SYNTHESYS Final Publishable Summary Report

Grant Agreement number: 226506

Project acronym: SYNTHESYS

Project title: Synthesis of systematic resources

1st September 2009 – 31st August 2013

Contents

Executive Summary 2

Summary description of the project context and the main objectives 3

Description of the main S & T results 6

NA2: Improving collections management, enhancing accessibility and conserving the unique value of European natural history collections 6

NA3: Consolidating the Information Network of European Natural History Collection 16

JRA1: PrediCtoR software tool 19

JRA2: DNA libraries applicable to skeletal museum specimens and soil-embedded archaeological remains 21

JRA3: Using Microsampling techniques to minimise damage to specimens 22

JRA4: Plants/fungi optimised DNA Extraction Techniques 23

JRA5: Development of high-throughput methods for DNA isolation from invertebrates with muco-polysaccharide rich tissue 28

Description of the potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and the exploitation of results 31

Executive Summary

SYNTHESYS has created a shared, high-quality approach to the management, preservation, and access to leading European natural history collections. These collections are a physical infrastructure supporting a variety of research including the impact of human activity on the diversity and distribution of biodiversity, and consequently the provision of ecosystem services essential to human well-being.

Present and future access to European natural science collections relies on those collections and their accompanying data being well preserved and easily and widely accessible both now and in the long-term. It has been the aim of SYNTHESYS Networking Activities (NA), and Joint Research Activities (JRA) to raise standards of collections care and access, to remove barriers to access, and to maximise the efficient use of limited resources available.

SYNTHESYS Access has provided access to the Consortium's vast collections, facilities and libraries with support from in-house researchers and curators. It provided Users with a total of 10,464 days of access over four years, with 1,551 recorded outputs from Users of which 722 are 'accepted', 'in press' or 'published'.

The SYNTHESYS JRA have assisted in meeting researcher demands for sequenceable DNA by enhancing and improving extraction tools and protocols. The JRA (i) developed non-invasive tools for estimating the presence of ancient DNA in specimens, (ii) investigated creation of DNA libraries to reduce the need to re-sample rare museum specimens by effectively immortalising one DNA sample, (iii) investigated the use of micro-sampling techniques to minimise impact on collections, (iv) tackled problems associated with extraction of DNA from plant material, and (v) opened up more effective access to DNA in formerly difficult organisms such as molluscs. By focusing the JRA on DNA extraction, SYNTHESYS has exploited a largely untapped facet of the Consortium's 337 million strong collections.

NA2 assessed the state of collections management in Europe, and provided resources and training to help institutions raise standards and minimise risks to long term preservation and access. It reviewed management tools such as performance indicators to help institutions use limited resources efficiently and measure performance in a meaningful way. NA2 improved collections management through the following mechanisms: (i) an assessment tool and audit facility to enable institutions to complete a self assessment on the status of their collections management; (ii) training courses in priority areas of collections management; (iii) a web-based forum for exchange of information and advice on best practice, techniques and policies; and (iv) sets of performance indicators in collections management.

NA3 has improved electronic access and sharing of collections information across Europe, previously held in disparate institutional databases. Barriers remained to efficiently and accurately transfer data from specimen labels etc. into electronic form and for specialists to feedback data such as new determinations of specimens to the original providers. NA3 has addressed these barriers through: (i) investigating innovative ways to populate databases; (ii) provision of state-of-the-art tools for specialist Users; (iii) improving the quality of content supplied by the global scientific community through implementation of best practice and the setting of global data standards; (iv) providing mechanisms for remote annotation of specimen records; (v) provision of a Helpdesk for researchers entering data via data portals.

Many hundreds of thousands of specimens held by SYNTHESYS Participants collections, previously considered too scientifically valuable to sample, provide the means to address questions about changes in biodiversity, impacts of climate change and the origins of human culture within Europe. The impacts of the SYNTHESYS NAs and JRA have ultimately been to vastly increase the range of specimens available to researchers across the ERA, and to improve efficiency of collections use for DNA extraction, thereby maintaining the balance of using collections to keep them relevant but also preserving them for future generations.

Summary description of the project context and the main objectives

SYNTHESYS is designed to produce an accessible, integrated European resource for research Users in the natural sciences. The SYNTHESYS Consortium has created a shared, high quality approach to the management, preservation, and access to leading European natural history collections. This has been achieved by providing physical access to collections of museums and herbaria (including increased accessibility to DNA within the specimens) and also the wealth of electronically stored data associated with those collections.

The aim was to significantly improve the capacities and operational accessibility of natural history collections for European researchers. While much has been achieved in earlier EU Framework Programmes, collections across Europe remain, in part, fragmented. SYNTHESYS has taken integration to a new level and will have a lasting effect on the capabilities of the infrastructures and broaden the wide spectrum of research that can be carried out using them. The main beneficiaries of this ambitious integration process are Users in the European bio and geoscience research community, particularly researchers active in the field of biological, molecular and geological diversity.

Recent advances in molecular and information technology research are already being implemented in natural history research and collections management, but there still is scope for co-ordinated “quantum leaps” in both these areas. SYNTHESYS has provided a sustainable framework for integration via the Joint Research Activities (JRA) and the Networking Activities (NA) in these exciting areas. The JRA have delivered new tools to enable Users to more efficiently extract DNA from valuable archive material. The NAs have provided enhanced quality and quantity of online collections information to virtual Users and have implemented best practice benchmarks in collections care to raise standards and improve accessibility to many of the Beneficiaries’ collections for all physical Users.

A core element in SYNTHESYS has been to deliver physical access to 337,204,000 specimens housed by SYNTHESYS Beneficiaries. In particular, 4,058,500 type specimens (the one original specimen to which all others are compared for the purpose of correct naming). The Transnational Access programme (TA) involving 16 Beneficiaries with collections has contributed to delivering this aim.

Transnational Access (TA)

SYNTHESYS aimed to deliver over 10,000 User days of TA via a fully integrated online application and evaluation system. This has been coordinated through four annual calls for proposals. All Users with awards have concluded their visit within 12 months and their research outputs have been collated and disseminated by the SYNTHESYS Coordinator. The Beneficiaries have been organised into 10 Transnational Access Facilities (TAFs).

Networking Activities (NAs)

Well-preserved, well-documented collections constitute a *sine qua non* for biodiversity-related research. Establishing and promoting best practice in collection development and maintenance across Europe will enhance the quality of the collections and their usability for an ever-growing spectrum of Users. The NAs have ensured usability of collections is enhanced by increased adoption of common standards and protocols. In addition, the NAs have made collections information available electronically to Users by addressing a number of bottle necks to populating the electronic infrastructure.

Joint Research Activities (JRA): Establish efficient sampling strategies for ancient DNA analysis of museum specimens

The development of high throughput DNA sequencing has led to a sharp increase in User demand for samples of rare and unique materials. Curators have faced a serious problem in balancing these burgeoning User demands for destructive analysis in pursuit of ground-breaking science against the long-term stewardship and preservation of collections of global importance. The overall objective of the JRA was to meet the contemporary Users' demand for sequencable DNA (via improved extraction tools and methodologies) from accurately identified archive material.

Resultant User benefits have included non-invasive tools for estimating presence of ancient DNA and optimal extraction protocols for ancient and more modern DNA which has been a major step towards the production of high quality DNA sequences which will be made publicly available in databases enabling non-taxonomists to identify material potentially including vectors of disease, agricultural pests, and invasive species. Integration of DNA will facilitate systematics studies on morphologically challenging groups.

Summary of SYNTHESYS objectives:

NA2 Objectives:

- Provide Performance Indicators (PIs) in collections management
- Quality assessment in collections management
- Advanced training in collections management

NA3 Objectives

- To significantly improve the degree of mobilisation of specimen information and to create focussed interfaces to meet the increasing demands from both researchers and the public
- To assist institutions in implementing the developed "best practice" data standards for electronic data throughout the virtual research communities and thus achieve the highest degree of interoperability possible for the European Research Area
- To facilitate the technical access to specimen data for specialised User groups in research and society, that are not adequately covered by other initiatives such as GBIF

JRA1 Objectives

JRA1's main aim was to develop software to help collections managers and Users determine the likelihood of specimens containing recoverable DNA.

- Develop a curatorial 'shield', in the form of a Web-hosted software tool - *PrediCtoR* – which predicts the probability of successful ancient DNA recovery from archive collections in museums (predominantly focussing on fossilised and non-fossilised bone)
- Make the tool scalable and integrate with additional databases to increase the sophistication and robustness of prediction
- Investigate dominant factors influencing optimal DNA recovery from archive bone specimens

JRA2 Objectives

The goal of JRA2 was to develop, test, and optimize protocols used in combination with NGS technologies to be applicable for degraded DNA. In particular, protocols for the creation

of DNA libraries out of both skeletal museum specimen and soil embedded archaeological remains and in solution capture enrichment were investigated.

- Optimise Illumina library generation protocol for very small amounts of highly degraded aDNA that allows for further amplification and therefore multiple uses in different experiments.
- Reducing the amount of contaminating DNA introduced by fungi or microorganisms before library creation in order to increase the proportion of endogenous DNA in a library, improve sequence data quality and create favourable preconditions for multiple amplifications and enrichment.
- Creating a protocol that enables the enrichment of specific loci in a generated library

JRA3 Objectives

The central direction of the work of JRA3 has been to develop a system for the use of very small samples (microsamples) in ancient DNA extraction from museum bone specimens.

- Generate baseline data on the use of minimally-destructive bone sampling techniques for application in genetic studies of museum archive-held material
- Improve our understanding of the location and stabilisation of DNA in bone, to develop optimal methods for DNA recovery, pre-screening and decay modelling
- Provide Users with an improved protocol for the recovery of pre-amplified DNA from extremely small sample volumes

JRA4 Objectives

In JRA4 the focus was on exploring and optimising protocols for extracting DNA from archival plant (and fungi) specimens, ranging in age from ‘recent herbarium specimens’ to over 140 years old.

- Identify the main obstacles to successful DNA extraction/amplification of herbarium specimens based on SYNTHESYS User and other Users experiences
- Test protocols for DNA extraction (including scaling-up of DNA extraction and whole-genome amplification from herbarium specimens)

JRA5 Objectives

The objective of JRA5 was the development and testing of a new, safer and more effective DNA isolation procedure from mucopolysaccharide-rich tissue, particularly from museum specimens.

- Development and test safer and more effective DNA isolation procedure from mucopolysaccharide-rich tissue, focussing on museum specimens of varying age (5-150 years)
- The setup of a Web database as an efficient tool to disseminate protocols and experience of DNA extraction from non-entomological invertebrates reports among Users

Access Objectives

To provide access to the collections, expertise, facilities and services at 16 institutions split across 10 TAFs.

Description of the main S & T results

NA2: Improving collections management, enhancing accessibility and conserving the unique value of European natural history collections

Identification and prioritisation of areas for improvement in collections management required baseline data for to benchmarking purposes. Networking Activity 2 (NA2) assessed the state of collections management in Europe, provided assessment parameters and training tools to enable institutions raise standards and minimize risks to long term preservation, and to enhance access to Users. NA2 also reviewed management tools including performance indicators to help institutions use their limited resources efficiently and quantify performance.

NA2 has improved collections management through delivery of the following products:

- Set of performance indicators in collections management (Deliverable 2.2)
- An assessment tool and audit facility to enable institutions to complete a self - assessment on the status of collections management practices (Deliverable 2.5)
- Initial-stage development of a European Competency Framework for staff specialising in collections management (Deliverable 2.10)
- Training courses in priority areas of collections management such as emergency planning, access to collections, molecular collections, ethnobiology (Deliverables 2.11-18)
- A web-based forum for exchange of information and advice on best practice, techniques and policies (Deliverable 2.9)

Performance Indicators (PIs) in collections management

Institutions use performance indicators and other metrics to measure, for example, how many scientific visitors (Users) are accessing the collections. They are potentially an aid to efficient and effective management of collections. At the outset of SYNTHESYS, PIs across European institutions were highly variable in form and usefulness. In association with the museum community beyond the natural history sector, NA2 developed a set of 68 standardised PIs which institutions can apply to meet their local needs and also facilitating comparison with peer institutions across Europe. Examples of these can be seen in Figure 1 and are available to the wider community via the SYNTHESYS website and EU-CoM: http://eu-com.cybertaxonomy.africamuseum.be/. The SYNTHESYS PIs were assigned to four categories, relating to:

- long term trends in collections
- distinct collections management activities; and
- the level to which collections management activities are performed
- the efficiency with which activities are performed

	Long term trends	Distinct activities
Activity level	1	2
Efficiency level	3	4

This structured approach helps institutions apply the PIs appropriately in their existing management processes.

