access

###### C 1.5 Justification for the Authority of an Institution to Set Requirements and Reserve Rights

###### C 1.6 Why is a Guideline or Policy Needed? What Are the Benefits and Aims?

###### C 1.7 Target Groups of a MPDC

###### C 2 Establishment and Formal Implementation of MPDC

###### C 2.1 Formal Implementation of a MPDC

###### C 2.2 Authorities and Responsibilities

###### C 2.2.1 Paperwork and Forms

###### C 2.2.2 Property Rights and Licensing

###### C 2.3 Advantages of Digitization

###### C 2.3.1 Research

Wiki handbook  
 "Recommendations:  
*Management Policy  
 on Digital  
 Collections*"  
 (screenshot). List of  
 contents (in parts).

## Handbook of best practice and standards for 3D imaging of NH specimens

Besides its potential for public relations and educational topics, 3D imaging of NH objects can offer a powerful tool to undertake scientific research without the need for direct examination of the physical object itself. Thereby, this approach helps to reduce the risk of damage or loss during handling or shipping of the specimens and minimises travel costs and time. Most importantly, it supports the visibility of and democratisation of access to NH collections worldwide.

Proper 3D imaging of NH objects heavily depends on the objectives and the objects themselves; each type of object or specimen may be digitised using different techniques with different workflows and protocols. In conjunction with the JRA work package (see p.10), SYNTHESYS3 has produced a handbook of best practice and standards for 3D imaging and therewith provided an overview of all relevant 3D techniques, software, and equipment. The handbook serves as a guideline allowing any NH institution to create 3D models of their collection, including recommendations on which techniques should be used for different types of objects, as well as information on existing technical standards.

An international working group of experts created two types of handbook:

- 1) an online Wiki with comprehensive, comparative information on diverse 2D+ and 3D imaging techniques realised at different NH institutions and a focus on practical aspects and workflows.
- 2) a printed handbook with more basic and background information on 2D+ and 3D imaging techniques.

The Wiki “*Handbook of best practice and standards for 3D imaging of NH specimens*” ([http://biowikifarm.net/v-mfn/3d-handbook/3d\\_Imaging\\_Handbook:Main\\_Page](http://biowikifarm.net/v-mfn/3d-handbook/3d_Imaging_Handbook:Main_Page)) is a living, publicly available document, which is hosted and administrated by the MfN, Berlin. It can be updated and extended by SYNTHESYS3 participants and approved external experts at any time, beyond the conclusion of SYNTHESYS3. Therefore it has the potential to provide the latest and most comprehensive information possible, especially on practical workflows and protocols. The MfN have arranged the sustainable continuation and maintenance of the Wiki platform. The Wiki forms facilitate structured data entry as well as provide a filter to search for applicable techniques. For the sake of clarity and comparability, each chapter covers one specific technique realised at one specific institute. 2D+ and 3D imaging techniques covered are, for example, focus stacking, photogrammetry, and  $\mu$ CT scanning.

The screenshot shows the SYNTHESYS3 Wiki interface. At the top, there is a blue header with the SYNTHESYS3 logo and navigation links for 'English', 'Log in', and 'Request account'. Below the header, the page title is 'Erweitert Discussion' with 'View', 'View source', and 'History' buttons. The main content area is titled 'Contents' and contains instructions for creating a page. Below the instructions is a 'Create or edit' button and a list of categories: 'Handbook' (with sub-items: Protocols, Software, Techniques, Workflows, Link collection) and 'Digitization techniques' (with sub-items: 2D+ (Focus Stacking) RBINS & RMCA, 2D+ (Focus Stacking): Zoosphere MIN, Focus Stacking + Photogrammetry RBINS & RMCA, Infrared Scanners RBINS & RMCA, Micro-CT HCMR, Micro-CT Large RBINS, Micro-CT MIN, Micro-CT Small RBINS, Photogrammetry MIN, Photogrammetry RBINS & RMCA, Structured Light Scanning 2 RBINS & RMCA, Structured Light Scanning RBINS & RMCA).

Wiki “*Handbook of Best Practice and Standards for 3D Imaging of NH Specimens*” (screenshot).

The printed handbook provides the more basic and background information about a technique, including tips and tricks which are likely not to change over time. Since an excellent guidebook on  $\mu$ CT scanning was already produced by the HCMR, Crete, within the scope of the SYNTHESYS3 JRA (see p.10), this handbook of NA2 solely mentions outer surface 2D+ and 3D methods. As a printed guideline, the information given can be easily consulted and cited. Its chapters cover, for example, the tool *Zoosphere*, infrared scanning, and light scanning. The manuscript can be accessed through the link <https://drive.google.com/file/d/0B2yIFo9B44xfWk1WSHqxVU51WVvk/view>. The chapters include:

- 2D+ digitisation: 2D+ focus stacking DSLR; 2D+ focus stacking microscope; Zoosphere; RTI:MiniDome,
- 3D Digitisation: photogrammetry; infrared sensors; structured light scanning; laser scanning,
- Comparison of techniques,
- Challenging materials, and
- Workflows

## Strategic priorities for barcoding of NH collections

DNA barcoding involves the standardised use of DNA sequences to tell eukaryotic species apart. The need for this approach is driven by the uncertainty as to the total number of species on earth, the small proportion of those that have yet been described, and the common difficulties encountered in identifying unknown specimens. Using expertly verified, identified material of NH collections, the construction of a DNA barcode reference library for life on earth solves these major problems. Traditionally, recovery of DNA sequence data from preserved NH collections has been hampered by DNA degradation. However, large-scale studies recently demonstrated effective recovery of DNA barcode data from NH collections by new technologies and even by traditional Sanger sequencing approaches.

Thus the aim of this SYNTHESYS3 task was to create a common strategy for prioritising the barcoding of NH collections and therewith to ensure a coordinated approach and action across Europe. For this purpose, the work involved collaboration with members of iBOL (International Barcode of Life, [www.ibol.org/](http://www.ibol.org/)) and a workshop with delegates of NH institutions as well as key iBOL members.

Survey work showed that while there are a lot of active DNA barcoding programs in Europe, some even on a national scale, no institution had identified priorities for the barcoding of their collections except for collection-based barcoding within individual research projects.

A session on DNA barcoding NH collections was co-organised and held at the iBOL Conference in Guelph in August 2015. Like a subsequent workshop, also in August 2015, it focused on the retrievability of DNA sequences from NH collections. In September 2015, a workshop was held at the RBGE to outline strategic priorities for DNA barcoding NH collections.

The main result of these activities is the report "[Strategic Priorities for DNA Barcoding Natural History Collections](#)", which explores the feasibility and strategic use of NH collections for DNA barcoding and wider genomic approaches for species discrimination. It gives, amongst others:

- (i) a summary and key recommendations,
- (ii) an overview of DNA barcoding,
- (iii) a statement on the role of NH collections in DNA barcoding,
- (iv) factors influencing the recoverability of Sanger sequenced barcodes from NH collections,
- (v) DNA barcoding and NGS technologies, and
- (vi) strategic priorities for DNA barcoding NH collections.

The report includes an overview of ongoing major initiatives among SYNTHESYS3 institutions as well as case study examples of DNA barcoding being applied to NH collections.

Resulting strategic priorities for DNA barcoding of NH collections include:

- sequencing of type specimens to formalise the link between scientific names and corresponding sequences,
- targeted sequencing to fill gaps in phylogenetic coverage,
- DNA barcode surveys of large genera to facilitate taxonomic revisions of difficult groups,
- DNA barcoding of endangered species,
- DNA barcoding of key pollinator species,
- sequencing invasive non-natives, pests, and pathogens, and
- DNA barcoding species with impact on human health.

Practical considerations for sample selection and design of barcoding projects are given and discussed.

A further SYNTHESYS3 workshop "*DNA Sequencing Natural History Specimens Using New Sequencing Platforms and Protocols*" was held at the RBGE in July 2017. It focused on data types which can be routinely recovered from what type of specimens and on which protocols and platforms prove to be most efficient. The first workshop day was organised as open symposium and brought together key leaders in the field of DNA sequencing NH collections as well as the related field of ancient DNA research. It was live-streamed to delegates in 13 countries, with 85 delegates from six countries attending the symposium in person. On the second day, the invited speakers and five other experts met to exchange ideas. One outcome of these

discussions was to build on the content of the above report and produce a paper capturing the current state of play for DNA sequencing NH collections with a focus on the step-changes possible by using next-generation massively-parallel sequencing platforms. At the workshop the first draft of this paper was produced, including key ideas and practical suggestions (manuscript in preparation).

An additional output of the task's work is a paper focusing on plants, published in the *Philosophical Transactions of the Royal Society*. This includes an overview of the development of barcoding protocols and direction in light of NGS technologies and the potential for these approaches to be applied to herbarium specimens:

HOLLINGSWORTH PM, LI D, VANDERBANK M, TWYFORD AD. 2016. *Telling plant species apart with DNA: from barcodes to genomes*. *Philosophical Transactions of the Royal Society, Series B*. [371: 20150338](#).

## Strategic priorities for DNA library creation of NH collections

As a result of the JRAs in SYNTHESYS2 (project 226506, 2009-2013), DNA libraries were suggested as a means to perpetuate DNA from rare specimens in NH collections. Therefore, the feasibility to create and use DNA libraries at institutional scale by European NH institutions as well as to develop common protocols for their integration was to be assessed by SYNTHESYS3 NA2. For this purpose, a two-day workshop "[Develop Strategic Priorities for DNA Library Creation of NH Collections](#)" was held in February 2016 at the MfN, Berlin. The workshop [agenda](#) included two main aims: to obtain an overview of DNA library technologies (including costs and challenges) and to discuss priorities and integrated strategies for DNA library creation of NH collections.

Based on the workshop results, it was confirmed that DNA libraries are an essential part of the NGS workflow and that various alternative protocols with specific advantages and disadvantages exist. It was noted that technological change is rapid while the costs remain significant. Due to costs and technological issues, it was concluded that the systematic DNA library creation from NH collection samples independent of specific research projects is not worthwhile. However it was emphasised that DNA libraries which are created during research projects and which involve specimens from NH collections should be databased and stored in DNA banks.

As one significant result of the workshop and jointly with the Global Genome Biodiversity Network (GGBN, [www.ggbn.org](http://www.ggbn.org)), a SYNTHESYS3 pilot study was launched by the Royal Botanic Gardens Kew (RBGK, lead), Botanischer Garten und Botanisches Museum Berlin (BGBM), and MfN. It aimed to explore the integration of metadata on DNA libraries into the GGBN data platform and thereby to close a significant gap regarding the sustainable archiving of genetic/genomic materials derived from NH collections. However, High Throughput Sequencing (HTS) developments have outstripped current plans within SYNTHESYS3 and GGBN. HTS libraries, which are developed at great expense, run the risk of being single use – only for the current project they were developed for. There is potentially a great deal of information available within these libraries besides the use they were created for if there was a means of making this information more widely available. This short-term pilot project was developed to address this perceived gap, making these data readily available to other researchers through the GGBN data portal. In many instances even the knowledge that there was a library from a particular organism was not available outside of the group.

The pilot study on the availability of HTS libraries developed 142 terms to represent six use cases from three SYNTHESYS3 institutes: RBGK, NHM, and BGBM. These are not regarded as a final set, but are understood to be a working group of terms, which will change as more use cases are made available or new protocols come online (see for example Brace, S. *et al.* 2016 and DENTINGER, B.T.M. *ET AL.* 2016).

The selected terms were used to generate a prototype based on the GGBN Data Standard ([https://terms.tdwg.org/wiki/GGBN\\_Data\\_Standard](https://terms.tdwg.org/wiki/GGBN_Data_Standard), DROEGE *ET AL.* 2016). A use case collection from animals, diatoms, fungi and plants has been developed and is available in the GGBN Sandbox:

[http://sandbox.ggbn.org/ggbn\\_sandbox/search/result?institution=UseCaseCollection%2C+GGBN](http://sandbox.ggbn.org/ggbn_sandbox/search/result?institution=UseCaseCollection%2C+GGBN), see also the GGBN Wiki: [http://wiki.ggbn.org/ggbn/Use\\_Cases\\_HTS\\_library\\_samples](http://wiki.ggbn.org/ggbn/Use_Cases_HTS_library_samples). These examples will be improved and kept stable until at least 2020, enabling interested parties to see further development and contribute to this.



## Protocols for data collection from DNA extraction

A main objective of institutions with NH collections and holding biobanks (that is, tissue and molecular or DNA collections) is to preserve molecular genomic biodiversity (as both physical samples and associated data) and to maximise the accessibility of these new collections. To meet the expected quality criteria of a biobank the curatorial work within the respective collections should be done by specialists and professional staff. Furthermore the associated data should be recorded and sustainably stored in a database which follows defined standards.

However depending on the type of facility and on the objectives of the respective activities, data captured in the databases may differ greatly in terms of quantity and content. For long-term preservation of collections the main aim in data basing is to retain information related to the preservation, the extraction method, and the access for future uses, while the techniques applied (concerning, for example, amplification or genotyping) might not be as relevant. In contrast, for short-term, objective-driven research projects, retaining information on the preservation of molecular extracts and related information (like geo-reference data) may not be a priority and such data may not be included in the databases. However, this information would be of vital importance for NH collections and their biobanks, which often adopt and preserve orphaned molecular samples after conclusion of the original research project.

The SYNTHESYS3 team working on this task aimed to quantify and get an overview of the tissue and molecular collections (that is, NH biobanks) at the SYNTHESYS3 institutions and to investigate which kinds of associated data are retained and stored alongside the physical samples. Based on these findings the task also aimed to develop protocols for collecting data from sequencing activities on NH collections and to feedback on the success of different methodologies.

To reach the task's aims, an online survey on existing DNA and tissue collections (or biorepositories) at institutions housing NH collections as well as on current practices regarding the gathering, storage, and the accessibility of associated data was created and distributed among SYNTHESYS3 member institutions. At the same time, further information was gathered on current methods for DNA and RNA isolation, which would be compatible with Next-Generation Sequencing (NGS) techniques. In order to make a more compelling analysis of methods and protocols currently in use, new techniques for probe sequencing for genomes were taken into account. The search for information was realised through literature studies, online searches, and through personal communication with laboratory staff. Several institutions also provided information on the number of classical collection specimens sampled, the preservation method of these specimens, extraction and amplifications methods, and resulting papers.

The results of the online survey and the more general search for information were analysed and summarised in a [report](#). Over 90% of the surveyed SYNTHESYS3 institutions with biobanks currently provide free of charge access to their collections for non-commercial research worldwide. However, only half of the collections are actually visible and explorable by full digitisation and accessibility through external data portals or links with important databases like GBIF or GenBank. The current data situation concerning taxonomy, georeferencing and basic institutional identifiers is (in most cases) satisfactory to good. However, there are deficiencies regarding data on, for example, provenance, compliance, and links to results and publications.

Based on these results, the report proposes several common fields or kinds of data which should be routinely collected and stored along the physical samples. These are, for example, data providing information on unambiguous identity and access, geography, legal origin and on preparation and preservation methods. General issues and standards concerning the respective databases are addressed, such as the realisation of regular quality checks and backups. Recommendations on the treatment of DNA extracts and its associated data and protocols are given.

## Network of DNA and tissue banks and developing policies for new physical collections management

NH institutions share a common interest in long-term preservation of genomic samples that represent the diversity of non-human life on Earth. Genomic research depends on reliable and rapid access to high-

quality, well-documented DNA and tissue samples that have been legally obtained. Despite the importance of these samples, information about them has often been unavailable or fragmented across the diverse repositories. No central registry of repositories was available to simplify the search for genomic samples. The wasteful collection of new samples and the slow-down or even halt of research projects were the results of this fragmentation.

SYNTHESYS3 aimed to solve these difficulties and develop a network of DNA and tissue banks. However in the time since this original objective was written, this had been superseded by developments elsewhere: the Global Genome Biodiversity Network (GGBN, [www.ggbn.org](http://www.ggbn.org)) was set up in 2011 and in the meantime progressed so far that it made the original SYNTHESYS3 task superfluous. A considerable number of SYNTHESYS3 members were deeply involved in GGBN and were either in the process of or considering joining GGBN. As a logical consequence, the aims of the planned SYNTHESYS3 tasks were adjusted to the new developments: SYNTHESYS3 beneficiaries not already members of GGBN were to be encouraged to join GGBN. Also, SYNTHESYS3 aimed to collaborate with GGBN to further develop the existing network of DNA and tissue banks in Europe. This also meant to enable the easy access and exchange of DNA and tissue samples, to coordinate the use of protocols and thereby to improve the storage of and access to collections, and to enhance the information system for a better coverage of links between specimens, tissue collections, and DNA collections.

GGBN provides genome-quality samples and open access to a global data management system hosting aggregated primary specimen data and associated metadata for all member institutions. It also provides members and the scientific community at large with a platform for biodiversity biobanks from across the world to:

- collaborate to ensure consistent quality standards for DNA and tissue collections,
- improve best practices for the preservation and use of such collections,
- harmonise exchange and use of genetic materials in accordance with national and international legislation and conventions, and to
- enable targeted collection of samples.

As of June 2017, GGBN has 61 members of which 20 are core members, 23 associates, and 18 observers. It provides genome-quality samples and open access to a global data management system hosting aggregated primary specimen data and associated metadata for all member institutions. In terms of taxa and to date, these are 38,816 species and 2,819 families (as of August 2017).

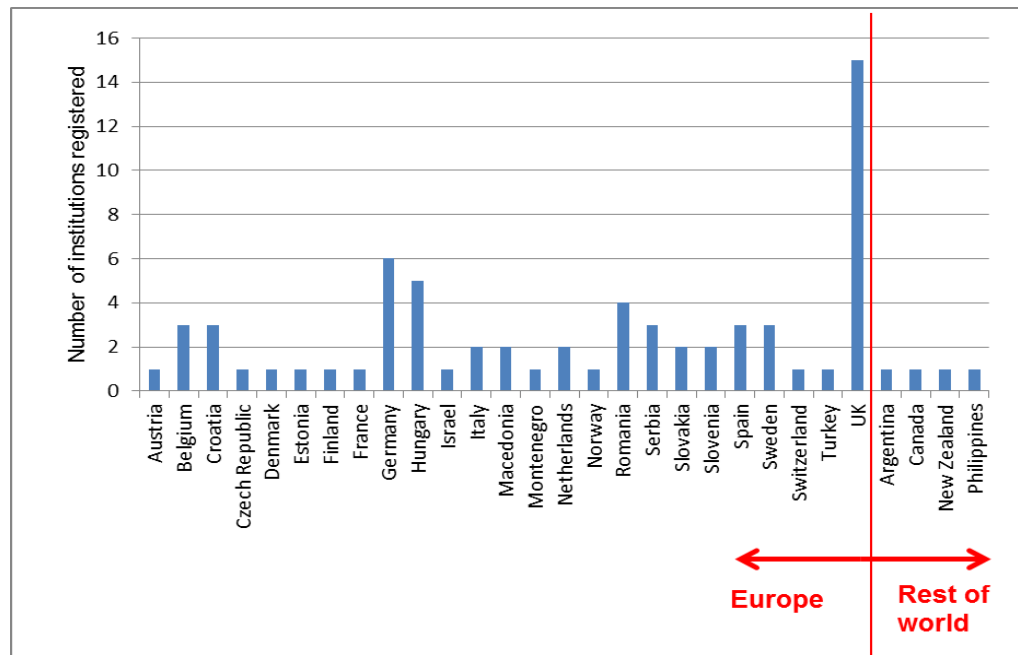
Through discussions on best practices, the code of conduct and the development of material transfer agreements (aligned with those of CETAF), SYNTHESYS3 significantly contributed to the further improvement and strengthening of GGBN. Also, the promotion and encouragement of SYNTHESYS3 beneficiaries to join GGBN was a key factor in establishing Europe as the best represented region in GGBN.

## Collections self-assessment provision

The SYNTHESYS [Collections Self-Assessment Tool \(CSAT\)](#) is a web application developed during SYNTHESYS2 (project 226506, 2009-2013) offering a free, simple to use online self-assessment for NH collections. The questions cover infrastructure (institution and building maintenance) and collections (management, care, access, molecular and digital collections). The tool identifies areas for improvement and provides advice on where improved standards can be adopted. The results of the assessment allow the respective institution to prioritise and implement changes, for example to enhance the collections accessibility. The valuable information gained can be used by the institution for planning management and also when seeking funding. Besides, the tool allows SYNTHESYS to identify key problem areas in collections management and collections management needs across Europe and beyond. Information on these was to be passed on to the CETAF Collections Policy Board.

The SYNTHESYS CSAT was presented on numerous occasions and successfully promoted across Europe and beyond. Dissemination activities were driven forward and the CSAT now has a total of 70 registered institutions.

*Geographical spread of institutions registered to use CSAT to date. Whilst there are registrants in Argentina, Philippines and New Zealand, the main focus of use remains within Europe, particularly the UK*



## Integration of new collections management approaches into the CSAT

During SYNTHESYS3 the range of CSAT benchmarks was expanded to enable institutions to assess the management of their digital and new physical collections. As a result the newly re-launched version of the CSAT comprises a new section on the management of molecular collections (such as collections of DNA, tissues, and proteins) with 12 new benchmarks as well as a new section on the management of digital collections with nine new benchmarks. Both sections do not aim at specialised institutions and specialist staff but rather target institutes with biodiversity collections such as museums and herbaria and areas under the responsibility of non-specialist collections staff.