Performance Indicator	What is it an indicator of?	Category	Comments
1. Collections			
1.1 Number of *specimens* in collection.	Collection size.	1	Estimated, based on SYNTHESYS collections survey. Updated each year.
1.2 Number of type *specimens* in collection.	Type richness.	1	Estimated, based on SYNTHESYS collections survey. Updated each year.
1.3 Number of species represented in collections.	Scope of collection.	1	For each taxonomic group. Updated each year.
2. Access and Use of Collections			
2.1 Number of *requests* for *specimens* on loan to other institutions for research in one year.	Interest in the collections by scientific community for visual examination at their home institution.	2	
2.2 Average time to process approved loan *requests* (working days) over one year.	Outgoing loan turnaround time.	4	Mean for all approved loan requests of the difference between the date loan request was received and date of loan despatch (in working days), over one year.
2.3 Number of *requests* for digital images in one year.	Interest in digital images of the collections.	2	
2.4 Number of *requests* for *specimens* on loan for exhibition in one year.	Interest in the collections by museum exhibition community.	2	Includes both internal and external requests, recorded separately.
2.5 Number of *requests* to visit the *collection* in one year.	Interest in the collections by scientific community for visual examination at the host institution.	2	
2.6 Number of destructive sampling *requests* for DNA extraction/other purposes in one year.	Interest in the collections by scientific community for testing beyond visual examination.	2	From both internal and external researchers. Sampling for DNA extraction and for other purposes counted separately.
2.7 Number of other requests.	Interest in the collections for other purposes.	2	Includes requests to visit for non-scientific purposes, collections enquiries etc. but not identification requests.
2.8 Number of research loans sent in one year.	Comparison with requests shows loan refusal rate (for any reason).	2	Number of approved loan transactions.
2.9 Number of *specimens* (or *curation units)* sent on loan to other institutions for research in one year.	Collections management activity (outgoing loans).	2	
2.10 Number of outgoing research loans returned (or partially returned) in one year.	Loan return rate, to compare with loan request rate.	2	
2.11 Number of specimens (or curation units) loaned for	Collections management activity (return of	2	

Figure 1 Examples of the set of Performance Indicators

For the full document, please see: http://www.synthesys.info/network-activities/synthesys2-na2/

Quality assessments in collections management

The first SYNTHESYS project (SYNTHESYS1, Grant 506117, 2005-2009) identified a need for improved consistency in collections management standards across Europe and specifically a need to improve the skillset and generate a self-sustaining, collaborative approach to staff development and succession planning for collections-related staff. SYNTHESYS2 has addressed these needs through training courses and the development of EU-CoM (the Collections Managers Forum; see next section). SYNTHESYS2 devised a set of European competencies to improve performance and promote staff training, self-directed learning plans and mobility within European institutions.

To assist in prioritising training and development needs in European natural history collections, SYNTHESYS1 developed an assessment tool to measure both museum and herbarium performance in collections management compared to a set of benchmarks. This tool required teams of assessors to visit each collection for several days. It was effective but was also very labour-intensive and hence costly. In SYNTHESYS2 a simpler, web-based self-assessment tool (based on the SYNTHESYS1 tool) has now been developed in which institutions can tick boxes on-line to indicate whether they fail to meet, meet (or exceed!) 63 benchmark statements in 5 categories

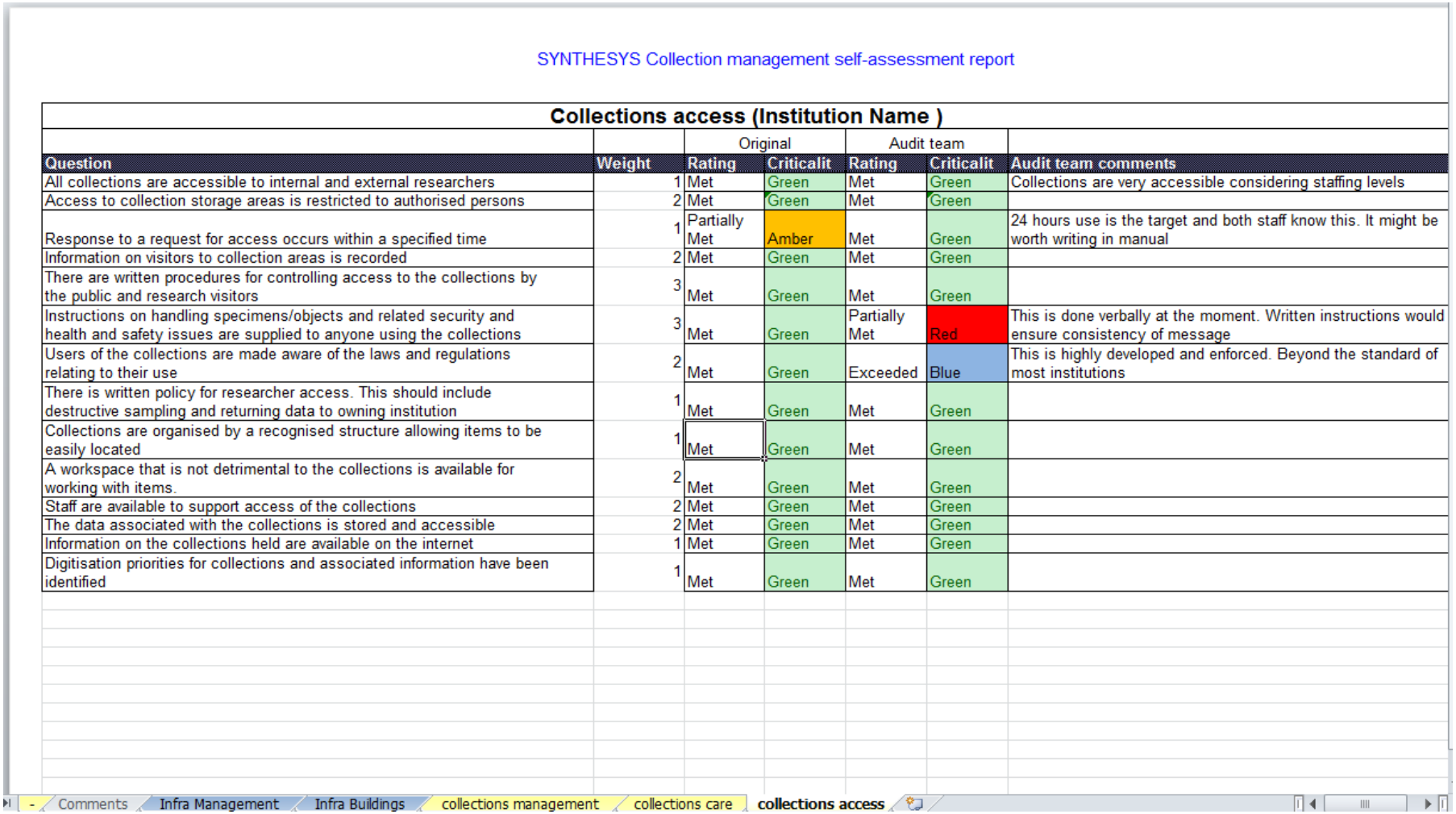
1. Infrastructure – management
2. Infrastructure- buildings
3. Collections Management
4. Collections Care and
5. Collections Access

The first two categories apply to whole institutions whereas the last three can be completed by individual departments giving a greater level of granularity than was previously feasible in the SYNTHESYS1 surveys. This makes the tool much more flexible.

A risk weighting highlights areas requiring immediate attention to reduce risks to collections. The method is intended for use beyond the SYNTHESYS2 Consortium and CETAF institutions and has been promoted across the Member States via national networks and relevant international conferences (including SPNHC). To ensure consistency and to help improve the tool an audit procedure was developed and 14 randomly selected institutions were visited by a small team of SYNTHESYS2 consortium collections experts (who had previously been participants in on-site collections assessments in SYNTHESYS1) to validate the self-assessment results during a 2-day site visit.

By the time the SYNTHESYS2 project concluded seventeen institutions had completed the self-assessment exercise and received reports (see Figure 2 for example) that include recommendations to help address sub-standard areas including using the resources available via the EU-CoM Forum described below.

The consensus from the participating institutions demonstrates these reports have been invaluable; several have used the results and particularly the additional advice provided by the audit assessors to support far-reaching improvements within their institutions. Fourteen institutions received audit visits and found the additional advice provided by the assessors of high value for the future development of their internal policies and processes.



SYNTHESYS Collection management self-assessment report

Collections access (Institution Name)

		Original		Audit team		
Question	**Weight**	**Rating**	**Criticalit**	**Rating**	**Criticalit**	**Audit team comments**
All collections are accessible to internal and external researchers	1	Met	Green	Met	Green	Collections are very accessible considering staffing levels
Access to collection storage areas is restricted to authorised persons	2	Met	Green	Met	Green	
Response to a request for access occurs within a specified time	1	Partially Met	Amber	Met	Green	24 hours use is the target and both staff know this. It might be worth writing in manual
Information on visitors to collection areas is recorded	2	Met	Green	Met	Green	
There are written procedures for controlling access to the collections by the public and research visitors	3	Met	Green	Met	Green	
Instructions on handling specimens/objects and related security and health and safety issues are supplied to anyone using the collections	3	Met	Green	Partially Met	Red	This is done verbally at the moment. Written instructions would ensure consistency of message
Users of the collections are made aware of the laws and regulations relating to their use	2	Met	Green	Exceeded	Blue	This is highly developed and enforced. Beyond the standard of most institutions
There is written policy for researcher access. This should include destructive sampling and returning data to owning institution	1	Met	Green	Met	Green	
Collections are organised by a recognised structure allowing items to be easily located	1	Met	Green	Met	Green	
A workspace that is not detrimental to the collections is available for working with items.	2	Met	Green	Met	Green	
Staff are available to support access of the collections	2	Met	Green	Met	Green	
The data associated with the collections is stored and accessible	2	Met	Green	Met	Green	
Information on the collections held are available on the internet	1	Met	Green	Met	Green	
Digitisation priorities for collections and associated information have been identified	1	Met	Green	Met	Green	

- Comments | Infra Management | Infra Buildings | collections management | collections care | **collections access**

Figure 2: Sample page from the self-assessment output report.

In order to identify trends or particular areas of strength and weakness in the institutions assessed, a matrix of results was compiled and analysed. An example of a matrix is shown in Figure 3. Even at this low level of granularity patterns can be seen with benchmarks being recorded as partially met or not for the majority of institutions demonstrating clear areas of priority for improvement across Europe. The detailed results are recorded in a report on the SYNTHESYS website (http://www.synthesys.info/network-activities/synthesys2-na2/). NB: The identity of individual institutions has been removed to preserve anonymity.

The results indicated a number of areas of strength in most institutions and these are discussed in more detail in the formal report available on the website (http://www.synthesys.info/network-activities/synthesys2-na2/).

It is clear that most institutions had, at the time of assessment, at least a basic collections management policy and importantly the staff have been trained in, and are aware of, current collections management and collections care procedures and best practice. Institutions have resources including a designated budget allocated specifically to collections management and care and also a clear staff structure that recognises collections management as a core task. However, whilst succession planning is acknowledged as important in most institutions, transfer currently is mainly by word-of-mouth reflecting a general lack of documented procedures.

An example of a benchmark met or exceeded in most institutions is “Staff have been trained in, and are aware of, current Collections Management procedures and best practice”. This is encouraging as SYNTHESYS1 identified training as a weakness in institutions surveyed in 2009. Hence there has been a qualitative improvement.

Conversely, circa 30% of organisations do not have an adequate written emergency preparedness plan. Thus there is a lack of established process for dealing with disasters and the salvage of collections.

A second area for widespread improvement relates to health and safety. Although generic training is strong in most institutions as it is embedded in national Law, specific threats to collections such as mercury and/or arsenic contamination are not always acknowledged and staff are currently inadequately trained and processes are not documented as part of best practice.

The detailed results are recorded in a report on the SYNTHESYS website (http://www.synthesys.info/network-activities/synthesys2-na2/) and a paper summarising this work will be submitted for publication in 2014. Lessons learned from this exercise will inform the improvement and modification of the self-assessment in SYNTHESYS3 (Grant 312253, 2013-2017).

The results highlight areas where there is scope for improvement to the self-assessment system and the benchmarks used. For example, discussion with the leaders of the SYNTHESYS2 JRA at a mini-symposium held in Leiden in August 2013 identified a number of additional benchmarks relating to molecular collections for inclusion in the self-assessment when it is developed further in SYNTHESYS3. The SYNTHESYS2 JRA developed protocols to minimise unnecessary sampling of specimens for DNA and improve extraction from “difficult” specimens such as plants and molluscs (See pages 19 to 30). Institutions need to be made aware of and be encouraged to use these improved protocols to preserve and guarantee the quality and integrity of objects such as unique fossil remains.