In addition, clarity of language and post-survey help comments were updated to include other SYNTHESYS3 outputs, for example the recommendations on Management Policies on Digital Collections. The scoring system was successfully modified. Following incorporation of feedback by members of SYNTHESYS3 organisations and by institutions which had completed the SYNTHESYS2 CSAT and did a re-survey, a new version of the CSAT was successfully launched in October 2015.

Additionally to the development and launch of a new version of CSAT, three other collections assessment platforms available to the museum and heritage community were reviewed and compared with each other and with the CSAT. It became evident that the CSAT has many strengths, for example regarding its practicability and relevance for the whole institute as well as the staff working directly with the collections, its suitability for institutions of different sizes, costs and efforts to be spent, and the experience behind its development.

The following **NA2 deliverables** were realised (all available on the NA2 pages of the SYNTHESYS website: <http://www.synthesys.info/network-activities/synthesys3-na2/>):

Title	Explanation	Lead	Link
"Virtual Collections" Management Policy Meeting	meeting to discuss virtual (digital) collections management policies	MfN	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.1-Virtual-CM-policy-meeting.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.1-Virtual-CM-policy-meeting.pdf</a>
"Virtual Collections" Management Policy	finalised policy for virtual (digital) collections management available	MfN	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del.2.2-Virtual-Collections-Management-Policy.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del.2.2-Virtual-Collections-Management-Policy.pdf</a>
New Physical CM Policy Meeting	meeting to discuss new physical collections management policies	UCPH	<a href="http://www.synthesys.info/wp-content/uploads/2014/01/NA2-Del.-2.3-New-physical-collections-management-policy-meeting.pdf">http://www.synthesys.info/wp-content/uploads/2014/01/NA2-Del.-2.3-New-physical-collections-management-policy-meeting.pdf</a>

Best Practice Handbook	handbook on best practice in 3D imaging for natural history collections	MfN → RBINS	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.4-Best-practise-handbook.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.4-Best-practise-handbook.pdf</a>
New Physical Collections Policy	finalised policy for new physical collections management available	UCPH	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/SYNTHESYS3_NA2_Deliverable_2.5_report.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/SYNTHESYS3_NA2_Deliverable_2.5_report.pdf</a>
Strategic Priorities for Barcoding	report on the strategic priorities for barcoding of natural history collections	NBC → RBGE	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.6-Strategic-Priorities-for-DNA-Barcoding-Natural-History-Collections.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.6-Strategic-Priorities-for-DNA-Barcoding-Natural-History-Collections.pdf</a>
DNA Library Construction Workshop	workshop on the construction of DNA libraries	MfN	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.7-Report-DNA-library-construction-workshop.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-2.7-Report-DNA-library-construction-workshop.pdf</a>
Protocols for DNA Extraction	protocols for DNA extraction made available	CSIC	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/DELIVERABLE-2_8-Protocols-for-DNA-extraction_230117.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/DELIVERABLE-2_8-Protocols-for-DNA-extraction_230117.pdf</a>
Updated CSAT	updated Collections Self-Assessment Tool (CSAT) launched	NHM	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del.2.9-Updated-Collections-Self-Assessment-Tool.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del.2.9-Updated-Collections-Self-Assessment-Tool.pdf</a>

## NA3: Innovation, Impact and Sustainability

### External experts and relationships with industry

During the course of the project the NA3 team have worked with external experts to review technological development and innovation and act as ambassadors to promote SYNTHESYS3 work. These include:

- **Debbie Paul** of US-based iDigBio initiative (Integrated Digitised Biocollections, [www.idigbio.org](http://www.idigbio.org)). Debbie brought knowledge of the North American network of collections digitisation activities to SYNTHESYS and in return was able to relay results back to iDigBio, strengthening US-European cooperation and maximising efficiency between these two major projects.
- **Pieter Holtzhausen** (Stellenbosch University, South Africa) is a specialist in automated image processing algorithms and has been instrumental in advising on the SYNTHESYS3 image segmentation software Inselect developed at NHM London. Peter co-authored the associated [paper on Inselect](#) that was recently published in PLoS One.
- **Paul Kimberley** manages various digitisation and informatics projects at the Smithsonian Natural History Museum. During a visit to the NHM London discussions were held on topics directly relevant to SYNTHESYS3 work including collections assessment, DoD, mass digitisation projects (prioritisation, planning, preparation, workflows and outputs), digital policies and standards, digitisation innovation areas (3D, spirit collections, data/text mining/OCR, applied computer vision, robotics, industrial / commercial collaborations, citizen science), linked open data and digitisation infrastructure.
- **Roderic Page** of the University of Glasgow was a key figure in the “Pit Stop” event co-organised by NHM and Cisco Systems Inc. aiming to tackle specific challenges through interactive debates and workshops that give new ideas and approaches (see below). Rod helped conceptualise the event and was a key participant, speaking and leading breakout discussions.

SYNTHESYS3 networking with industry led to three major activities with industrial partners: 1) digitisation pilot activities with the SME (small-medium enterprise) Picturae based in the Netherlands; 2) a “Pit Stop” event co-organised by NHM and Cisco Systems Inc. and facilitated by the Digital Catapult Centre; and 3) promotional opportunities with digitisation SMEs at conferences and meetings.

### *1) Picturae Digitisation Pilot*

One of the key digital objectives for natural history (NH) collection holding institutions is to pilot priority workflows in readiness for large-scale digitisation. A proposal was developed by RBGK and the NHM London to digitise Solanaceae (nightshade family of plants including potatoes) and Dioscoraceae (family of flowering plants including the yam) collections to test a scalable, high-throughput digitisation workflows for herbarium sheets. This provided a small scale test (110,000 specimens) of one of the fastest throughput digitisation workflows developed to date, piloting the SME Picturae conveyor-belt process used by Naturalis in the Netherlands to digitise their entire herbarium. The project aimed to be exploratory in its nature, looking at whether outsourcing of imaging and data transcription is a suitable model for museum herbarium sheets across the whole SYNTHESYS3 consortium. The pilot involved:

- Developing a tailored workflow for high throughput digitisation of herbarium sheets including curatorial preparation, pest management, shipping, imaging, barcoding, databasing, digital storage, transcription, return of data and images, validation and QA, incorporation into institutional systems;
- Identifying digitisation challenges and develop existing Picturae process to address them;
- Trialling a collaboration as a means of sharing curatorial, research and digitisation practises;
- Assess the suitability of the Picturae process as a long-term method of digitisation.

In addition to the successful digitisation of the collections material, the pilot led to the publication of a detailed lessons-learnt report, outlining the challenges of working with third-parties to develop mass digitisation workflows. The report was circulated and provided valuable lessons that can be embedded within future mass digitisation projects.

### *2) Cisco-Pit Stop, 25 - 26 February 2016*

The NH collections community have many established peer-to-peer collaborations: both SYNTHESYS and iDigBio are good examples of large-scale collaborative projects involving multiple large organisations. However, as a community we are less able to effectively engage with larger industrial partners, SMEs, start-ups and innovators. In order to address this problem SYNTHESYS3 organised a “Pit Stop” event co-organised by NHM and Cisco Systems Inc. and facilitated by the UK-based Digital Catapult Centre. The overall aim of this event was to create opportunities for participants to explore new partnerships, commercial collaborations and innovative services.

Cisco and the NHM identified key challenges for the event revolving around metadata transcription and data mobilisation / extraction, as well as tacking NH data quality and linking. A total of 60 participants from 31 organisations of multiple European countries took part in the Pit Stop. As a direct result of the event 13 proposals from SMEs were received, three of which have been taken forward.

The Pit Stop event format could be used by other SYNTHESYS partners to work with industry. In addition to presentations from SYNTHESYS3 members, there were numerous subsequent presentations and follow-up activities (for example Debbie Paul presentation: Cisco Pit Stop: [Digitising the Natural History Museum’s collections](#), 12 April 2016) and a number of blog posts were written during and after the event, highlighting the model and acting as a general call for innovators to become more involved with the challenges associated with NH collections digitisation.

### *3) Promotional opportunities with digitisation SMEs*

As part of a number of SYNTHESYS3 events, SYNTHESYS3 members and institutions have been reaching out to SMEs to encourage participation within conference activities. Historically this is unusual for the NH community because of the perceived risks associated with appearing to endorse one or other commercial activity to the disadvantage of another. By preparing careful guidelines to SMEs SYNTHESYS3 sought to break this taboo at the 2016 Society for the Preservation of Natural History Collections (SPNHC, [www.spnhc.org](http://www.spnhc.org)) meetings in Berlin, inviting key SMEs to participate in the meeting and showcase what they can offer the community.



The SYNTHESYS3 symposium at the SPNHC meeting included two talks from SMEs on high-throughput industrial digitisation services (*Witikon* and *Digitarium*) and a view on industry (Cisco) on the benefits of collaborating with NH collections:

- [Automating Insect Digitisation - Speed and Costs](#) - Riitta Tegelberg, Janne Karppinen, Zhengzhe Wu, Jere Kahanpää & Hannu Saarenmaa
- [WITIKON: Mass 3D digitisation at a national scale](#) - Mira Silanova
- [A Bridge from Enabling Infrastructure to Digitisation Priorities, a view from Industry](#) - Deb Paul, Philip van Heerden, Vince Smith, Laurence Livermore & Ehsan Alavi Fazel

## Optimal digitisation technologies and equipment

Information has been gathered from all SYNTHESYS3 partners on their current digitisation facilities, equipment and workflows, identifying the biggest challenges faced and how digitisation of collections is being prioritised. This survey, first of all, showed that approaches to digitisation vary greatly over the NH institutions. It is also apparent that only a small part of those collections has been digitised so far and that most organisations still prefer in-house digitisation, for various reasons, rather than outsourcing the whole process. The resulting [report](#) acknowledges the wide variety of approaches, requirements and equipment needs, recognising that “a one size fits all” approach is not a viable solution. However, a number of common themes were present enabling the delivery of a list of key recommendations.

Common themes on digitisation facilities and prioritisation:

- Institutions house between five and six types of collection on average, with an average collection size of around 6 million. Prioritisation for collection digitisation is 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.
- Approximately two thirds of the institutions performed at least some minimal curation, specimen conservation or collections management steps prior to digitisation.
- There is a vast range of digitisation equipment in use, reflecting the wide variety of collections in the different institutions.
- Most institutions did not appear to follow any official imaging standard and it appeared that their awareness of such standards was under-developed.
- The majority of institutions are still capturing full specimen metadata prior to the imaging step in their main digitisation workflows.
- The main reported factor limiting the rates of digitisation was not related to equipment but to the challenge in securing funding for digitisation and a lack of available human resource.
- Suitable digitisation infrastructure was seen as a greater impediment to digitisation rates than the lack of suitable digitisation equipment.
- 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 would 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 for achieving optimal digitisation technologies and equipment:

- 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, stakeholder 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 – there are many different types of equipment and workflows that are currently in place within the various institutions. In the USA iDigBio is coordinating the national digitisation effort through the Resource for Advancing Digitisation of Biodiversity Collections (ADBC) program. Activities include collating 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](https://www.idigbio.org/wiki/index.php/Digitization_Resources)). The SYNTHESYS3 report recommends that institutions in the EU also share their information and workflows in a shared repository so that lessons learned are shared throughout the EU.
- Set of harmonised guidelines and standards be agreed upon in order to set a benchmark for quality assurance of digitisation procedures. There may be potential to implement image checks against image standards using automated software to speed up the process of QA.
- Successful adaptations to workflows that increase digitisation rates should be disseminated more widely so they are more likely to be implemented in other institutions.

## Optimal crowdsourcing processes for NH institutions

A further piece of work completed under NA3 was a study into optimal crowdsourcing processes for NH collections, a highly productive way to increase active public engagement in science alongside large-scale data gathering. This consisted of a study into the target audiences that are likely to engage in crowdsourcing and identify the most appropriate categories of project for crowdsourcing efforts in the context of NH collections.

An inventory of current practices in crowdsourcing and more general citizen science in NH contexts was made, resulting in 62 relevant projects distributed over various countries. By understanding crowdsourcing as one subset of engaging volunteers, different forms of volunteering could be distinguished with related organisational issues, (skill-based) types of volunteers, and types of projects. A study about motivation of crowdsourcers and citizen scientists using relevant literature and a series of interviews has given a first insight into what motivates participants of crowdsourcing projects and how to strengthen their link with volunteer activities. A key finding is that belonging to a community is a critical driver and these community and social interactions are critical to the successful organisation of crowdsourcing projects.

The report on crowdsourcing investigated the methods that have been successfully used across different scientific communities and have the potential to become an effective method for transcribing and georeferencing NH data. Current crowdsourcing platforms and communities such as Atlas of Living Australia, DigiVol, herbaria@home, Les Herbonautes, Notes from Nature and the Smithsonian Digital Volunteers Transcription Center have been reviewed and their potential role in improving digital access to NH collections worldwide discussed.

Based on the feedback gathered from crowdsourcing case studies, experience from running citizen science projects and guidance issued by some of the platforms, some general recommendations for creating effective crowdsourcing projects were summarised:

- Each project should have a clear design rationale with cultural or scientific benefits;
- Projects should be actively promoted (PR strategy/comms plan);

- Institutions should make it clear how and when the data will be publically available through an open access policy;
- Projects should be broken up into discrete parts - volunteers enjoy completing projects rather than having endless tasks
- Platforms should utilise existing data standards where possible ([TDWG/ISO](#));
- Ensure there is adequate help and documentation on both platform and project level;
- Always display specimens along with labels where possible (volunteers enjoy seeing specimens);
- Record and analyse project activity (e.g. using Google Analytics);
- Allow volunteers to work on projects and tasks of interest (e.g. specific location or taxa);
- Provide post-project updates (e.g. resulting papers, data sets, public outreach etc);
- Allow volunteer communities to establish and support themselves (e.g. the wiki/forums model of herbaria@home and object-orientated forums within Zooniverse);
- If appropriate and feasible, consider differentiating between users of different skill/experience levels (e.g. reviewing contributions or harder tasks);

## Improved access to new collections

SYNTHESYS3 aimed to improve access to virtual and new physical NH collections by making all data outputs freely available and open access.

Consejo Superior de Investigaciones Científicas (CSIC) staff prepared a document presenting [guidelines](#) to sharing data via four major data aggregators:

- 1) Europeana (<http://pro.europeana.eu/>)
- 2) Global Biodiversity Information Facility (GBIF, [www.gbif.org](http://www.gbif.org))
- 3) The European Bioinformatics Institute (EMBL-EBI, [www.ebi.ac.uk](http://www.ebi.ac.uk))
- 4) International Society for Biological and Environmental Repositories (ISBER, [www.isber.org/](http://www.isber.org/)).

### SYNTHESYS 3 Synthesis of Systematic Resources

a Seventh Framework Programme funded project

Project Number 312253

Date: February 23, 2016

Work package number: WP3 Innovation, Impact and sustainability

Objectives 2: Impact

Task 2.2: Improved access to collections and contribution to ERA

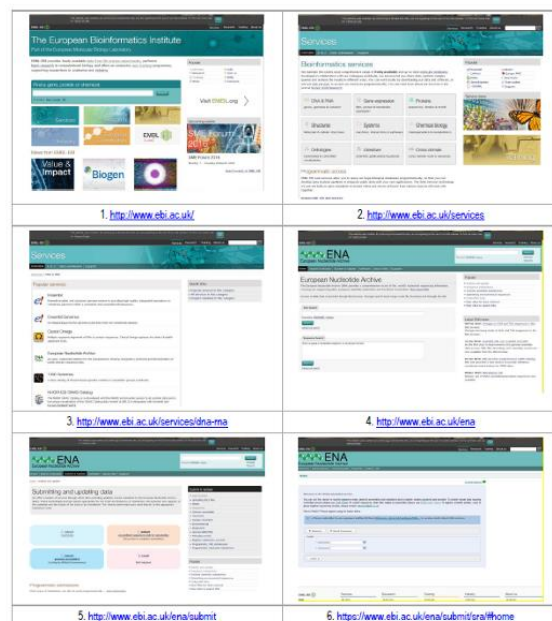
Partner: MNCN-CSIC

Reporter: Celia M. Santos Mazorra, Isabel Rey Fraile, Manuel Sánchez Ruiz, Antonio Gª Valdecasas, Josefina Barreiro Rodríguez & Marian Ramos



#### 5.1.- Workflow to submission of sequence information

To make raw sequencing data accessible through ENA, you can follow this path



Note: All the links within this document are active at February 20, 2016

9

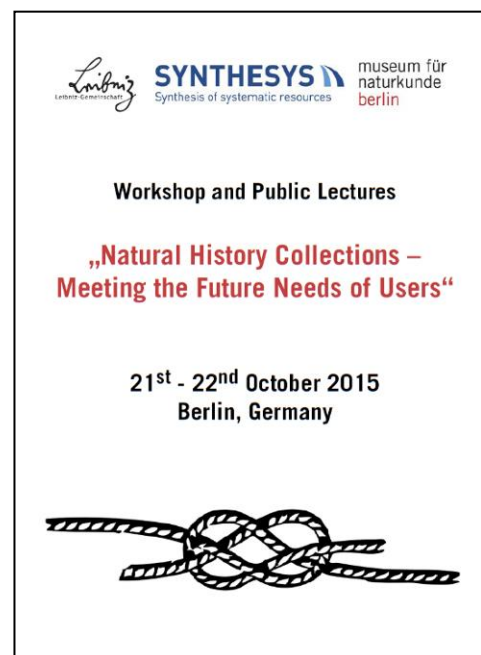
Screenshots from guidelines to sharing data via four major data aggregators.

## A sustainable future and a European roadmap for NH collections

The SYNTHESYS3 consortium recognises the importance of the societal challenges identified by the Horizon2020 programme and relevance of NH collections for solving problems outside traditional and conventional fields. The consortium therefore acknowledges the need to provide for new uses and new user groups of NH collections. In order to discuss and explore these topics a SYNTHESYS3 workshop took place at the MfN Berlin in October 2015: [“Natural History Collections – Meeting the Future Needs of Users”](#).

One fundamental aim of the workshop was the compilation of traditional and current uses of NH collections as well as the identification of non-conventional potential future uses and user groups. Examples for non-conventional uses relate to the fields of e.g. agriculture and plant protection, human health and epidemiology, environmental monitoring and estimation of anthropogenic impact, biodiversity distribution modelling, archaeology, biomimicry and bionics, arts and design, social sciences and wildlife protection regulations. Experiences with innovative approaches and projects which benefit from NH collections or even heavily rely on these were presented, discussed and analysed at the workshop. Besides evaluating the potential of NH collections, possible challenges and limitations were outlined and discussed. Possibilities of collections management to respond appropriately to these needs were evaluated and ways discussed how to meet the requirements of new user groups. Another aim of the workshop was to define means to promote innovative uses and to reach potential new user groups. Possibilities to communicate the diverse, broad application fields of NH collections and to promote their importance for society as a whole were outlined.

- Dr. Vincent Smith (Natural History Museum, London): *“Building on the successes of SYNTHESYS: creating sustainable access and business models for natural history collections”*;
- Dr. Sandra Knapp (Natural History Museum, London): *“Collections for understanding environmental change”*;
- Prof. Dr. Günter Rolf Fuhr (Fraunhofer-Institut für Biomedizinische Technik IBMT, Sulzbach): *“Low-temperature biobanking – possibilities, potentials and limitations of cryopreservation”*;
- PD Dr. Jens Amendt (Universitätsklinikum Frankfurt der Goethe-Universität, Institut für Rechtsmedizin, Frankfurt am Main): *“Interactive natural history collections as a possible tool in forensic and medical entomology”*;
- Prof. Dr. Herbert D.G. Maschner (Center for Virtualisation and Applied Spatial Technologies, University of South Florida, Tampa): *“Open access natural history and the democratisation of science”*.



Sustainability workshop speakers

During the second workshop day participants exchanged experiences, opinions and ideas on:

- I. Current and future uses and user groups
  - a. Overview of current uses and user groups
  - b. Future uses and user groups as well as aims and purposes
  - c. Needs (e.g. infrastructure, access, data, services)
- II. Implications
  - a. Benefits (for NH collections community)
  - b. Unmet needs, barriers, challenges, limitations
  - c. Strategies to overcome these: joint actions
  - d. Cost-benefit analysis

### III. Outreach mechanisms (joint actions)

- a. Priority list of new user groups
- b. Strategy to approach them
- c. Future actions: promotion

The workshop concluded that virtual or digital collections increasingly are becoming an important service. However since digital collections cannot provide all information needed by researchers, physical collections remain to be the main source of information. It was agreed that the research community remains the core user group. Nevertheless a need for widening the user base of the NH collections was acknowledged: new user groups and uses should be considered more strongly and access to NH collections for these groups facilitated and promoted in order to reflect the current challenges of the Horizon 2020 work programme. As main guiding principle towards a “customer first approach” it was suggested to consider a re-orientation of institutions housing NH collections towards putting priorities on customers first and to analyse trends of customers’ interests and means for their satisfaction.

Following recommendations of the sustainability workshop the SYNTHESYS3 consortium agreed it should gather robust documentary evidence how the NH collections, their expertise and services can contribute to solutions to a range of societal challenges. A set of 26 relevant use cases addressing these urgent challenges was developed, providing an overview of type of data and skills involved in these projects. Furthermore potential user groups (collaborators and beneficiaries) who would be stakeholders within this research have been identified.