By adding such protocols to the list of benchmarks and to the self-assessment tool, institutions will be made aware of and strongly encouraged to adopt these as standard procedures for managing the controlled use of specimens for molecular studies. Table 1 shows examples of the new benchmarks developed.

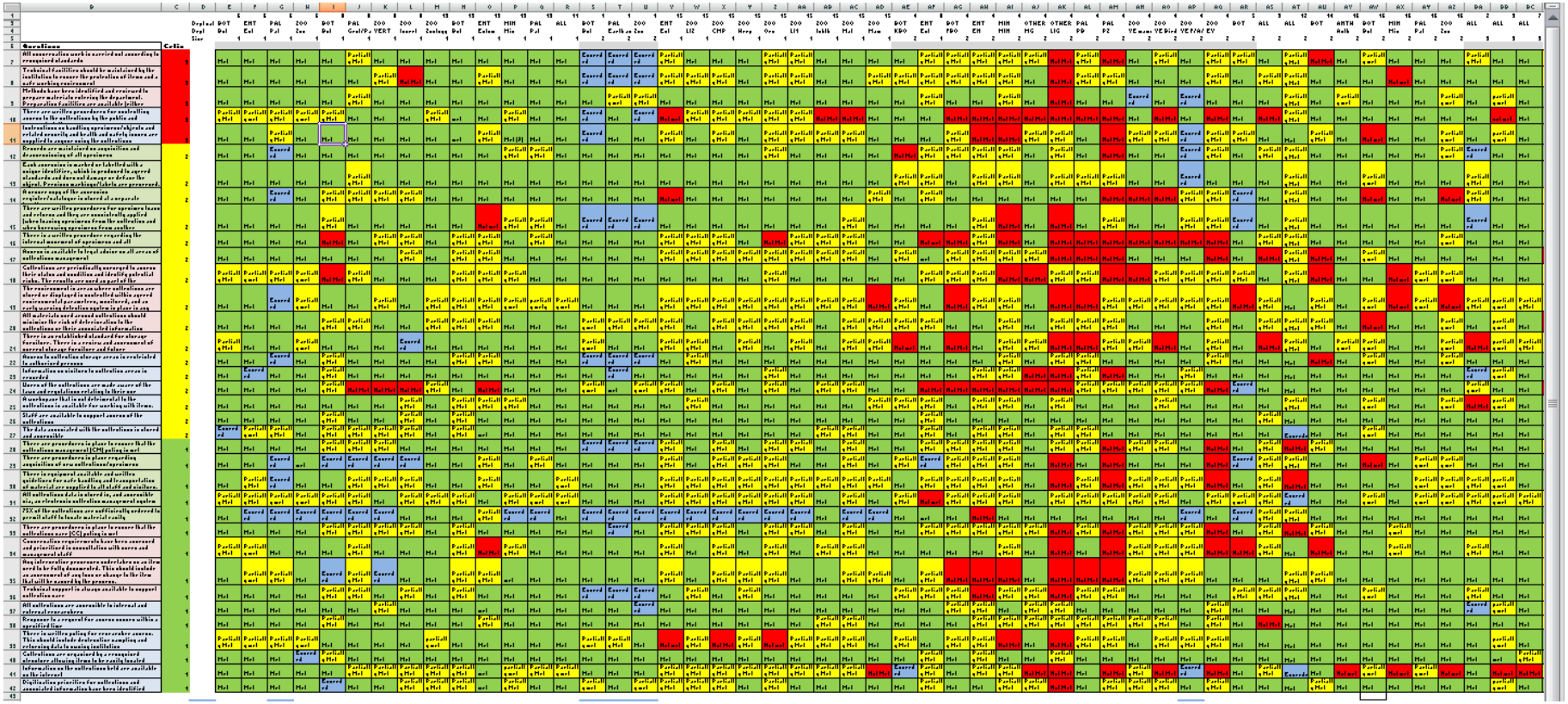


Figure 3. Matrix of institutions by department (for illustration only). Blue = “exceeded”, Green = “Met”, yellow “partially met” and Red = “Not met”

Table 1 Proposed additional benchmarks to add to the SYNTHESYS standards list

Collections staff are aware of protocols developed by the SYNTHESYS project to maximise extraction of DNA from and encourage users to use them: JRA 5 methods for Muco-polysaccharide rich tissues (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra5-development-of-high-throughput-methods-for-dna-isolation-from-invertebrates-with-muco-polysaccharide-rich-tissue-home-page/), JRA 4 protocols for extraction form plant and fungal specimens (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra4-plantsfungi-optimised-dna-extraction-techniques/)
Collections staff are aware of protocols developed by the SYNTHESYS project and other research to minimize destructive sampling for DNA and other molecules and encourage users to use them.
All users wishing to extract DNA and/or proteins from specimens are expected to use the PrediCtoR software tool developed by SYNTHESYS to determine likelihood of successful extraction of DNA (http://thermal-age.eu)
The online PrediCtoR is the standard non-destructive decision making protocol used to avoid unnecessary destructive sampling of palaeontological material for molecular studies (http://thermal-age.eu)
Sampling for DNA from bone specimens is carried out using techniques that minimise damage to the specimens such as the SYNTHESYS JRA2 DNA library protocol (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra2/) or the SYNTHESYS JRA3 microsampling protocol (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra3-mordor-methods-for-optimal-recovery-of-dna-from-osteological-remains/)
Extraction from herbarium and fungal material is carried out using SYNTHESYS JRA4 protocol (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra4-plantsfungi-optimised-dna-extraction-techniques/)
Extraction from muco-polysaccharide rich specimens such as molluscs is carried out using protocols that avoid use of highly toxic materials and can be scaled up to meet high-throughput demands such as the SYNTHESYS JRA5 protocol (http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra5-development-of-high-throughput-methods-for-dna-isolation-from-invertebrates-with-muco-polysaccharide-rich-tissue-home-page/)

Lessons learned from this work will inform the improvement and modification of the self-assessment in SYNTHESYS3 and as more institutions add to the dataset trends relating, for example, to size of institution will become more evident. SYNTHESYS3 and the CETAF Collections Policy Board (http://cetaf.biodiv.naturkundemuseum-berlin.de/activities.php) will seek means to improve these priority areas

EU-CoM - Raising collections management standards and encouraging best-practice in European Institutions

After the SPNHC (Society for the Preservation of Natural History Collections) conference in Leiden 2009 and a follow-up questionnaire, a strong need was revealed for an overarching network or forum for European natural history collection managers to facilitate the sharing of best practice and trouble-shooting.

During a brainstorming meeting at the Natural History Museum in London in August 2010, it was decided that this new forum be called EU-CoM, and that it would be developed with the support of the EU projects EDIT (European Distributed Institute of Taxonomy) and SYNTHESYS (Synthesis of Systematic Resources). A draft plan and provisional website were created and development started towards the end of 2010.

In the electronic discussions following the initial meeting, it was decided that the forum would be created as a sub-domain to the Cybertaxonomy portal of the Royal Museum for Central Africa, all the while remaining independent.

The idea was to have a working Forum with a substantial amount of initial collections management content before being launched to serve as the primary resource for Natural History Collections in Europe and is intended to be driven by the community.

This work was carried out in association with the CETAF European Collections Policy Board (Formerly EDIT Directors of Collections Committee).

The EU-CoM forum is now live (http://eu-com.cybertaxonomy.africamuseum.be) and allows institutions to access and share a variety of data and support on collections management through a single portal. It has been populated with a variety of content of use to the collections management community including new techniques, examples of policies, disaster plans etc. It is one site where collections managers can go to for information on best practice and innovative ideas. Members take responsibility for areas of content. The forum creates a conduit between those in need of advice and guidance on the one hand, and the body of literature institutional and individual expertise in Europe and beyond on the other.

For the final report on the EU-CoM website, see http://www.synthesys.info/network-activities/synthesys2-na2/. As EU-CoM is a site to which documents and procedures will be uploaded by the natural history community, it will become a living resource which will play an important role during SYNTHESYS3 where it will be further developed to ensure that it matches the needs of both the established physical collections and latterly the “new” i.e. digital and molecular collections.

Figure 4 An example of the forum pages on EU-CoM.

Competency Frameworks

SYNTHESYS1 identified a need for improved consistency in collections management standards across Europe and specifically a need for a consistent approach to the learning and development planning and implementation for collections-related staff.

SYNTHESYS2 aimed to devise a set of European competencies to improve performance and promote staff training and mobility within European institutions.

Knowledge and experience gained during the progress of the SYNTHESYS2 led to the conclusion that delivery of a pan- European Competency Framework would require resource additional to that being provided via NA2, such as the development of a corresponding training curriculum linked to ECVET (European Credit System for Vocational Education and Training), and expertise on the sociological elements to implementation of the Framework.

NA2 sponsored a workshop to define what was needed to achieve a comprehensive Competency Framework and to develop a workplan for delivery.

A component of the workplan was for the NHM, London to prepare (in association with the Collection Policy Board under CETAF) a bid to the EU-funded Leonardo de Vinci Transfer of Innovation Programme to promulgate the NHM Collections Competency Framework to other European institutions. The funding model of Leonardo da Vinci is more highly suited to vocational training modules.

The bid made to Leonardo da Vinci Transfer of Innovation was successful with €285,000 awarded for a two year project which started in October 2013 which will deliver a multi-language set of competencies for collections management staff, a curriculum of VET opportunities, examine the feasibility of gathering ECVET accreditation for the competencies, trial the Competency Framework in a number of institutions, and identify the feasibility of extending the products beyond the natural history community.

The report of the NA2 workshop is available on the SYNTHESYS and Eu-CoM web-sites. (http://www.synthesys.info/network-activities/synthesys2-na2/).

The ground work established by SYNTHESYS 1 and 2 has been significant in building a robust case for funding. The Leonardo da Vinci project will deliver a universal multi-language set of competencies that can be arrayed to suit the needs of institutions of varying size, culture and governance. It describes the requirement of each role in collections management and also the individuals' personal competence in those areas. Its ultimate aim is to raise the standards of collections management across Europe and hence ensure scientific and other access to these collections for the foreseeable future. Further details of this project can be found on the ADAM Leonardo Projects and Products Database

www.adam-Europe.eu/adam/homepageView.htm

Training

NA2 training courses assisted institutions to share best practice in collections management and raise awareness. Priority training needs were identified using evidence gained from SYNTHESYS1 and through a questionnaire sent to over 1000 European researchers and collection managers. Five courses have been delivered on Emergency Planning and Response, Molecular Collections and Collections Access with some 50 people attending from more than 15 countries. In addition a "Train the Trainer" course gave representatives from a number of institutions a suite of skills and techniques for cascading broad ideas and concepts in collections management to staff within their own institutions in local languages to help enhance dissemination.

In the framework of SYNTHESYS, 7 training modules were organized in 5 different countries (http://synthesys.cybertaxonomy.africamuseum.be/). These modules were attended by 110

participants from 20 countries, involving the expertise of 43 trainers from 9 countries. A detailed report on the training modules is available on the SYNTHESYS website (http://www.synthesys.info/network-activities/synthesys2-na2/).

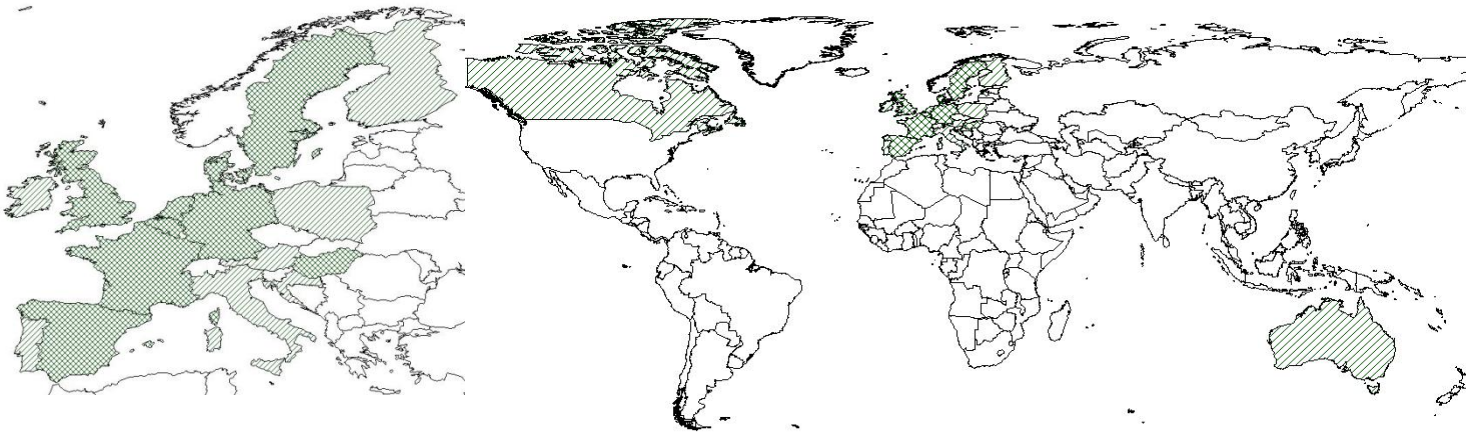


Figure 5: Country of origin of participants (right-angled hatch) and trainers (left-angled hatch) for the SYNTHESYS Advanced Training in Collections Management.