SYNTHESYS3 reviewed these 26 use cases backed up by documentary evidence demonstrating how the consortium’s collections, expertise and services can be used to find solutions to a range of societal challenges. The use case review has been made available online in a format which allows updating and the inclusion of additional examples. It constitutes the first element for building a [roadmap](#) for the access policy of the European NH institutions. This will enable further discussion on the prioritisation of solutions by a wider community.

#### **Example user groups of NH collections:**

- Environment:
  - Urban planning;
  - Environmental impact assessment;
  - Deep sea mining;
  - Conservation planning and monitoring;
- Agriculture:
  - Identification;
  - Future domestication;
  - Land use change;
  - Industrial insect farming;
  - Forestry;
- EU border control:
  - Invasive species and pests;
  - CITES;
- Health:
  - Pathogen identification;
  - Medicine and food supplement verification;
  - Pharmaceutical industry;
- Biobanking:
  - Research;
  - Government;
  - Industry (medicine, biotechnology);
- Forensic ecology;
- Arts;
- Cultural history and archeology.

Both the SYNTHESYS3 sustainability workshop and European roadmap have significant impact on future development in the European NH community. They have provided documentary evidence and recommendations on new uses of NH collections. They also represent a critical contribution to preparatory phase of the DiSSCo initiative (Distributed System of Scientific Collections) developed under the European Strategy Forum on Research Infrastructures (ESFRI, <http://esfri.eu>, see below).



## Facilitating access beyond SYNTHESYS3

### A new pan-European research infrastructure for NH collections

As an advanced community the SYNTHESYS3 consortium has been working on providing sustainable post-project physical access to NH collections and it is acknowledged that this requires an intellectual effort and unified approach from Europe's NH collections-holding institutions. One of the aims of SYNTHESYS3 was to publish a strategic sustainability report including mapping out a series of national activities, building on current institutional investments, securing the commitments of national governments to commit to supporting the long-term sustainability of SYNTHESYS outputs and providing a vision of an integrated European NH collection. This work has culminated in DiSSCo: the Distributed System of Scientific Collections (<http://dissco.eu>).

The DiSSCo initiative is an ESFRI proposal for a new pan-European research infrastructure that will streamline access to collections and transform a fragmented model into an integrated European system. The DiSSCo ESFRI proposal builds on a mature community of institutions including all SYNTHESYS3 collection-holding partners. It is a collaboration of 114 national facilities in 21 countries, the largest ever formal agreement between natural science collection facilities. This strategic collaboration is underpinned by the sound governance and decision-making structures founded within SYNTHESYS.

The DiSSCo research infrastructure proposes to unify access to information, to provide massive new linked data associated with collections, and to drive policy and process harmonisation. This new infrastructure achieves the economies of scope and scale necessary to maximise impact for science and society. Through digitisation, aggregation and the linking of European collections it will be possible to draw critical new insights, enabling scientists to address some of the world's greatest challenges on a scale above and beyond what has been possible through SYNTHESYS1-3. Critically, DiSSCo brings a transfer of authority from facilities to a central hub for all key operations, a clear decision-making mandate, and binding institutional commitments as part of a new independent legal framework. While SYNTHESYS3 has been working toward these goals, it has taken the proposal of a more inclusive infrastructure and changes in institutional governance to make DiSSCo possible. Having developed this agreement during SYNTHESYS3 throughout 2015-16, the DiSSCo proposal was submitted in August 2017 for inclusion on the 2018 ESFRI roadmap.

### Effects of digitisation on physical access to NH collections

Recent development in digitisation methods and building of virtual collections in NH institutions may gradually change the face of access to NH collections and reduce the need for lengthy physical visits. NH institutions will increasingly be able to send digital surrogates that are of excellent research quality. This new form of access will be of benefit not only to the European research community but also to researchers in Third Countries who wish to undertake collaborative research with European scientists using collections based in Europe.

However the digitisation practices and amount of available digital data significantly vary across European NH institutions. A survey was carried out in order to compile evidence about how digitisation of NH collection affects physical access to and physical loans from the collections, and this was summarised in a subsequent [report](#). It has been assumed that digitisation would reduce the need for loans and visits to NH collections, but against all expectations, as digitisation has proceeded, many custodians experienced the opposite effect: when information on the collections became available more researchers became aware of what was in a given collection leading in some cases to more loans and visits.

However the overall pattern from the SYNTHESYS3 survey is a decrease in number of physical loans and visits. "Digital loans", i.e. requests for digital information, on the contrary, show an increasing tendency. There is also an indication that for vertebrate collections, especially collections of mammals, the number of physical loans does not show the same decline as seen for other types of collections.

The questionnaire replies from custodians and users do not give a simple picture and the effect of digitisation is thus not simple, but the survey showed several clear tendencies:

- Availability of digital images (online or on demand) reduces the number of *loans*.

- According to the users, but not to the custodians, availability of digital images also reduces the number of *visits*.
- The highest degree of reduction of numbers of loans and visits is seen in herbaria.
- Experienced custodians more often experience a decrease than an increase in loans and visits, whereas more junior custodians more often experienced an increase.
- Availability of digital information leads to more specific loan requests.
- Declines in loans and visits may not be caused by digitisation – other factors such as declining economy (including reduced collection staff), increased travel costs, and difficulties with international shipping of material (increased regulations) may play decisive roles.
- Digitisation cannot fully replace physical loans and visits. There is still a need of support for transnational access, such as has been provided by the SYNTHESYS programmes.

Finally, bearing in mind the finding by GBIF’s task force: “The major realised benefits of digitisation are: increased use, exposure and knowledge of the institution’s collections”, one might argue that the most important fact is that collections continue to be the essential infrastructures they are, irrespective of digitisation.

## Training for the future

During the SYNTHESYS3 project it has become clear that difficulties in the installation and implementation of software tools and resources is a major obstacle to the ability of institutes to benefit from the developments taking place. There appears to be several issues involved, and these may vary between institutes. However one of the most common issues is the lack of availability of technological expertise relating to the software.

A training workshop was jointly hosted by the EU-based SYNTHESYS3 project and the US-based iDigBio project in June 2016. A main aim of the workshop was to increase awareness of some of the digitisation tools available and to make them more accessible for institutes with more limited ICT and informatics availability. Sessions were led by NHM, Symbiota (<http://symbiota.org>) and ABBYY (<https://www.abbyy.com/en-gb/>). It was a mix of informative presentations, practical training and open discussion with an aim to make these tools more accessible to institutes of all sizes and to encourage collaboration between Europe and the wider community. NHM led the session on Insect, which currently supports automated recognition, cropping and annotation of scanned images of items such as drawers of pinned insects and trays of microscope slides. ABBYY presented ABBYY FineReader: an optimal character recognition (OCR) tool which has been found to perform well for specimens, enabling the automated capture of specimen label data. Symbiota presented its virtual platform which incorporates OCR, natural language processing (NLP) and crowdsourced transcription modules.



*Photo from JRA training workshop – Symbiota (Anne Barber). Photo by Elspeth Haston.*

## Instruction videos

SYNTHESYS3 partners prepared a series of short videos providing information and instruction on tools and protocols for the digitisation and management of new collections developed in other areas of the project:

1. [Tea-bags for collecting organic material in silica gel for DNA extraction](#)
2. [Stable URIs for specimen images](#)
3. [Digitisation software Insect](#)

These films are available on-line and disseminated via various channels.

Collecting plant material in silica gel:  
the teabag method

Dr. Peter Wilkie  
Royal Botanic Garden Edinburgh

1:07 / 7:10

Collecting plant material in silica gel: The teabag method (SYNTHESYS)

Royal Botanic Garden Edinburgh

Subscribe 1.2K

283 views

Views from the SYNTHESYS training video on collecting plant material in silica gel

The following **NA3 deliverables** were realised (all available on the NA3 pages of the SYNTHESYS website: <http://www.synthesys.info/network-activities/synthesys3-na3/>):

Title	Explanation	Lead	Link
Promotion NA2	Promotion of the NA2 outputs	NMP	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del.3.1-Promotion-and-Outputs-for-NA2.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del.3.1-Promotion-and-Outputs-for-NA2.pdf</a>
Promotion JRA	Promotion of the JRA outputs	NMP	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Del.3.2-Promotion-and-Outputs-for-JRA.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Del.3.2-Promotion-and-Outputs-for-JRA.pdf</a>
Optimal digitisation pilot study	Report on optimal digitisation methods pilot study	NHM	<a href="http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.3-Review-of-Digitisation-workflows-and-equipment.pdf">http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.3-Review-of-Digitisation-workflows-and-equipment.pdf</a>
Crowdsourcing reports	Report on findings of the crowdsourcing pilot study	NHM	<a href="http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-1.pdf">http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-1.pdf</a> <a href="http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-2.pdf">http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-2.pdf</a>
	Phase 1: Phase 1: Crowdsourcing and Citizen science in biodiversity research	VU	
	Phase 2: Making molehills out of mountains: crowdsourcing digital access to natural history collections	NHM	
Sustainability workshop	Workshop on sustainability of NH Collections	NMP	<a href="http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-2.pdf">http://www.synthesys.info/wp-content/uploads/2014/01/NA3-Del.-3.4-Crowdsourcing-report-Phase-2.pdf</a>
European roadmap	Roadmap for European natural history collections	NMP	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/D3.6-EU-Roadmap.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/D3.6-EU-Roadmap.pdf</a>
SPNHC symposium	International symposium to promote SYNTHESYS activities	MfN	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/S3_NA3_Deliverable_3.7_SPNHC_Report.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/S3_NA3_Deliverable_3.7_SPNHC_Report.pdf</a>
JRA training workshop	Training workshop of JRA outputs	RBGE	<a href="http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-3.8-Report-on-the-SPNHC-Workshop.pdf">http://www.synthesys.info/wp-content/uploads/2018/04/Deliverable-3.8-Report-on-the-SPNHC-Workshop.pdf</a>
Instruction videos	Instruction videos for new techniques, collections and software support	RBGE	<a href="https://vimeo.com/195812974">https://vimeo.com/195812974</a>  <a href="https://www.youtube.com/watch?v=xDvzJYmdG38">https://www.youtube.com/watch?v=xDvzJYmdG38</a>  <a href="https://www.youtube.com/watch?v=lqMwTToOoik">https://www.youtube.com/watch?v=lqMwTToOoik</a>
	Tea-bags for collection organic material in silica gel for DNA extraction	RBGE	
	Stable URIs for specimen images	RBGE	
	Digitisation software Inselect	NHM	

# **The potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and exploitation of results**

## **Impact of the Joint Research Activity**

A main goal within the JRA of developing techniques to digitise natural history (NH) objects is to give researchers a tool to undertake their research without the need of the physical object. The improvements to NH virtual collections as a result of improved digitisation methods are already changing and reducing the need for physical access, improving accessibility to collections to a wide range of scientists. NH institutions are now increasingly able to provide Users with digital data of high quality. This will help reduce costs and time for travelling to collections, damage due to excessive handling or the risks of the sending of physical specimens. Most importantly, the access to the digitised collections is greatly increased, allowing for the democratisation of research, due to the fact that all researchers have an equal opportunity to do high quality research with the digital objects. Researchers have wider, simplified, and more efficient access to the best NH research infrastructures they require to conduct their research, irrespective of location.