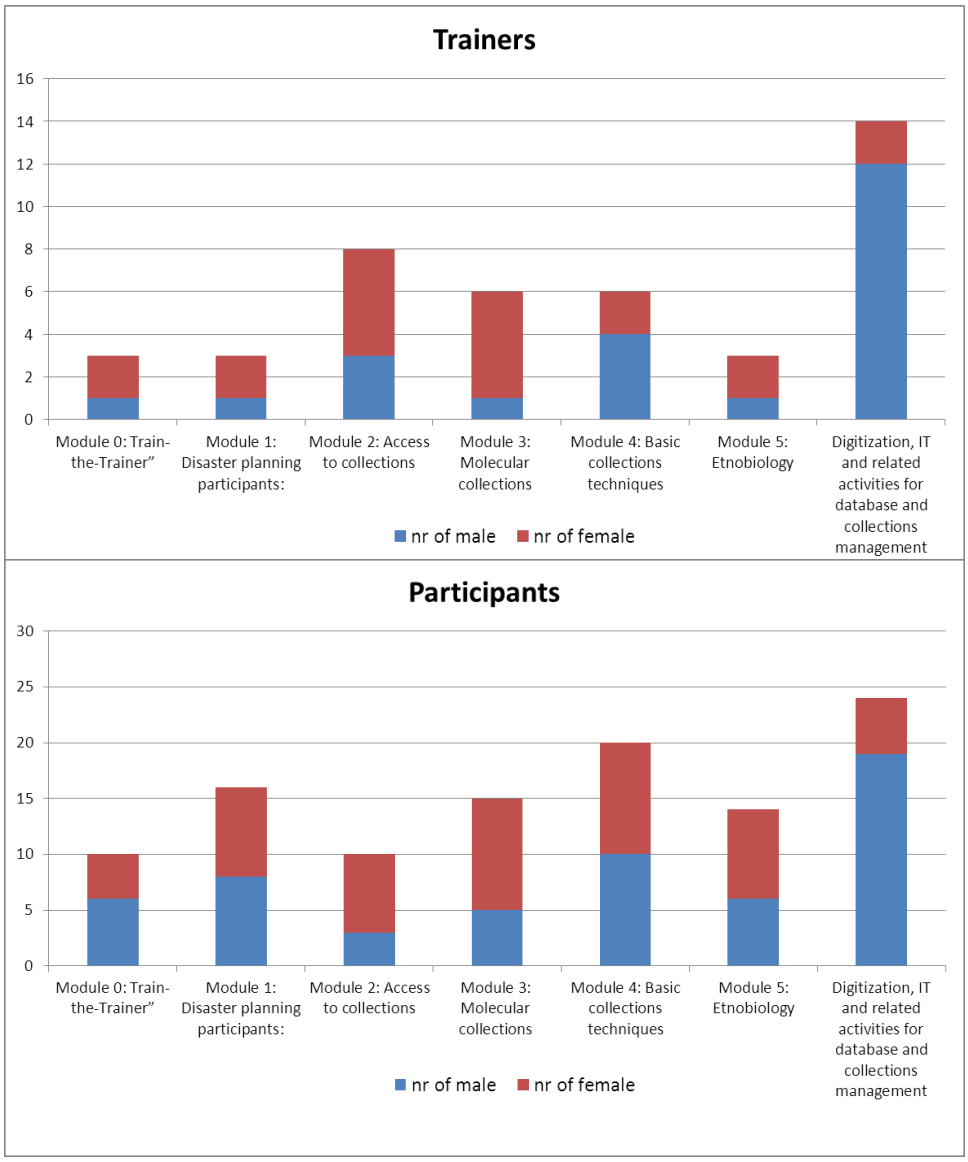


Figure 6 Numbers attending and gender balance of trainers and participants for each module.

Feedback was gathered from course participants and graphical representation is included in the detailed report. One example is given in Figure 7 which shows that in general participants considered that the courses provided a comprehensive overview of the topics (some organisational difficulties in the “Access to Collections” course may have led to the lower scores in this course, but overall the scores are positive).

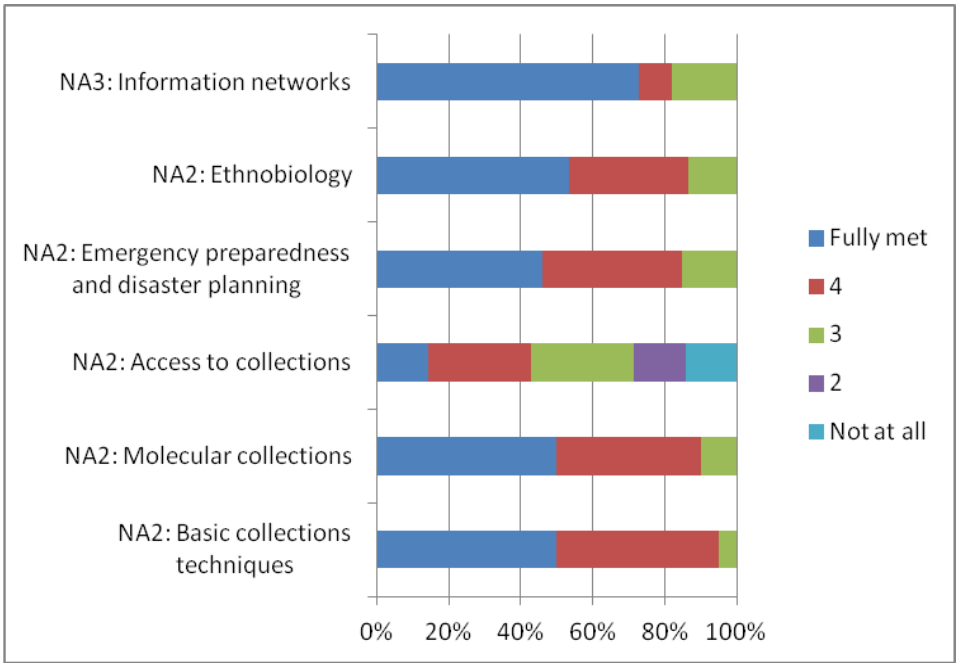


Figure 7 Participant feedback - “Did the course provide a comprehensive overview of the topic”

NA2 delivered:

- A set of Performance Indicators for collections management
- A web based self-assessment tool
- 17 assessments with recommendation reports
- 14 Audits of assessed institutions
- A work plan for a competency framework which was developed into a successful bid using the core resources of NHM for a Leonardo da Vinci Transfer of Innovation project
- 7 training modules attended by 110 participants from 20 countries, involving the expertise of 43 trainers, from 9 countries.
- The EU-CoM Forum for collections management in Europe

NA3: Consolidating the Information Network of European Natural History Collection

There has been an exponential increase in the amount of data captured from museum and herbarium specimens over the last 15 years. Virtual access to this data has been impeded firstly by their distributed nature, held in individual museums and herbaria, and secondly by the heterogeneous nature of systems used and standards for data entry and online access mechanisms. In 2004 SYNTHESYS1 began to address these barriers by creating universal data standards, protocols, user interfaces and a help desk to assist data providers and users. These technical innovations greatly increased the access to data held in collections

across Europe in a way that allowed institutions to maintain ownership of their data whilst giving access via the BioCASE (Biological Collection Access Service for Europe: http://www.biocase.org) portal. Currently, a relatively small percentage of the specimens held in European museums and herbaria are in electronic form and therefore electronically inaccessible and not available to the user community.

In SYNTHESYS2 the data networking activity NA3 investigated ways to assist institutions to speed up the task of capturing specimen label data and produced a report highlighting ways to automate this process (http://www.synthesys.info/network-activities/synthesys2-na3/). Also tests have been carried out on data capture mechanisms by volunteers (crowdsourcing). This report documents the rationalization possibilities of data capture through for example using mechanical production lines to reduce handling of specimens and speed up the rate of scanning of specimens; optical character recognition (OCR) to read the information on labels automatically, and feature detection - software that can recognise automatically which part of a herbarium sheet is the label so that it can then be scanned and translated with OCR.

Crowdsourcing has become an increasingly popular means of capturing data using the power of thousands of online volunteers who can enter data remotely from their chosen location.

One challenge to creating “virtual collections” has been ensuring that the datasets are up-to-date and accurately reflect the latest information known on the physical specimen such as its correct name, type status etc. For example, if a specialist views an image of a specimen held in the owning institution and re-identifies it then this new identification needs to be fed back to the owning institution where the specimen is kept and a label with the newer identification added to the specimen. NA3 developed a simple prototype “reverse wrapper” mechanism to feedback annotations to the owning institution. Through interviews with collections holders a number of psychological and practical problems were identified and a report published (http://www.synthesys.info/network-activities/synthesys2-na3/). As a prototype the Berlin Botanic Garden has tested a system that reads annotations on the JSTOR platform (http://www.jstor.org/) used by a number of large digitization projects and writes these back to the Berlin Botanic Garden database. For example, if an external expert identifies a plant from an image on JSTOR the identification is fed back to the owning institution who can then update the information on the physical specimen.

The report produced on this work details an improved annotation system and has led to the funding of the AnnoSys project (an annotation data repository for networked and highly complex biodiversity data, https://annosys.bgbm.fu-berlin.de/) by German DFG.

ABCD (Access to Biological Collections Data) schema is a common data specification for biological collection units, including living and preserved specimens along with field observations that did not produce voucher specimens. It was created with support from SYNTHESYS2 funding to aid the exchange and integration of detailed primary collection and observation data and was been further developed during the lifetime of the project.

NA3 has extended ABCD coverage through an extensive helpdesk via email, telephone or in-situ assistance. The BioCase Provider Software has now been extended further to comply with the latest developments in the occurrence networking community (XML and DarwinCore archives). In addition a similar schema and Portal (GeoCase) was developed in SYNTHESYS1 for geological collections such as fossils and in SYNTHESYS2 this was expanded and additional features have been added.

The following figures demonstrate the impact of SYNTHESYS2 investment on the number of collections items now available virtually:

- GBIF / BioCASe network: 28m records in March 2009 → 30.6m records by August 2013

- OpenUp network: mobilised 1.9m multimedia documents from 19 content providers
- GeoCASe: mobilised 460,000 records from palaeontological and mineralogical collections
- DNA-Bank Network: linked 49,571 DNA samples with their original specimens in 10 institutions
- Italian Biodiversity Network: 1.2m records
- Australian Virtual Herbarium: c.4.5m records

NA3 has developed tailored web portals to give access to specimen data for specialists such as taxonomists who require more data than provided by GBIF (Global Biodiversity Information Facility). This NA has tested the portal with real world examples to ensure that the process is easily accessed and intuitive to use.

Many users want access to broader information on collections, e.g. which organisations hold palms from Borneo. The Biodiversity Collections Index began integrating existing metadata sets such as the long-standing, botanical Index Herbariorum and initiated a data gathering exercise to add new data and update existing data. This system has now been merged into the Global Register of Biodiversity Repositories (http://grbio.org/) to create a large set of data on a wide spectrum of collections from museum specimens to collections of frozen tissues and DNA in biobanks.

Many technical improvements have been made to the BioCase provider software to optimize performance and improve compatibility for example with Internet Information Server. It is also important that the SYNTHESYS portal is up-to-date with the GBIF network and important changes have been made to introduce the GBIF DarwinCore Archives as a publication method and to match the GBIF's registry requirements.

BioCase and GeoCase are portals to data sets in individual institutions; they do not provide the software for management of collections and storage of a data. However NA3 has provided assistance in selecting appropriate software for their needs by reviewing a number of packages and producing a report. There are many software packages available for collections management and to help users identify appropriate packages 22 systems were identified and documented on the BD-Tracker (Biodiversity Service & Application Tracker, a Drupal site initiated by EDIT, http://www.e-taxonomy.eu/).