## **Impact of Innovation & Development in Imaging**

### **Inselect image segmentation software**

Inselect has led to collaboration and stronger links with institutes outside Europe both through the development of the software and then through its release and subsequent use across at least six countries. The collaboration between NHM London and the Stellenbosch University in South Africa has the potential to produce additional tools and software in the future.

The software was originally designed for segmenting images of specimen drawers in order to automatically separate each tiny insect from the hundreds of other insects in a single drawer and it is proving to be successful in this task. Interestingly, however, there is an ever expanding range of other kinds of tasks for which it is being tested and it is proving to be more versatile than expected. It is therefore having a broader impact on a wide range of digitisation workflows, including trays of microscope slides, malaise trap samples and palaeontological specimens. This impact will include a radical rethink of how some collections can be digitised as a large scale process.

The software and new workflows is also having an impact by reducing the cost of digitisation for some of the collections.

### **Zoosphere 3D imaging and website**

The startling 3D images which can be viewed on Zoosphere enable users to zoom in to examine images of insects in incredible detail in an interactive way. The images are often beautiful or striking and have therefore appealed to a broad audience. This appeal has resulted in significant media coverage in mainstream press, including the New York Times. The coverage has opened up these collections to new audiences and engaged the public with important scientific collections in a novel way, providing an innovative research service to answer the needs of a broader scientific community of users.

The Zoosphere technique enables researchers to work with specimens in a new way, opening up the collections and facilitating research, whilst minimising risk to the physical specimens.

A broader impact is also being seen, such as the Department of Agriculture and Food in South Perth Australia stating interest in adapting the ZooSphere technology for pest control services. This is an example of how the software can be used as a solution to ecological and economical issues, and as such has the potential to be taken up by industry.

There is also the potential for it to be used more widely by new audiences. This includes the opportunity to showcase representative species for a certain insect group of which the sequences can be annotated and used in educational programs. Excitingly, for several artists involved in 3D sculpting, gaming or animation industry these 3D sequences can be used to create fantasy models, art or 3D printed sculptures.

To date more than 110 specimens are digitised using the ZooSphere and placed online, including both pinned and amber preserved insect specimens. Some of these image sequences have been requested by scientists from other Institutions for their research. Google Analytics logged 10,847 users and 17,731 page views from 20.09.2015 to 20.09.2016 with 25% returning visitors.

### **Affordable 2D+ & 3D imaging solutions**

The tests carried out during the assessment of affordable 2D+ and 3D imaging solutions have been incorporated into the 3D Handbook of Digitisation and are openly available for other institutes to use. The results will have a huge impact since most institutes housing NH collections are keen to start digitising their collections in 3D, but don't know where to start. This work will enable these institutes to start 3D digitisation more easily, and at affordable costs.

A wide range of the techniques tested have now been incorporated to the routine workflows of SYNTHESYS3 institutions, opening up these collections in ways that are informative to the researcher and also exciting for the public.

The range of objects that have been digitised using these methods includes entomology, invertebrates, vertebrates, geology, anthropology, prehistory and palaeontology.

### **Micro-CT imaging solutions**

The potential of the micro-CT imaging technology to create accurate virtual representations of both internal and external features of an object at sub-micron resolution, without destroying the specimen, makes this technique an interesting tool for the digitisation of valuable and unique NH collections. It also provides the discipline with the potential to revolutionise its current functioning towards a true omics one, the phenomics and metaphenomics. This will inevitably have a great impact on the International Code for Zoological Nomenclature (<http://iczn.org/>) in the near future.

Micro-CT imaging technology will greatly contribute to the rise of the future "virtual museum" where digital data are shared widely and freely around the world and 3D models created through micro-CT scanning will be either printed or made available via interactive touch screens to be used for public display, education, awareness, dissemination and outreach efforts.

The Micro-CTlab (<https://microct.portal.lifewatchgreece.eu/>) as well as the protocols created during SYNTHESYS3 project refer to a wide audience from scientific community to students, artists, animators and everyone who is interested in natural sciences and digitisation methods. The collaboration that took place between a number of European institutes for the 3D work in SYNTHESYS3 was extremely successful. This 3D network will aim to maintain the 3D Handbook into the future. The additional links with Lifewatch through the Micro-CT virtual laboratory (Micro-CTlab) has also been an important part of this work and, again, this relationship will continue into the future.

### **Impact of Innovation & Development in Data Capture**

JRA work developing software that can detect and capture image properties, recognise and classify depicted specimens, and ultimately generate descriptions of properties like colour, size and shape is not only speeding up the digitisation process, but is making digital surrogates of specimens more useful to researchers and easier to find through online portals.

### **Using Optical Character Recognition (OCR) to capture data**

Closer links have been forged with a major software company ABBYY ([www.abbyy.com/en-gb](http://www.abbyy.com/en-gb)) to discuss how their products are used for NH collections and the needs from our community.



The work undertaken here provided some of the framework for another successfully-funded EU project within EUDAT ([www.eudat.eu/](http://www.eudat.eu/)), [Herbadrop](#), which has incorporated OCR into an aggregated workflow, processing specimen images from several partner institutes. The success of both these projects will help shape the data pipeline within a European research infrastructure.

The impact of OCR on label transcription has been significant. Over 400,000 specimens were processed through OCR at RBGE and over 150,000 specimens have been transcribed through this process at RBGE. It is now used on a routine basis for additional data entry.

## Engaging the crowd to capture data

Engaging new audiences for NH collections has had a major impact on the amount of digitisation that institutes can undertake. It has opened up the potential of carrying out massive digitisation programmes which initially capture minimal data. The power of the crowd is then harnessed for the additional data capture from the specimen labels. This is a hugely important step, being a huge part of bringing curation and research staff to accept the initial minimal data capture which does not give them the data that they need.

The high numbers of the public who are becoming engaged with NH collections shows the attraction of the collections when we are able to open them up.

This work is continuing to grow rapidly, with more institutes becoming involved, both in the Notes from Nature platform as well as the Les Herbonautes software, and the Atlas of Living Australia platform, DigiVol.

The opportunity to work with the public to transcribe the label data has made this part of the digitisation process affordable.

The numbers of specimens which have been transcribed using some of these platforms is impressive. Around 100,000 microscope slides containing chalcids were imaged and their labels transcribed by 1,000+ 'digital volunteers' over a 10 month period. So far 2,071 slides of foraminifera have been transcribed with a third and final batch released in autumn 2017 and currently being transcribed.

Global collaboration between NH institutes through the [WeDigBio](#) annual event has been an exciting outcome from this work. This enables a major media push from many institutes around the world in a co-ordinated event, which is growing each year.

The crowdsourcing activities are continuing to contribute to a public engaged in science directly through individual participation. By integrating this activity Europe-wide, SYNTHESYS3 has fostered public engagement across Europe, moving beyond national boundaries to creating an informed and participatory European citizenry, better able to contribute to and influence environmental initiatives.

## Impact of Innovation & Development in Access & Sharing

### Digitisation on Demand (DoD)

A high-level plan for a European networked DoD service was drafted in SYNTHESYS3. This has the potential to provide an integrated service which would form part of a European research infrastructure.

The barriers to DoD have been identified through a survey and there is now the potential to investigate how these are overcome to enable more institutes to join a networked approach. This would have a major effect on reducing the amount of work and time currently needed by researchers to contact each institute individually to a) determine the extent of the relevant collections and b) to request a loan or digitisation.

The potential impact of DoD for NH collections is significant. There is a fundamental cost to providing such a service that many institutes cannot cover at present, and yet the value of this service for taxonomic research is large. The implementation of a service at RBGE has provided an opportunity to assess the costs and impact of integrating this within core work.

A DoD pilot study at the NRM was successfully carried out. During the first four months of the project, c. 10,000 specimens were sorted and c. 7,000 photographed, resulting in >12,000 digital images. Of those,

1,810 images were included in a crowdsourcing project for label transcription, which was completed after 50 days, involving 125 volunteer transcribers. This has provided valuable data for developing services, particularly in potential language issues involved in crowdsourcing across a multi-lingual Europe.

An alternative approach was explored based on the concept of crowdfunding and the requirements to develop a pilot, including TreeMap visualisations, to support the crowdfunding work. The interactive TreeMap enabled users to understand what has been digitised, what is still to be completed, and what the costs of digitisation are in an effort to attract sponsorship. The use of crowdfunding for this work has the potential to enable many more institutes to incorporate this service more widely.

The use of a crowdfunding model for DoD also has the potential to engage the wider public with the collections. The action of providing financial support for these collections would encourage a closer connection and a stronger feeling of involvement.

## **Access and sharing collections**

The work carried out within the NA2 work package: recommendations for a Management Policy on Digital Collections (MPDC), was closely linked to the work within the JRA on providing open access to captured data. The MPDC has now been adopted by the CETAF Digitisation Working Group and will provide a strong foundation for developing a collaborative and consistent approach to digital collections management within Europe.

The recommendations have already been used within RBGE during the development of policy documentation.

The policy statement on digital data produced by guests or other external users of NH collections that was written alongside the MPDC is being adopted by the CETAF Collections Working Group and will have an influence on the development of future policies, in terms of legal framework, general principles for digital collection data and key principles for digital collection data produced by users.

These policy recommendations and statements have encouraged stronger and more effective collaboration between collections managers across Europe. They have required discussion, consensus and agreement and this has helped strengthen this community.

The harmonisation across domains of standards for sharing data will have a significant impact on opening access to collections and enabling users to work effectively with the data. This is an area that needs constant review and discussion to ensure that standards are in place and are consistently used.

The application of standards will enable a wider range of users, particularly those who are not as familiar with the data, such as graphic artists, software/application developers, education programmes, etc.

A key portal for access to occurrence data, Global Biodiversity Information Facility (GBIF), currently contains a total of 136,311,486 aggregated records of preserved specimens including fossils from institutes around the world. A comparison of different data portals was carried out for one institute (RBGE) which found some remarkable figures to demonstrate the level of use of NH collections data.

The impact seen when collection data and images are made available through a single portal is exceptional. Researchers are more than ever having to work with large datasets to answer some of the big questions, such as the impact of climate change on biodiversity and the environment. This requires institutes to ensure that their data are accessible and available through these aggregated portals. The progress made so far has been shown to be impressive and the work carried out within the JRA has contributed to the rapid expansion of the data available.