In summary over the four years of the NA3, the online access to specimen information has been improved significantly, both in respect to quantity and richness, contributing over 10.7million records in natural history collections across Europe and beyond. This has been achieved through:

- documenting the existing technical possibilities, standards and best practices
- defining precise requirements for upcoming projects
- training of IT managers and staff
- improvements and extensions of existing implementations
- extensive support of institutions in using the existing standards and software

Joint Research Activities (JRA): Establish efficient sampling strategies for DNA analysis of specimens

Museum and herbarium collections are a potentially a vast store of un-extracted DNA and demands for access to specimens for molecular sampling are increasing exponentially. There are a number of barriers to DNA access: i) intrinsic factors such as effects of storage and preservation treatments, past history and cell products that interfere with the extraction process; ii) extrinsic factors such as the reasonable concern of curators to avoid unnecessary sampling damage to rare and irreplaceable specimens.

The SYNTHESYS JRA (1-5 sub-projects) aimed to address some of these barriers to access. Three activities specifically addressed the issue of minimising damage through avoiding sampling specimens that are unlikely to yield usable DNA, minimising the amount of material removed by using micro-sampling techniques and providing ways to ensure that sampling is not only minimal but only needs to be done once by immortalising DNA samples. A further two activities developed protocols to improve efficiency of extraction from "difficult" material, specifically, dried plants and molluscs.

JRA1: PrediCtoR software tool

JRA1 developed software to help curators and users determine the likelihood of specimens containing recoverable DNA. These predictions can help to quantify the risks associated with destructive analysis of specimens and avoid unnecessary sampling.

The user enters data on a PC, tablet or smartphone such as how long the specimen in question was buried and at what site and altitude. The likelihood of successful recovery of DNA is displayed as a traffic light system - green for samples likely to yield DNA, amber suggests that there may be risks and red suggests that there is insufficient DNA for analysis.

When PrediCtoR was first conceived it was designed to assess the likelihood of successful DNA extraction and amplification using the Polymerase Chain Reaction (PCR) technique commonly used for amplifying small amounts of DNA to provide sufficient material for analysis. PCR has been widely used in biological and medical science for many years; however technologies in this field are developing rapidly and so called Next Generation Sequencing platforms are replacing PCR as the prime method of analysing DNA. The JRA1 team changed the focus of the web-site to match this technological change and the site now reports predicted length of DNA fragments rather than the copy number of a PCR fragment of a specific length. The site was re-purposed as Thermal-age.eu taking these changes into account.

Thermal-age.eu calculates the thermal history of a site using established climate databases and uses this to predict fragment lengths. Temperature is a major factor determining DNA survival and so past climate data can give an estimate of the temperatures that the specimen was exposed to throughout its history. When the software is used the records are kept and so can be published and subjected to scrutiny. An important feature of Thermal-age.eu is that it collects data from users to help refine the quality of predictions. It also provides detailed explanations and supporting materials to help users understand the strengths and limitations of the numbers we can produce.

During the course of the project a number of improvements were made to the software such as improved temperature resolution and improved altitude correction; altitude is an important factor in determining the thermal history of a specimen.

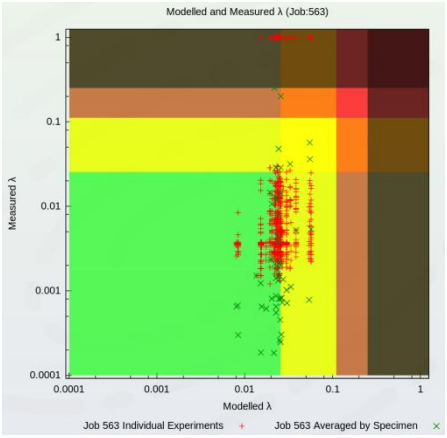
The user is helped to establish the locality of the specimen they are testing by a link from the site to Wikipedia which gives latitude and longitude – this site includes many archaeological sites and geographical features in addition to place names. Altitude can either be entered by hand or automatically filled in.

Soil type is another important factor and this can be added to the calculation – the kind of soil a specimen was buried can affect its thermal history and thus affect DNA survival.

In response to user feedback, the following additions were made:

- software now runs on PCs and tablets and also allows data to be entered on spreadsheets allowing users to enter large numbers of specimens in one "batch".
- an interactive report in PDF form, which can be printed.
- a means of making searches publicly accessible (including an ability to embargo the results to a set date - and change this at will).

An important feature for the future improvement of the model and accuracy of prediction is the ability to enter actual results of DNA sequencing (sequencing provides the precise order of individual nucleotides within DNA molecules) to be uploaded against previously run predictions, allowing data capture and comparison between predicted and measured DNA survival. The figure to the right shows a comparative graph generated for the user showing predicted and actual results.



Improving access to collections

Thermal-age.eu enables collections managers to assess the likelihood of DNA preservation, which was the original aim of the project. However the SYNTHESYS2 JRA1 team realised that a more time effective method of using the tool was to provide a self-service model for the users. Therefore an interactive PDF Report was produced (which enables each of the figures to downloaded as a PDF, PNG, or SVG format).

This has a number of advantages:

1) If the sample is identified as unsuitable the request will never be made.

2) The users get an insight into where and when samples s/he is interested in analysing are predicted to fail.

3) The reporting tool offers a way to report on thermal-age.eu data highlighting instances in which the predicted fragment length is incorrect.

4) A report detailing the predicted suitability of the sample was added to each request for sampling. This 8 page report is generated after each run and provides information on the model set up details and how the final estimate was made.

The tool is currently slightly ahead of its time, and at the point of conclusion of SYNTHESYS2 there have been relatively few studies which have reported DNA fragment length. GoogleScholar lists three articles which have reported having used the tool in publication. A total of 145 unique visitors spend on average of 6 minutes each time they visit the site (a total of 237 visits), this is the time to run a minimum on one analysis.

However, the tool has now come to the attention of leading researchers in the field of ancient DNA and we confidently anticipate future greater uptake of use. Experience of its use to date and feedback from collections researchers indicate that the best way to use thermal-age.eu is for users to supply a thermal-age.eu report to any request to sample museum specimens.

Both collections and research communities are keen for the principle to be extended to other types of tissue other than bone such as dried plants and insects.

Thermal-age.eu is thus heading to become an accepted and common tool in the management preservation and access to museum collections.

JRA2: DNA libraries applicable to skeletal museum specimens and soil-embedded archaeological remains

JRA2 investigated creation of DNA libraries to reduce the need to re-sample rare museum specimens by effectively immortalizing one DNA sample. Ancient skeletal material was used as the model as the DNA is often both highly degraded and in demand from users.

This activity originally aimed to investigate the use of spectroscopic methods to detect levels of DNA in bone specimens without the need for destructive procedures. Early on in the work it became apparent that the great variety of responses depending on the particular animal species being tested and the kind of tissue made this unfeasible. Thus, with the prior approval of the EU Project Officer, JRA2 was repurposed to investigate the use of DNA libraries as a means to reduce the need to re-sample rare museum specimens by effectively immortalizing one DNA sample.

A number of protocols were successfully tested to reduce the detrimental effects on extraction of contaminants such as DNA from soil fungi attached to specimens. The principle is to avoid amplifying contaminant DNA but still amplify the small amounts of DNA in bone specimens to create libraries. Parallel experimental work targeted endogenous (non-contaminant) DNA and enrich it in comparison to contaminants.

JRA2 has developed, tested, and optimized new protocols which used in combination with next generation sequencing technologies can be applicable to the degraded DNA found in museum and herbarium samples. In particular, protocols for the creation of DNA libraries out of both skeletal museum specimens and soil embedded archaeological remains and in-solution capture enrichment (capturing genomic regions of interest from a DNA sample prior to sequencing using magnetic beads) were investigated.

To be of maximum value the protocols were designed to work with a broad range of samples in different states of preservation.

The protocols would also allow particular regions of DNA to be enriched to support studies on the evolutionary history of a particular taxon.

For highly degraded samples for which whole genome sequencing is not feasible, JRA2 developed a new capture strategy for the inference of the evolutionary and/or demographic history of a species using humans and cattle as showcases.

Several existing protocols were modified and optimized and new protocols were developed. Modified protocols include i) a refined extraction method for highly degraded DNA, ii) an efficient and robust library construction method, iii) the comparison of different DNA enrichment strategies for NGS sequencing and iv) a pipeline for data analysis to remove sequencing errors and guarantee the correct assignment of various individual samples run on a single lane.

In collaboration with JRA3, JRA2 refined an existing method to deliver the most effective ancient DNA extraction method available. Dozens of samples can be analysed in a short period of time and also specimens that are considered to contain only minimal amounts of DNA.

Three different library protocols were modified and tested to find the one with the highest amount of incorporated DNA library molecules. The strategy here was to use a protocol that can be successfully applied to a wide range of specimens from differing preservation conditions and not just for the “one big genome” application.

A significant problem with the use of next generation sequencing of highly degraded samples is that there is relatively little useful information contained in any single library. Even after the use of capture array techniques, which pull out a pre-specified set of target sequences, it is still common to have insufficient information for population genetic analyses. We have experimented with generating multiple NGS libraries from a single DNA extract, and then running those through a single capture and sequencing step. This seems to work well, and significantly increased the amount of information generated, without requiring additional sampling from museum or archaeological specimens.

In comparison, we have shown that using capture arrays to enrich for particular target sequences is possible even for highly degraded DNA, through the application of our improved protocols and strategies. A good capture array protocol will provide much better sequence data than an unenriched sequencing run. The newly developed capture enrichment protocol can be used to enrich for either nuclear or mitochondrial DNA, and in conjunction with the new bioinformatic protocols was applied in the recently published study of material from the Mesolithic/Neolithic site at Blatter Cave (Bollongino *et al.* 2013).

Protocols developed in JRA2 will, in combination with the work from JRA 1 and 3, provide a standardized NGS-pipeline for museums and herbaria that can reliable be applied to different sorts of bone specimens to generate population genetic or phylogenetic sequence data with relative small amounts of sample material. These protocols that have been shown to work also on highly degraded archaeological material will routinely work on at least 75% of skeletal museum material if applied correctly, based on the findings of JRA2.

JRA3: Using Microsampling techniques to minimise damage to specimens

JRA3 investigated the use of micro-sampling techniques to minimise intervention in bone and other materials. Changing technologies resulted in a shift of emphasis with this JRA and the research focussed on the minimal amount of bone needed to obtain DNA using both PCR and Next Generation Sequencing methods. It also included the creation of DNA libraries (means of immortalising DNA to reduce the need to sample again); work carried out in association with JRA2. The JRA extended its original remit which focussed on bone to include small amounts of insect remains often encountered in archaeological studies, again comparing PCR and NGS methodologies. The central direction of the work of JRA3 has been to develop a system for the use of very small samples (microsamples) in ancient DNA extraction from museum bone specimens. The plan was to investigate issues arising from the use of micro-samples, including the practicalities of sampling, extraction, and data interpretation. The intention was to open up for sampling the large number of bone specimens that are currently not available, as they are too culturally or aesthetically valuable to permit User access. As in other areas of the JRA, the enormous changes in DNA sequencing technologies (Next Generation Sequencers) brought about a shift in emphasis of the JRA. With these new technologies the JRA3 team were able to sequence much larger amounts of DNA in a single reaction. However this required substantial changes in laboratory methodologies. JRA3 accommodated these technical changes by altering the approach to the core tasks whilst keeping the primary aim; the use of micro-sampling to enable greater access to collections. The JRA3 team principally worked on the development and application of a DNA sampling, extraction and Illumina library protocol, suitable for use with hard tissue (bone or tooth) samples that are under 10mg in mass and as a natural development of this work were able to investigate the potential for sequencing libraries

derived from minimally destructive invertebrate samples, including single legs from a range of beetle remains, and micro-samples from a soft tissue deficient isopod species.

Using these new protocols the JRA3 team were able to make massive reductions in the amount of sample required from museum specimens for DNA sequencing (from c. 500mg to c. 10mg) and in the sample size required for DNA sequencing. This was the case with both traditional PCR beads methods and also the new NGS technologies. This should enable collections managers to make better-informed decisions when providing access to extract DNA from the most valuable samples in skeletal collections.

The work extended beyond bone samples to insect collections and demonstrated that a single, uncrushed beetle leg provided sufficient DNA to enable both multiple traditional PCR-Sanger based sequences, and for NGS.

These results have been communicated to both the molecular evolution and curatorial communities via conference presentations, workshops and peer-reviewed publications and are expected to become standard protocols for sampling through user uptake.

A number of other research projects have evolved from this work including studies on *Bathynomus* (a giant bathyal isopod), *Amara alpina* (Holarctic arctic ground beetle), *Nesophontes* (Caribbean insectivore), and Oryzomyini (rice rats) of the Lesser Antilles. These projects begin to demonstrate the improvement in provision of access to museum materials to support future molecular studies.

JRA4: Plants/fungi optimised DNA Extraction Techniques

Testing commercial available DNA extraction kits on plant herbarium material

Herbarium collections are increasingly used as ‘ancient DNA repositories’ for DNA-based systematic and ecological studies, and represent years of collecting efforts; especially as collecting permits for international scientists for fresh material are becoming increasingly difficult to secure from biodiverse nations.

Extraction of DNA from plant specimens and some families in particular is problematic as plant cell walls and cell products (secondary metabolites) interfere with the extraction process. Previous successful extraction has relied on time-consuming and expensive procedures that also involve toxic chemicals. Tests on fresh and herbarium material carried out in this research project identified commercially available kits that can effectively recover DNA from these “difficult” herbarium specimens.

JRA4 has developed new protocols for users in the systematic community to enable them to carry out the efficient mining of herbarium collections for DNA utilising state-of-the-art DNA technology, thus enhancing the value of specimen collections and increasing contributions to DNA banks.

Despite previous studies conducted on specific aspects of herbarium DNA, no large scale studies have tested the effect of different DNA extraction methods on DNA extraction efficiency from herbarium specimens across a broad range of taxa. And although innovative methods of DNA extraction and PCR amplification are being developed that function well with low quantity DNA, few are currently available for large-scale projects aimed at working across plant taxa using automated or semi-automated protocols. The available methods include (1) silica binding, (2) magnetic bead binding, (3) salting-out precipitation, and (4) anion exchange purification. Each is based on a different DNA extraction technique and chemistry, and hence the methods are expected to vary in their DNA extraction efficiencies.

JRA4 carried out a systematic study of DNA extraction efficiency from historic herbarium specimens representing a broad range of phylogenetic diversity of vascular plant species and preservation histories. The team tested eight commercially available protocols and protocols that are commonly used in laboratories specialised in processing herbarium samples, representing each of the above-mentioned four DNA extraction methods. Tests were also carried out on the effect of preservation method on DNA extraction efficiency through experimentally drying specimens using silica drying, natural air-drying, artificial air-drying, and alcohol drying (both quick and slow). Results are shown in Table 2.

In summary, results from the study together with evidence from previous investigations on herbarium DNA, strongly suggests that there are five important factors to be considered when working with herbarium DNA (in no particular order):

(1) amplification success is higher for shorter target regions due to severe fragmentation of herbarium DNA;

(2) DNA purity is more important as a predictor for amplification success than DNA yield, and hence DNA extraction techniques which maximize DNA purity should be used;

(3) BSA (bovine serum albumin) should be routinely used in PCR reactions in high concentrations when working with herbarium DNA;

(4) specimen preparation method strongly affects PCR success through DNA fragmentation, where specimens treated with alcohol have generally more fragmented DNA; hence shorter target amplicon sizes (c. 100 base pairs) are recommended for alcohol dried specimens in order to maximize success rates; and

(5) there is no biased degradation of nuclear DNA and that organelle (plastid and mitochondrial) and nuclear DNA are equally available in herbarium samples compared to fresh tissue.

This work was published in the online journal PLOS 1 (Särkinen et al 2012) (Deliverables 7.1 and 7.2):

> *How to Open the Treasure Chest?* Optimising DNA Extraction from Herbarium Specimens. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. PLoS One 7 (2012) e43808. doi: 10.1371/journal.pone.0043808. Epub 2012 Aug 28.

Table 2: Performance of different DNA extraction methods with historic plant herbarium specimens; results of the three best performing methods are in bold under each category (published in Särkinen *et al.*, 2012)

Protocol used	Extraction method	Reference/supplier (catalogue #)	Median DNA yield (ng)	Median DNA purity (A260/A280)	No of samples with pure DNA (%)[3]	No of positive amplicons (%)			
						rbcL	*LEAFY*	P6 loop	Combined
DNeasy Plant Mini Kit	Silica binding	Qiagen (69104)	434	**2.118**	7 (14.5)	**10 (21)**	**16 (34)**	**45 (96)**	**45 (96)**
NucleoSpin Plant Kit II	Silica binding	Machery & Nagel (NZ74077050)	593	**1.884**	**19 (39.6)**	**9 (19)**	**19 (40)**	41 (87)	**42 (89)**
Canadian Centre for DNA barcoding Guelph, Ontario	Silica binding	Ivanova et al. [25]	285	1.564	**8 (16.7)**	2 (4)	14 (30)	35 (74)	29 (62)
CTAB[1,2]+ Wizard DNA Clean-up System (Promega)	Silica binding + Salting-out precipitation	Doyle and Doyle [26] + Promega (A7280)	**1712**	**1.620**	**11 (22.9)**	3 (6)	**31 (66)**	**47 (100)**	**47 (100)**
Urea pre-treatment[2]+ DNeasy Plant Mini Kit	Silica binding + Salting-out precipitation	Qiagen (69104)	330	2.116	4 (8.3)	2 (4)	5 (11)	41 (87)	41 (87)
CTAB[1,2]	Salting-out precipitation	Doyle and Doyle [26]	**3004**	1.350	2 (4.2)	0 (0)	4 (9)	38 (81)	37 (79)
ChargeSwitch gDNA Plant Kit	Magnetic bead binding	Invitrogen (CS18000)	**2076**	1.459	**8 (16.7)**	3 (6)	8 (17)	14 (30)	16 (34)
Genomic Tip 20/G	Anion exchange purification	Qiagen (10223 & 19060)	7	0.602	3 (6.3)	**6 (13)**	12 (26)	**42 (89)**	34 (72)

[1]CTAB (cetyltrimethylammonium bromide lysis) with PVP and β-mercaptoethanol.
[2]See File S1 for details of extraction protocol used.
[3]A260/A280 ratios 1.7–2.0 were considered pure.
doi:10.1371/journal.pone.0043808.t002

DNA damage, miscoding lesions, and locus bias in herbarium DNA

The JRA4 project also looked at the preservation and storage of plant specimens to see what effect these had on survival of DNA. By comparing fresh material from old living trees in a botanic garden and herbarium specimens made from the same trees up to 100 years ago, it has been shown that most damage is caused by the preservation methods at time of collection rather than long-term storage in the herbarium. Plant specimens are usually heat dried and/or sprayed with alcohol to prevent decomposition. This supports modern practice of collecting and storing both a small leaf sample "cold" dried with silica for later molecular study and a heat-dried herbarium specimen for morphological examination (Staats et al 2011).

Another question answered by JRA4 is whether there is any difference in the amount of damage to DNA from herbarium specimens between DNA from different cell compartments. Results show that nuclear b, chloroplast, and mitochondrial DNA are equally degraded and this happens at time of preservation.

The first aim of this study was to assess DNA damage as a result of polymerase non-bypassable damage using quantitative real-time PCR for multiple plastid, mitochondrial, and nuclear DNA regions. Secondly, levels of miscoding lesions in herbarium DNA were assessed using 454-sequencing of amplicons derived from each of the three genomic compartments.

Results indicate that there is no difference in the degree of DNA fragmentation between plastid, mitochondrial, and nuclear genomes in herbarium specimens, as the reduction in nDNA copy numbers is not statistically different from those in plastid or mitochondrial DNA. Organelle DNA therefore does not appear to be preferentially degraded in herbarium tissue both directly following drying, or after long-term storage. Copy number reduction directly after drying was more pronounced than that following long-term storage.

The study confirms that herbaria are incredibly rich sources for reliable DNA sequence data. Various next-generation sequencing approaches can now be applied to severely fragmented DNA, and although more difficult it is still possible to generate DNA sequence data from them. The adoption of silica-gel drying of specimens generally yields higher quality DNA than most herbarium specimen preparation methods, and the JRA4 team recommended that all future collections stored for subsequent genetic analysis should continue to use this approach. However, results confirm that herbarium DNA is a readily available resource that will be invaluable for future phylogenetic and genomic studies not least in view of the current biodiversity crisis.

Results (Deliverables 7.3, 7.4 and 7.5) are published in:

> *DNA damage in plant herbarium tissue.* Staats M, Cuenca A, Richardson JE, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT. PLoS One 6 (2011), e28448 doi: 10.1371/journal.pone.0028448. Epub 2011 Dec 5.

These are major advancements in unlocking the vast amount of molecular data held in herbarium specimens.

Plant family-level herbarium DNA extraction database (HDED)

Sharing results in extracting and sequencing DNA from herbarium specimens is essential if methods are to develop and the resources held in herbarium specimens unlocked. In this way the community can learn from successes and failures. As a product of a herbarium DNA symposium held by JRA4 in 2011, an online Herbarium DNA Extraction Database (HDED) was launched and hosted by RBG, Kew (Deliverable 7.6). By using questionnaires, data has been gathered from the community on successful DNA extraction and PCR.

HDED provided an important basis for future sampling and extraction design from herbarium collections, as it includes negative results that normally go unpublished yet are highly valuable in preventing the same hypothesis being tested twice, allowing workers to anticipate adjustments in procedures.

(see http://www.kew.org/science-research-data/directory/projects/DNA-extraction-Herbarium.htm).

In a paper to Taxon (Bakker & al., in prep.), the JRA4 results are basis for prompting the herbarium community to continue to submit herbarium DNA extraction results, both positive and negative, to HDED, thereby eventually creating a self-sustained community-driven database in which contributors remain 'visible' along with their submitted data.

Whole genome amplification from trace amounts of herbarium DNA

Combining research strands looking at extraction and new techniques for DNA using micro amounts of plant material and using the latest sequencing technologies led to a spectacular success - the first successful full Angiosperm nuclear genome sequence from herbarium tissue (Staats et al 2013) (Deliverable 7.7):

> *Genomic Treasure Troves*: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. (2013) PLoS ONE 8(7): e69189. doi:10.1371/journal.pone.0069189

In today's next-generation sequencing (NGS) world, opportunities and prospects for historical DNA have changed dramatically, as most NGS approaches do not rely on large, intact DNA templates but are actually designed for taking short fragmented molecules (100–400 base pairs) as templates. While the application of NGS technologies to ancient DNA from paleontological and archaeological records has been firmly established, its application to historical museum specimens is rare and so far limited to mammals, snails and plants.

This JRA4 project set out to investigate the feasibility of obtaining genomescale sequences using the Illumina HiSeq 2000 platform from a wide-range of historical plant, fungal and insect museum specimens. 'Typical' museum specimens were selected, in order to keep as close as possible to the reality of museum specimens and their preservation histories. Findings showed that complete organellar and nuclear genomes can reliably be generated from low quantities' of historical DNA using NGS.

Thus the study has shown that using a standard multiplex and paired-end Illumina sequencing approach, genome-scale sequence data can be generated reliably from dry-preserved plant, fungal and insect specimens collected up to 115 years ago, and with minimal destructive sampling. Using a reference-based assembly approach, the team were able to produce the entire nuclear genome of a 43-year-old Arabidopsis thaliana (Brassicaceae) herbarium specimen with high and uniform sequence coverage.

Test protocols & formulations in participating herbarium institutes and make recommendations for future specimen preservation

A joint effort has been made between partners in the JRA to assemble a complete plastid genome sequences for 94 specimens of a range of plant families and some fungal groups (see report http://www.synthesys.info/joint-research-activities/synthesys-2-jras/jra4-plantsfungi-optimised-dna-extraction-techniques/, Deliverable 7.8).

The significance of this work is that it will demonstrate to the community that these sequences can be generated at relatively low costs.

The results of JRA4 will aid the herbarium community by arming users with a data-driven justification of DNA extraction procedure and hence building confidence in persuading collections managers to provide access to DNA sample from unique or rare specimens. In addition, results will be of great importance as they reassure users that post-mortem damage artefacts in herbarium DNA sequences obtained (as frequently observed in animal archival DNA) is negligible. Already, the interest shown by the user community for the three JRA4 PloS ONE papers, plus recent invitations for three international symposia, indicates that herbarium collections (both Angiosperms and Fungi) could undergo a hitherto unprecedented exploitation for their molecular resources this greatly enhancing access which is the fundamental aim of the JRA.