## **Impact of the Networking Activities**

The SYNTHESYS3 Networking Activities (NAs) are integrated packages that support and develop the NH infrastructure in Europe. They have produced collections management protocols, best practice guides and policies relating to management of new virtual and physical NH collections that are openly available. This integration has led to a better management of the continuous flow of data collected and produced by NH

institutions and enabled them to maximise the use of their existing resources, coordinate efforts and avoid duplication of effort. They ensure usability of collections that is enhanced by increased adoption of common standards and protocols. Links between SYNTHESYS and international organisations is fostering the use and deployment of global standards, for example in management of molecular collections via the collaboration with the Global Genome Biodiversity Network (GGBN). As a result the NAs have helped develop a more efficient and attractive European Research Area for researchers in the natural sciences.

## **Policies for virtual collections management**

Sharing common standards and harmonising policies are essential basics in a successful culture of co-operation. The common principles on data produced by external visitors are another important step in the series of common principles developed and shared by European NH collections. Users add value to the collections; the data they produce can be a real asset for the institutions. On the other hand, each visit will cause extra efforts and costs. The principles provide a common approach and will balance value and costs involved. They will improve transparency of processes and standards within the community but also for Users to the collections.

NA2 work collated already-existing European virtual (digital) collection policies and used these as a basis for recommendations on common European digital data management policies and procedures. The results revealed that many NH collections-holding institutions do not yet have management policies on digital data. In the era of digital transformation, this becomes more and more important. Since policies are determined by national and institutional laws and regulations, there is no 'one-fits-all' solution. The Wiki handbook produced by SYNTHESYS3 will be a helpful and effective tool for the NH collections community to understand what is really necessary and why, and to develop a policy that meet the requirements of the institution. Through the CETAF DWG who will continue the implementation of the handbook produced in SYNTHESYS3, more and more institutions will be encouraged to work on and complete their own set of policies in this area. This will lead to more transparency and reliability within the community and also for external Users.

Both the handbook and the policy statement will build an important basis for joint projects, for example on collections digitisation, marketing, and value chains and in particular for the DiSSCo proposal to the ESFRI roadmap (<http://dissco.eu>). Both these SYNTHESYS3 outputs will be critical building blocks in DiSSCo's organisational structure. Apart from CETAF, common standards and protocols are becoming more and more important in global and interdisciplinary cooperation as well. Through presentations and discussions in these settings (for example SPNHC and TDWG), new fields for application are explored. This will also help to further improve and develop especially the handbook to open it up to an even broader community.

## **3D imaging best practice and standards**

Thanks to the work in the SYNTHESYS3 JRA, almost all the existing 3D techniques suitable for NH collection digitisation are tested and protocols are written. Therefore, the 3D imaging Wiki and Best Practice handbook are very strong products. Because most institutions started with massive 2D digitisation, many are only starting to discover what 3D techniques can be used for. In both outputs all the guidelines, the dos and don'ts, tips and tricks for successful 3D digitisation with any kind of budget are listed, which is a timely tool for those institutions starting out on 3D digitisation, as well as those already in progress. Already now before the publication of the printed handbook, large organisations like iDigBio and CSIRO have shown to be very interested in the content of it. This work will be a great starting point for any organisation wanting to digitise their collections in 2D+ and 3D.

## **Impact of strategic priorities in molecular collections**

### **Strategic priorities for barcoding of NH collections**

Unlocking the genetic and genomic information in the world's NH collections has the potential for immense impact. The many millions of specimens housed in these collections represent an extremely valuable resource for genetic approaches to species discovery and identification (DNA barcoding), studies of

ecological interactions, evolutionary biology, impacts of environmental change, conservation biology, and varied practical applications supporting human utilisation and/or control of biodiversity. However, key barriers to this full use of genetic/genomic data from preserved collections include tractability (for example, technical feasibility of success) and awareness of what is possible.

The report, workshops and subsequent papers produced in SYNTHESYS3 NA2 all aim to address these barriers by promoting increased understanding of what is possible and sharing methodologies to improve project successes at recovering DNA sequence data from NH collections.

The outputs are of high-quality, demonstrated by, for example, the fact that the paper published in the *Philosophical Transactions of the Royal Society* in 2016 has already attracted 17 citations in the first few months since publication and 4,738 views of the full article and 2,333 PDF downloads. The workshop focusing on new sequencing platforms featured world-leaders in the fields of genetics, NH collections and ancient DNA researchers. The interest in this workshop reflects the cutting edge nature of the topic. The new sequencing platforms are enabling transformative changes in accessing DNA from NH collections and we expect there to be continued interest in viewing the presentations online and the resulting paper to be very highly cited and influence working practices.

### Priorities for DNA Library Creation of NH Collections

The potential impact of the pilot study will be along similar lines to that of GGBN:

- increased access for HTS library metadata and data,
- increased use of HTS libraries as a genetic preparation of an organism,
- increased collaboration between different branches of life sciences utilising NH collections and environmental collections,
- increased accessibility of environmental DNA (eDNA) sampling,
- reduced destructive sampling (and requests to do so) of NH specimens, and
- increased access to biodiversity data, in particular eDNA.

The work of the pilot study will be collated into a paper once sufficient data is gathered from the public consultation (public consultation running until 2020). As well, it is planned to publish a number of short communications and to hold oral presentations, for example at the GGBN meeting 2018.

### Network of DNA and Tissue Banks and Policies for New Physical CM

GGBN has become a key player in the world of biodiversity repositories and has been present at many international meetings and symposia, including those of ISBER (International Society for Biological and Environmental Repositories, <http://www.isber.org/>) and ESBB (European, Middle Eastern and African Society for Biopreservation and Biobanking, <http://www.esbb.org/>). GGBN is working closely with both SPHNC and CETAF on many issues, not least on the consequences of the ratification of the Nagoya Protocol and the EU ABS regulations (*Regulation (EU) No 511/2014 of the European Parliament and of the Council on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation in the Union*). SYNTHESYS3 members are involved in creating the 4<sup>th</sup> edition of ISBER's Best Practices Document as contributors and editors. The close co-operation of GGBN members, facilitated by SYNTHESYS3 and its members, is ensuring operators of DNA and tissue banks develop synergies and complementary capabilities, leading to improved and harmonised services. The MoU, best practices, code of conduct and material transfer agreements developed by GGBN with SYNTHESYS3 input means there is less duplication of services, leading to an improved use of resources across Europe. Economies of scale and saving of resources are also realised due to common development and the optimisation of operations.

### Partnership with industry

SYNTHESYS3 has been working with industrial partners to provide innovative solutions to key challenges faced by NH institutions serving the scientific user community. Mutually beneficial relationships have been established, combining technologies and scientific challenges into longer-term partnerships, with industrial

partners showcasing what they can offer the NH community, whilst also promoting the use of research infrastructures by industrial researchers. Preliminary discussions have already taken place and it is hoped these exciting areas of innovation can lead to more ambitious partnerships in the future.

## Sustainability workshop and European roadmap

Both the NA3 sustainability workshop and European roadmap have already demonstrated significant impact on development in the European NH community. They have provided documentary evidence and recommendations on new uses of NH collections and represent an important contribution to the preparatory phase of DiSSCo. SYNTHESYS3 has been pivotal in building a socio-cultural, governance and technical consensus around DiSSCo, which ensures long-term sustainability of the integrated services developed during SYNTHESYS3.

## Impact of Transnational Access

The impact of the **TA** can be measured by the number of additional Users accessing the infrastructures via SYNTHESYS3 and by the resultant publications.

During the four years of the SYNTHESYS3 project, 1,110 awards from 2,805 applications have been spread relatively evenly across the four funding calls, demonstrating sustained high demand, and in total 11,030 User days were delivered. Demand has been geographically far-reaching, with applications from institutions received from 39 EU and Associated Countries.

SYNTHESYS3 offered priority Access to first time Users in order to ensure that the benefits of funding were as widely distributed as possible. Aligned with the restriction of 10% on Users who have had previous funding, the NHMMT and TAF Leaders also ensured that awareness of the funding schemes was continually promoted to new applicants. Wide-scale marketing methods used earlier on in the project meant that the easy-to-reach research communities provided straightforward access to potential new applicants.

Entering details of User outputs on the SYNTHESYS website was a condition of the Users' awards, and also future funding, which provided an effective system for gathering data. In total, 1,428 unique output records are now registered on the SYNTHESYS3 database, though many of these are still 'in preparation' (764) or 'submitted' (69) [to a publisher/journal]. Of the 595 'accepted', 'in press', or 'published' User outputs, 460 are in peer-reviewed journals, with a further 18 contributing to books/monographs. Further analysis shows that Call 1 User visits account for 57% of the outputs, whereas Call 4 only 2%, with Calls 2 and 3 in between. This is to be expected given the inevitable time lag between project initiation and publication and suggests that the final total of outputs from SYNTHESYS3 will be far higher as publications continue to be produced over the next few years.

As well as quantity, the quality of publications from SYNTHESYS3 Users (using Journal Citation Report Impact Factor as a measure) continues to be high. Despite the fact that researchers working on systematic and taxonomic subjects generally publish in subject specific journals in order to reach their target audience, Users continue to demonstrate that their research is of interest to a broad range of journals, many with high impact factors. The range of research topics covered by SYNTHESYS Users remains broad: recent examples funded from SYNTHESYS3 include projects to understand the distribution and transmission of West Nile Virus from mosquito collections (AT-TAF-3844); indicators of plastic and metal pollution by studying the dental and skeletal pathology of marine mammal collections (DK-TAF-5825); archaeological studies on the use of flax products for textile and oil production ancient societies (DE-TAF-3837); studies on 16th century Spanish and Portuguese botanical illustrations to identify cultural links between the two countries (ES-TAF-3724); detection of contaminant seeds in Sardinian irrigated crops to decrease the impact of alien plant species (GB-TAF-4276); and the detection and timing of iron oxide-copper-gold type mineralisation in the SW East European craton to assess mining and mineral potential (SE-TAF-7040).

It is worth noting that a good proportion of the outputs are written jointly with the Host and/or other staff from the hosting institutions. Furthermore, many publications are the result of visits to multiple TAFs. This demonstrates that User visits are creating collaborations that will continue beyond the life of the project.