Especially the results from Tasks 7 and 8 (testing DNA extraction kits for scaling up, and whole genome amplification from trace amounts of herbarium DNA) are (potentially) spectacular in that complete genome sequences from archival DNA can be obtained for modest costs – around €250 for well-covered 160kbp plastid genomes. It is expected that with the 'demonstrator projects' (Tasks 8a and b: Whole genome amplification from trace amounts of herbarium DNA: plastid genome sequences from a wide range of Angiosperm herbarium: Task 8a; and Whole genome amplification from trace amounts of herbarium DNA: a Fungal test case: Task 8b) a significant larger proportion of herbarium community will be persuaded to embrace the power and economics the latest sequencing technology has to offer. With the dissemination of bioinformatics training of JRA4 partners so far, necessary for proper assembly of all data generated, also comes the notion that data storage and handling is going to be an increasingly important issue for collections managers and users alike.

JRA5: Development of high-throughput methods for DNA isolation from invertebrates with muco-polysaccharide rich tissue

Extraction and amplification of DNA from muco-polysaccharide rich tissues such as molluscs is difficult and requires use of toxic chemicals (CTAB protocol). JRA5 has tested a number of alternative protocols and commercial kits on fresh and museum preserved mollusc specimens and has identified commercially available solutions using magnetic beads that effectively replaces the CTAB protocol and is particularly effective for older specimens.

Also the toxic β-mercaptoethanol used has little effect on yield of DNA and PCR success and can therefore be omitted from extraction protocols for mucopolysaccharide-rich tissues. These are high throughput protocols and lend themselves to automation opening up considerably more effective access to DNA in these formerly difficult organisms.

Four DNA extraction methods representing three techniques available in the market today were tested on a standardized set of four species comprising the most species-rich groups of molluscs (bivalves, prosobranch and pulmonate gastropods):

- CTAB with and without beta-mercaptoethanol followed by Chloroform-isoamyl alcohol (precipitation)
- Qiagen Blood and tissue extraction kit (Silica matrix)
- Nucleospin extraction kit (silica matrix)
- Qiagen Biosprint plant extraction kit and Qiagen Biosprint blood and tissue kit (Magnetic beads – Solid Phase reversible Immobilisation).

DNA amplification from ethanol preserved (older) museum specimens was tested and optimized using in-house designed techniques (internal primers) for the set of all four

standard taxa. In addition, techniques that enrich particular areas of interest in the genome (target enrichment- bait capture and hybridisation techniques) were used on museum collections of the snail Cepaea nemoralis.

A high-throughput DNA extraction protocol based on the Biosprint blood plant extraction kit and tissue kit method was tested in three other major European mollusc collections (London, Madrid and Vienna). This protocol has made the extraction of DNA from molluscs much more efficient taking less time and giving better results.

The biosprint protocol was successfully tested on samples of the four target species as well as additional taxa from the three other major European collections; no major modifications of the protocol will be necessary, indicating that collection-specific influences are negligible.

All four methods of DNA extraction (see above) tested on freshly collected mollusc specimens as well as museum preserved specimens worked well on freshly collected specimens (based on the amplification and sequencing of the 660bp COI barcoding region). However, the full barcoding region could not be amplified from the museum specimens, regardless of extraction method used.

The JRA demonstrated that the chemical βME is not essential in the extraction of DNA from molluscs using the buffer CTAB; it has since been removed from that protocol without need for replacement.

The extraction of mollusc DNA was optimised using a high throughput platform (Qiagen Biosprint using the plant extraction kit).

The failure to amplify the full barcoding fragment from most museum samples prompted a closer look at the post-extraction workflows. Analysis of DNA of the museum specimens, using a fragment analyser (Advanced Analytical), showed the DNA to be highly degraded. Better results were achieved using three pairs of internal primers for the amplification of 200bp fragments, designed for each of the four standard species. Repeatability of results was an issue making it difficult to obtain the full complement of replicate samples.

Standard PCR methods used on specimens are largely trial-and-error resulting in specimens being sampled but not always yielding DNA. Instead the JRA5 team used a protocol published by Maricic et al., (2010) which gives a definite outcome avoiding unnecessary sampling of specimens. In a parallel project, the method was also successfully applied to dried insects specimens, thus demonstrating the possibility of extending this technique to other collections in the museum.

Greater success was achieved using a technique that enriches areas of DNA of interest (target enrichment method). This technique was originally used on bones and not on ethanol-preserved specimens, but the results obtained were promising. Larger fragments of DNA were recovered using the bait capture technique, providing a viable alternative for sequencing alcohol preserved museum specimens.

The outcome of JRA5 substantially increases the value of existing museum collections for groups comprising at least 200,000 species and millions of specimens in European institutions by providing protocols and workflow options for opening them up for DNA research.

The results also moved museum labs towards a safer working environment. The demonstrated uselessness of β-mercaptoethanol removes one of the more harmful reagents in molecular biology laboratories and thus decreases long-term health risks. Also, the replacement of organic extractions eliminates potential associated health hazards (e.g., by removing chloroform from the protocol).

Results have been written up and are in the final stages of preparation for submission: R Danabalan and T von Rintelen ““The need for 2-mercaptoethanol in DNA extraction of molluscs” *Molecular Ecology Resources* (Deliverables 8.2 and 8.3).

Summary of JRA deliverables achieved:

Deliverables 4.1-4.3 (JRA1): Production of the http://thermal-age.eu website, proof of concept using both temperature plus large DNA datasets and DNA validation, and software refined to incorporate thermal modeling, normalized to sample size, development and testing of User-entered database, analysis and reporting.

Deliverables 5.4-5.9 + 6.2 (JRA2): Protocols for library preparation for highly degraded DNA, characteristics of hominin aDNA, immortalization of aDNA libraries, showcase capture-NGS, mitochondrial and nuclear capture enrichment, contamination-free DNA extraction from human bone for subsequent sequencing, and application of mtCapture protocols to various museum and archaeological specimens.

Deliverables 6.1 + 6.3 (JRA3): microsampling handling and validated optimal DNA extraction for microsamples protocols.

Deliverables 7.1-7.8 (JRA4): comparison of DNA extraction kits on herbarium material, DNA extraction protocols for removal of secondary compounds from 'difficult' plant/fungal families, DNA damage and miscoding lesions and locus bias in herbarium material, assessment of post-mortem damage in herbarium DNA, plant family-level herbarium DNA extraction database, whole genome amplification from trace amounts of herbarium DNA, and testing of protocols in herbarium institutes and recommendations for future specimen preservation.

Deliverables 8.1-8.4 (JRA5): presentation of alternative methods of DNA extraction from mucopolysaccharide tissue, replacement component for β-mercaptoethanol, protocol adapted to high-throughput methods, and parallel testing in several institutions of safe high-throughput method for DNA extraction from mucopolysaccharide-rich tissue.

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

The key impact of SYNTHESYS has been to aid in the evolution of a European resource through the creation of an accessible, integrated infrastructure for researchers in the natural sciences in Europe and globally.

The natural history collections held within the museums and herbaria of Europe are world-class in terms of their magnitude and taxonomic coverage. They represent a resource unique in Europe as a model of the diversity of life on earth and are an irreplaceable physical dataset enabling Users to research how human activity (including climate change) is having an increasingly negative impact on the diversity and distribution of biodiversity, which is threatening the continued provision of ecosystem services essential to human well-being. Improved access to these collections, therefore, means that SYNTHESYS has already had, and will continue to have, a far-reaching positive impact on the advancement of environmental science now and in the future.

A range of new services and improved access – both physical and digital – have been provided to a broad range of scientific Users (from biological and geological related disciplines) in a consistent and more easily accessible way. The new tools developed and disseminated have given Users the chance to pursue new avenues for independent research at the leading edge of biodiversity and environmental research.

NA2 has developed Performance Indicators which have been designed in such a way as to allow institutions to not only create their own subset of indicators for internal management but also feed them into an integrated European set. The quality assessments carried out in NA2 have led directly to institutions improving their internal collections management policies, and in some cases establishing policies that were previously absent.

One of the main drivers of NA2 has been to promote best practice in collections standards across Europe so Users receive a similar high level of support at each infrastructure. The NA2 deliverables have given all collections managers (regardless of whether they were SYNTHESYS Participants) the tools to assess and benchmark the quality of their collections management and User access provision. NA2 has developed and promoted the use of collections management standards. The self assessments have identified where these standards are not being met, and the audits of the self-assessments, the EU-CoM forum, and various training courses have offered advice on how institutions can attain the relevant standard.

The survey tools have been made available as a resource not only for European natural history and herbaria collections, but with adaptation, to the needs of other collections. All European collections–based institutions (i.e. art collections via the Museums Association) are able to use the tools as a mechanism to efficiently measure the standard of management of their collections.

NA2 has trained European collections managers to be trainers themselves in collections management techniques. Those staff trained can now run training courses in their home institutions, spreading the knowledge of best practice in collections management to a broader audience. Collections management training will therefore continue beyond the life of the project and outside of the SYNTHESYS Participants. Delegates have been given publications and presentations on what they have learned to pass on to staff in their institution, thus broadening the impact of the training. This has resulted in a well-trained

collections workforce appropriate to meeting User needs in a 21st century scientific infrastructure.

NA2 has generated best practice sets of performance indicators and management polices. These have in turn increased the efficiency of use of resources and enabled improvements in collections' conditions, management and accessibility, particularly in infrastructures where resources are limited.

NA2 dissemination of the self-assessment tool has involved presentations at in-house seminars and international conferences, and promotion via established networks of contacts such as CETAF. In addition, presentations at meetings such as the SciColl programme on collections management, DNA Bank Network and SPNHC (Society for the Preservation of Natural History Collections) have promoted the workpackage and the project as a whole.

NA3 NA3 has promoted the use of data standards by means of focussed training (e.g. for IT managers) and direct assistance through the helpdesks for technology updates and standard implementation.

All ten of the SYNTHESYS Consortium Participant nations are signatories to the UN Convention on Biodiversity. Signatories have an obligation to share data originating from their collections, and in particular, with the countries that the organisms originate from. Hence, SYNTHESYS was required to operate at a European rather than a national level as responsibility is shared by all and the complementary geographic coverage of the holdings makes collaboration a necessity.

This has given SYNTHESYS a powerful mandate to digitise, network and disseminate collections information for the benefit of the European and global biodiversity research User community. NA3 has strived to make this a reality through the provision of tools maximising the amount of data made available via a distributed 'IT network' by increasing the rate of data entry e.g. through use of OCR and by establishing priorities for data capture (aimed for instance at projects such as FPVII-funded PESI). NA3 has delivered a mechanism enabling acceleration of adding/updating specimen data. The global User community now has far greater access to a critical mass of relevant specimen records incorporating the maximum amount of verified information.

The functionality of the 'IT network' has been enhanced through the Helpdesk and jointly implemented standards which have been used to monitor and improve the quality of data that are made available to (and by) Users. This establishment of operational best practice has helped to make a high-quality, large and relevant body of data available for researchers and facilitate evidence-based biodiversity policy development (e.g. the data provide information on temporal changes in species distribution patterns).

The data standards that have been consolidated and harmonised during SYNTHESYS will have a long-term structuring effect as they will continue to be implemented via TDWG to the benefit of Users long after the conclusion of SYNTHESYS.

In order to optimise utility of the digital information for Users NA3 has also created specialist interfaces to maximise the usefulness of collections data to researchers engaged in taxonomic studies. This has greatly increased the quality of data held in GBIF. The data that have been entered into GBIF during SYNTHESYS will continue to be available to all Users well beyond the life of the project.

NA3 dissemination activities include presentations at global conferences such as the TDWG (Biodiversity Information Standards; in New Orleans, USA, in 2011 and Beijing, China, in 2012), various GBIF meetings, International Congress of Systematic and Evolutionary Biology (Berlin, 2011) and the International Conference for Environmental Specimen Banks, Berlin, Germany (2010).

JRA1 has developed the novel, easily-accessible and free PrediCtoR (http://thermal-age.eu) decision-making tool to determine the likelihood of archaeological and fossil specimens containing recoverable DNA and other molecules, therefore enabling collections managers and Users to quantify the risks associated with destructive analysis of specimens. Current DNA extraction policies are often restrictive as institutions have a duty of care to protect and maintain their collections. The impact of this tool therefore is that Thermal-age.eu avoids unnecessary destructive sampling by identifying samples with low likelihood of DNA recovery. The quality of these predictions is constantly improved and refined by collecting the data from Users of the tool. Researchers across Europe will be contributing their data to building the new software. The larger the pool of data contributors, the more robust the model will be. A further consequence of the tool is to improve the relationship between curators and Users and thus to improve confidence in scientific use of collections, improving the degree of access granted to collections for DNA molecular work and thus advancing scientific knowledge. Thermal-age.eu has a further positive impact on those seeking access to collections other than bone, hence liberating further rare specimens for DNA research.

JRA1 dissemination activities have included the creation of a YouTube video to demonstrate how to use the tool, and this is included on the http://thermal-age.eu website. Use of the tool has been promoted at in-house seminars in many of the Consortium institutions (for example at the NHM's Collections Management seminar series) as well as at European and international conferences and meetings, including the Fourth International Symposium on Biomolecular Archaeology (ISBA4 2010), the European, Middle Eastern and African Society for Biopreservation and Biobanking (ESBB 2013), and the Society for the Preservation of Natural History Collections (SPNHC 2013) where live demonstrations were carried out.

JRA2 has investigated the creation of DNA libraries to reduce the need to re-sample rare museum specimens by effectively immortalising a single DNA sample. The specificity and efficiency of library preparation protocols have been improved. The work has developed, tested, and optimised new protocols which, used in combination with next generation sequencing technologies, can be applicable to the degraded DNA found in museum and herbarium samples.

Protocols developed in JRA2 will, in combination with the work from JRA 1 and 3, provide a standardized NGS-pipeline for museums and herbaria that can reliable be applied to different sorts of bone specimens to generate population genetic or phylogenetic sequence data with relative small amounts of sample material. These protocols that have been shown to work also on highly degraded archaeological material will routinely work on at least 75% of skeletal museum material if applied correctly.

The existing publication relating to JRA2 is in the high-impact journal Science (Bollongino *et al.* 2013) and is an example of how the new developed capture enrichment protocol can be flexibly adjusted to work with mitochondrial or nuclear ancient DNA. Further publications are due from this workpackage and the new protocols can be expected to be used more widely in the User community in future years.

The impact of **JRA3** results include the ability to make an order of magnitude reduction (from c. 500mg to c. 10mg) in the sample size required for DNA sequencing, both with traditional PCR-Sanger based approaches, and in library builds for NGS technology. This should enable curatorial staff to make better-informed decisions when providing access to the most valuable samples in skeletal collections.

In the allied area of entomological collections, JRA3 has demonstrated that a single, uncrushed beetle leg provides sufficient DNA to enable both multiple traditional PCR-Sanger based sequences, and for NGS library builds. This ensures improved efficiency of use of museum collections and protection against unnecessary destructive sampling.

JRA3 dissemination activities included presentations to both the molecular evolution and curatorial communities at national and international conferences such as Quaternary Vertebrate Association Extinctions meeting (London 2012), UK Archaeological Sciences Conference (Cardiff 2013), BIOCOLD workshop (Stockholm 2012), International Environmental 'Omics Synthesis (Cardiff 2013), British Association of Biological Anthropology and Osteoarchaeology (York 2013). Results have been published in high-impact journals including Proceedings of the National Academy of Sciences of the US, Molecular Ecology, and Systematics and Biodiversity.

JRA4: The JRA4 study showed that a standard sequencing approach could be reliably used to generate genome-scale sequence data from a wide variety of dry-preserved plant, fungal and insect specimens from historical collections. Results are significant for the following reasons:

i) material otherwise not available, such as rare or extinct species, or costly to obtain is now in reach for comparative genomic analyses without fully destroying the original specimen;

ii) availability of previously inaccessible genetic information from old type specimens that are crucial for resolving taxonomic uncertainties and for providing DNA barcodes for various applications (e.g. ecological studies, conservation, control of agricultural pests and pathogens);

iii) accuracy of nuclear genome based phylogenetics, especially at the Angiosperm species level, is expected to be greatly enhanced as resolution will increase and organelle transmission related artefacts (a problem with commonly-used phylogenetic markers) at the species-level can be avoided, and

iv) including historical samples in demographic reconstructions will significantly increase accuracy of, for instance, estimating past effective population size.

JRA4 will also provide higher performance methodologies and protocols. Widespread collaboration across Europe has been essential for JRA4 as invaluable DNA expertise has been building up in larger European herbaria (including UK, Denmark, Netherlands) in the past decade, and needs to be disseminated in practical ways. Proper testing of resulting protocols by a broad range of herbaria was essential hence requiring further collaboration across Europe. The JRA4 protocols will have a significant global impact on the progress of DNA barcoding as implemented in iBOL activities by the Consortium for the Barcode of Life's Working Groups (JRA4 Leader is chair of CBOL's Science Advisory Board)

JRA4 findings have been published in three separate publications in the Plos ONE journal, as well as being promoted at international conferences such as the Biosystematics meeting (7th International Congress of Systematic and Evolutionary Biology (ICSEB VII) of IOSEB: International Organization for Systematic and Evolutionary Biology).

JRA5 investigated the use of Beta-mercaptoethanol which has been a component in the extraction of mucopolysaccharide rich tissue. This was the first study to date ascertaining the need for the reagent in the lysis of molluscan tissue and the impact of extraction done without the use of BME measured in paired samples. The results from this study indicated that specimens could be extracted using with and without βME without subsequent impact on PCR amplification and sequencing. Eliminating the need for chemicals such as βME,

phenol and chloroform brings us one step closer towards the use of automated extraction techniques, and automated techniques such as the Biosprint Plant kit (Qiagen) used in this study could be used in high-throughput extraction of molluscs.

So the JRA5 results have proved to be promising for the User community. Firstly, a high-throughput extraction method can be applied to mucopolysaccharide-rich tissue. The DNA obtained from this method yielded viable sequences from specimens as old as 1964. In addition, through testing at partner Beneficiary institutions, the method of extraction does not seem to have a direct impact on the DNA obtained. Secondly, internal fragments worked to a certain degree, producing the full 660bp fragment or at least up to 400bp of the COI (Cytochrome Oxidase 1) barcoding region. Further optimisation or perhaps even redesigning one of the internal primer pairs would be necessary. Finally, protocols and primers designed in the MfN, Berlin (JRA5 Leader institution) could be applied to other laboratories on similar museum specimens, which was one of the key deliverables set out for this particular JRA.

JRA5 work has been written up in a manuscript currently in the final stages of preparation to submit: R Danabalan and T von Rintelen “The need for 2-mercaptoethanol in DNA extraction of molluscs” *Molecular Ecology Resources* (SYNTHESYS Project Deliverables 8.2 and 8.3).

Most JRA deliverables have been published in the scientific literature in high impact journals to encourage adoption by collections/laboratory managers, plus future Users of DNA from collections. JRA2 has published in Science (Impact Factor 31.027), JRA3 in Proceedings of the National Academy of Sciences of the United States (9.737) Systematics and Biodiversity (1.884), Molecular Ecology (6.275), JRA4 in PloSOne (3.730) and Proceedings of the Royal Society B: Biological Sciences (5.683) and JRA5 in Molecular Ecology Resources (*in prep.*; 7.432).

By focusing the JRA on DNA extraction, SYNTHESYS has increased the opportunities for Users to exploit a largely untapped facet of the 337 million strong collections. Users have been able to play an active role in generating new knowledge based on molecular and morphological studies. Providing the means for efficient DNA liberation from collections has greatly increased their accessibility to Users resulting in wider application for the research community, such as DNA-based systematic, ecological or population genetics studies plus applications to invasive/pest species and disease vector control.

Current DNA extraction policies are often very restrictive as institutions have a duty of care to protect and maintain their collections. The new standards for extraction developed by the JRA will hopefully lead to changes in DNA extraction policies, leading to better access to data held in collections. As a result there will be less need for researchers to go to into the field on collecting trips, there will be less damage to specimens and it will improve European research.

The impacts, via policy change effected by NA2, of JRA 1, 2 and 3 has ultimately been to vastly increase the range of specimens available to researchers across the ERA. These specimens include type specimens currently considered too scientifically valuable to sample. Many hundreds of thousands of specimens held by SYNTHESYS Participants collections fall into these categories, and provide the means to address questions about changes in biodiversity, the impacts of climate change and the origins of human culture within Europe. This is only possible using the resource represented by such specimens conserved in museums as a basis of comparison.

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

During the 4 years of the SYNTHESYS project, 1,002 awards from 2,246 applications have been spread evenly across the 4 funding calls, demonstrating sustained high demand, and

in total 10,464 User Days were delivered. Demand has been geographically far-reaching, with applications from institutions received from 90% of eligible countries (37 countries).

SYNTHESYS 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 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. Marketing in later calls was aimed at smaller, harder-to-reach research communities which, although successful, would not be expected to generate the same numbers of new applicants as earlier rounds. In spite of this, the percentage of new applicants plateaued in Call 3 and 4, showing that even in the later funding calls well over half of all applicants were new to SYNTHESYS.

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, 1,551 outputs have been entered on the SYNTHESYS site to date, though many of these are still 'in preparation' (757) or 'submitted' (84) [to a publisher/journal]. Of the 722 'accepted', 'in press', or 'published' User outputs, 536 are in peer-reviewed journals, with a further 28 contributing to books/monographs (see *Table 3* for summary). Further analysis shows that Call 1 User visits account for 38% of the outputs in *Table 3,* whereas Call 4 only 14%, 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 SYNTHESYS *FPVII* will be far higher as publications continue to be produced over the next few years.

As well as quantity, the quality of publications from SYNTHESYS 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. For example, following the trend from previous reports, *Zootaxa* and *Zookeys* (specialist taxonomy journals with JCR values of 0.974 and 0.864 respectively*)* account for over 15% (82 out of 536)) of SYNTHESYS User peer-reviewed publications. Additionally, 33 peer-reviewed articles (over 6%) are in journals with an impact factor rating of over 4.00, including one paper in Nature (details can be seen in *Table 4*).

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.

Table 3 Summary of selected User outputs (as of 3rd March 2014)

Type	**Accepted**	**In press**	**Published**	**Grand Total**
Book/Monograph	14	2	12	**28**
Database, CD or DVD	1	2	8	**11**
Non Peer Reviewed	31	5	90	**126**
Peer Reviewed	68	82	386	**536**
Thesis	3		18	**21**
Grand Total	**117**	**91**	**514**	**722**

Table 4 The Top 15 journals according to Journal Citation Reports (JCR) Impact Factor in which SYNTHESYS-funded TA Users research has contributed to publications (with a status of accepted / in press / published).

Journal	**JCR**	**Total**
Nature	38.597	1

Systematic Biology	12.169	1
Nature Communications	10.015	1
Proceedings of the National Academy of Sciences of USA	9.737	1
Molecular Ecology	7.432	1
Earth-Science Reviews	7.339	1
Global Change Biology	6.910	1
Proceedings of the Royal Society B	5.683	1
Journal of Ecology	5.431	1
Ecography	5.124	1
Cladistics	5.043	2
Journal of Biogeography	4.863	3
Journal of Human Evolution	4.094	7
Journal of Quaternary Science Reviews	4.076	1
Molecular Phylogenetics and Evolution	4.066	7
Grand Total		**30**

It is expected that the volume of SYNTHESYS-funded outputs will continue to rise after the end of the Project as Users continue to disseminate their work long after research visits are completed. Moreover, outputs that are published by high impact factor journals are often the result SYNTHESYS User visits contributing to larger scale long-term research projects. This can be exemplified by a number of recent publications in Nature resulting from SYNTHESYS *FPVI* funding (i.e. Access visits that took place between 2004-2009).

A comprehensive list of User outputs is included within the transnational access database.

Please note that the publications list uploaded on SESAM includes published outputs from the SYNTHESYS JRA only and not this extensive list from the Access Users.

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 USP was to consider the quality and the impact of the research being performed along with the plans for result publication. All output details have been collated online throughout the course of the project; and all scientific papers (and other outputs) acknowledge financial contributions made to support the research by the European Commission. Once published, all the output references have been presented back to the Directors (copied to the TAF Leaders) of the hosting institutions to facilitate a feedback between the TA Users and the senior staff of the hosting institutions.

IP resulting from the two NAs and the JRA have all been developed on an ‘open access’ basis; all documentation and software have been made freely available to Users.

The activities of the SYNTHESYS have also increased the quality of Participants’ collections: collection management standards have been increased, electronic data capture and processing have been enhanced, consistent and increased, access to previously restricted DNA has been increased, and as a result, Access to the collections has been much enhanced. Many Users have learned new skills during TAF visits which has been transferred to their home institutions. All this has led to an increase in the quality and quantity of European collections-based research. This positive effect is self-reinforcing, since Users having access to the collections has increased the quality of the collections, e.g., by identifying previously unidentified specimens, updating nomenclature. Furthermore, Users of the collections have created new information which has been incorporated in the collections databases. Thus, increased use of the collections has led to enhanced value in the collections.