## Dissemination activities and exploitation of results

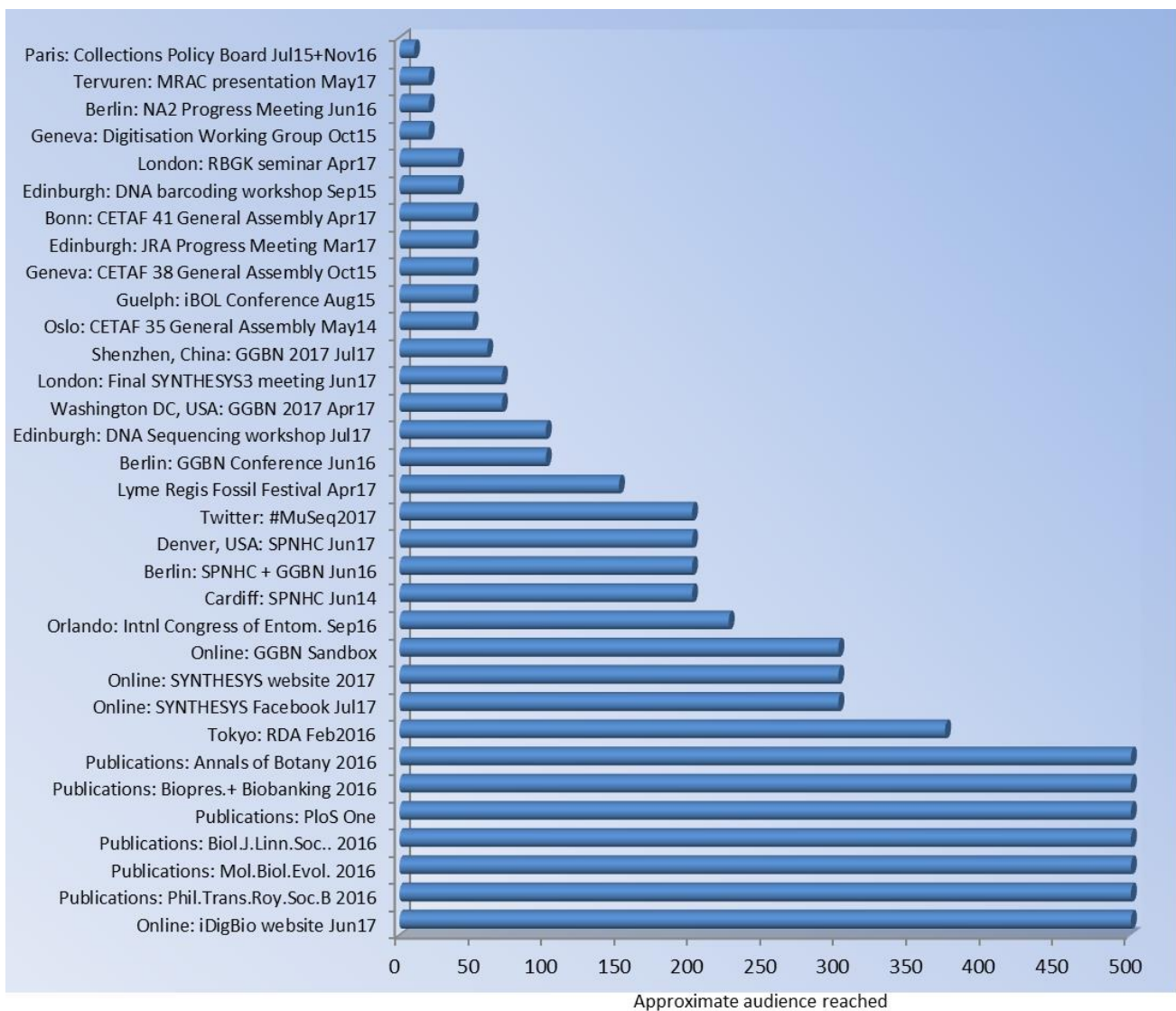
Promotional activities of the consortium members have been carried out on various levels, including presentations on scientific conferences, published articles, workshops, videos, short on-line contributions and presentations to the public. Dissemination activities were recorded on the project management site and collated: <http://www.synthesys.info/network-activities/synthesys3-na3/>. Besides official dissemination activities, the project was also promoted on numerous informal occasions during scientific conferences and other events.

The SYNTHESYS Facebook profile: <https://www.facebook.com/synthesysproject/> was launched in December 2014 followed by the new project Twitter: <https://twitter.com/SynthesysEU/>. Both profiles have been regularly updated with information on interesting research projects undertaken at TAFs together with NAs and JRA outputs. The project Facebook site also frequently informed about major events and provided general information on project institutions.



A selection of SYNTHESYS3 dissemination activities are shown graphically overleaf. Main dissemination routes have been presentations and discussions at conferences and workshops of established networks both within Europe (CETAF, DWG) and internationally (SPNHC, GGBN). Also at ad hoc workshops (DNA barcoding) and public SYNTHESYS project meetings. Publications in journals (e.g. Philosophical Transactions of the Royal Society of London, Series B; PLoS One) and online (SYNTHESYS Facebook, SYNTHESYS website, iDigBio website) have reached even larger audiences and continue to do so.





*Selection of SYNTHESESYS3 JRA and NA2 dissemination activities.*

## Access dissemination and results exploitation

All TA Users have been actively supported by their hosts and TAF Leaders to produce scientific papers and other outputs (e.g. conference posters, publicly-available databases, and training materials) after their visits have concluded. A critical part of the application evaluation by the User Selection Panel was to consider the quality and the impact of the research being performed along with the plans for result publication. The previous SYNTHESESYS2 project (no. 226506) developed an online tool for collecting all outputs details; Users are reminded annually to update the information online. Subsequently the updated listing of all publications is posted onto the public-facing side of the SYNTHESESYS3 website. All Users have been encouraged to give seminar presentations of their work whilst at the TAF of their visit. Financial contributions by the European Commission have been acknowledged in scientific papers (and other outputs).

Two sets of promotional leaflets have been printed and distributed to partners. One flyer promoted access to TAFs and its highlights, the other promoted project outputs.

**SYNTHESYS**  
Synthesis of systematic resources

SYNTHESYS is an EU-funded initiative, comprising of 20 natural history museums, universities, botanic gardens, and research organisations across Europe. It aims to create a shared, high quality approach to the management, preservation, and access to leading European natural history collections through the implementation of three core strands of activity:

- **Access**  
A core element to SYNTHESYS is to provide funded researcher visits (Access) to the 390 million specimens housed by SYNTHESYS institutions – in order to carry out high-quality research projects in the natural sciences that address global challenges.
- **Joint Research Activities (JRA)**  
We are improving the quality of and access to digital collections and data within natural history institutions by developing virtual collections.
- **Network Activities (NA)**  
We are enhancing the quality and quantity of online collections information available to virtual users by implementing best practice benchmarks in collections care. This is raising standards and improving accessibility to collections for all users.

SYNTHESYS3 builds on the success of two EU-funded predecessor projects:

- SYNTHESYS1 (February 2004 - July 2009)
- SYNTHESYS2 (September 2009 - August 2013)

To learn more, please visit our website:  
[www.synthesys.info](http://www.synthesys.info)

synthesys@nhm.ac.uk  
<http://www.twitter.com/1K0D2K6>  
@SYNTHESYS3EU

Partners:

- The Natural History Museum, London
- Botanic Garden and Botanical Museum, Ghent
- Herbario Central de México, Mexico City
- Hungarian Natural History Museum, Budapest
- Museo Nacional de Ciencias Naturales, Madrid
- Museum für Naturkunde, Berlin
- Museum Historialo Nationalo, Paris
- National Museum, Prague
- National History Museum, Vienna
- National Herbarium, Leiden
- Naturhistoriska riksmuseet, Stockholm
- Polish Indigo Institute of Natural Sciences, Kraków
- Royal Botanic Garden, Edinburgh
- Royal Botanic Gardens, Kew
- Royal Museum of Central Africa, Tervuren
- Staatliche Gärten für Naturforschung
- Stuttgartia
- State Museum of Natural History, Budapest
- University of Copenhagen
- VU University Amsterdam

Creating an integrated European infrastructure for researchers in the natural sciences

Our three core strands of work are:

- Access
- Joint Research Activities
- Network Activities

See how we tie these strands together →

Financed by the European Commission through the Seventh Framework Programme

## Dissemination plans and outlook

Within CETAF, JRA outputs will be taken on by the CETAF Digitisation Working Group and the CETAF Collections Group to ensure the continuation of work and further dissemination of results, tools and recommendations. Work on writing up of results in the form of papers continues, including bringing the results from SYNTHESYS3 into other projects which will also lead to longer-term impact and additional papers. Many of the software outputs will continue to be used and developed and the number of overall users is expected to continue to rise. Online outputs such as the 3D Handbook are considered to be living documents and will be updated as more solutions are tested and developed.

## Glossary:

### Abbreviations and Institutions

BGBM	Botanischer Garten und Botanisches Museum Berlin	NGS	Next-Generation Sequencing
CSIC	Consejo Superior de Investigaciones Científicas	NH	Natural History
HCMR	Hellenic Centre for Marine Research	NHM	The Natural History Museum London
HNHM	Magyar Természettudományi Múzeum	NHMW	Naturhistorisches Museum Wien
HTS	High Throughput Sequencing	NMP	Národní Muzeum
JRA	Joint Research Activity	NRM	Naturhistoriska Riksmuseet
MfN	Museum für Naturkunde, Berlin	NRSG	Networking & Research Steering Group
MNHN	Muséum National d'Histoire Naturelle	RBGE	Royal Botanic Garden Edinburgh
MPDC	Management Policy on Digital Collections	RBGK	Royal Botanic Gardens, Kew
MRAC	Musée Royal de l'Afrique Centrale	RBINS	Royal Belgian Institute of Natural Sciences
NA	Network Activity	SGN	Senckenberg Gesellschaft für Naturforschung
NA2	Network Activity 2	SMNS	Staatliches Museum für Naturkunde Stuttgart
NA3	Network Activity 3	UCPH	Københavns Universitet
NBC	Naturalis Biodiversity Center		

### Initiatives and Societies

**CETAF:** Consortium of European Taxonomic Facilities (<http://cetaf.org>)

**DISSCO:** Distributed System of Scientific Collections (<http://dissco.eu/>)

**CPB:** Collections Policy Board (a CETAF body) (<http://cetaf.org/about-us/cetaf-bodies>)

**DWG:** Digitisation Working Group (a CETAF body) (<http://cetaf.org/about-us/cetaf-bodies>)

**ESFRI:** European Strategy Forum on Research Infrastructures

([https://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri))

**EUColComp:** EU Collections Competencies Project (<http://eucolcomp.myspecies.info/>)

**GGBN:** Global Genome Biodiversity Network ([www.ggbn.org](http://www.ggbn.org))

**iBOL:** International Barcode of Life (<http://www.ibol.org/>)

**iDigBio:** Integrated Digitised Biocollections ([www.idigbio.org](http://www.idigbio.org))

**ISBER:** International Society for Biological and Environmental Repositories ([www.isber.org](http://www.isber.org))

**SPNHC:** The Society for the Preservation of Natural History Collections ([www.spnhc.org](http://www.spnhc.org))

**TDWG:** Biodiversity Information Standards (<http://www.tdwg.org/>)

### The address of the project public website

# SYNTHESYS

Synthesis of systematic resources

[www.synthesys.info](http://www.synthesys.info)

[synthesys@nhm.ac.uk](mailto:synthesys@nhm.ac.uk)