SYNTHESYS+

Synthesis of Systematic Resources a DiSSCo project

D2.3 CATALOGUE AND RECOMMENDATIONS FOR THE DEVELOPMENT OF A PROACTIVE, EFFICIENT AND EVOLVING DiSSCo TRAINING PROGRAMME

Magalie Castelin, Jonathan Blettery (MNHN), Carole Paleco (RBINS), Isabel Rey, Celia Santos (MNCN-CSIC), Gergely Babocsay (HNHM), Patricia Mergen (MBG), Peter Giere (MfN), Nesrine Akkari, Edmund Schiller, Silke Schweiger, Karin Wiltschke (NHMW), Dan Holtstam, Jörgen Langhof (NRM), Tim Fulcher, Alan Paton (RBGK), Larissa Smirnova, Franck Theeten (RMCA), Heléne Aronsson, Matthias Obst (UGOT-GGBC), Hugo de Boer (NHM, UiO – DEST), Maarten Van Steenberge (RBINS – DEST), Ana Casino, Laura Tilley (CETAF)

Grant Agreement Number | 823827

Acronym | SYNTHESYS PLUS

Call | H2020-INFRAIA-2018-2020

Start date | February 2019

Duration | 27 months

Work Package | NA2

Task 2.3 Lead | Magalie Castelin & Jonathan Blettery (MNHN) – Carole Paleco (RBINS)

Delivery date | April, 30th 2021

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Table of contents

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Abstract

The competencies and skills to provide digital access to Natural History Collections (NHC) data, as well as curate and use that data are fundamental to meet the goals of the Distributed System of Scientific Collections (DiSSCo) Research Infrastructure (RI). Training activities targeting DiSSCo users and service providers (e.g., researchers, curators) constitute a cornerstone to these unprecedented objectives. This Deliverable 2.3 (D2.3) from task 3 within work package NA2 of the EU-funded SYNTHESYS+ project, provides the first catalogue of existing training, supplied by the NA2.3 partners and a few additional SYNTHESYS+ beneficiaries, that are potentially available to the DiSSCo community. The catalogue includes the following information on the training courses: training title, institution, city, target audience, length, in-take capacity, frequency, the position of the trainer, language, format, fees, and certification. In addition, D2.3 presents an in-depth assessment of the catalogue's training landscape in view of DiSSCo training needs. Based on this assessment, recommendations have been formulated for the construction of an efficient and proactive "DiSSCo training programme". More specifically, the assessment involved organising the catalogue of training into Key Training Areas (KTAs), which is a classification developed in NA2.3 to define the broad educational categories, domains and areas on which NHC trainers are currently concentrating. Using the KTA classifications, the training landscape was mapped to the DiSSCo Implementation Readiness Levels (IRLs) (i.e., Technological, Data, Organisation, Scientific and Financial), which correspond to the areas that DiSSCo needs to become proficient, in order to be fully operational. This analysis allowed for the identification of gaps, in terms of areas that are not covered yet or barely covered by the NHC related training landscape, but are critically needed to fulfil the IRLs. In general, it is advised that within the next 3 years the following activities are undertaken: (1) Establish an exhaustive inventory of the available training activities by broadening the criteria and widening the communities inventoried, for instance including the training programs of other related infrastructures and international initiatives as well as courses from the academic sector; (2) Address further training topics in complementary areas linked to the Natural History (NH) community, such as general IT skills or legal and financial aspects; and (3) Adjust and expand the boundaries of the catalogue according to the DiSSCo training objectives and needs once DiSSCo services are fully defined and integrated.

To coordinate efforts and better disseminate the catalogue of training developed, the collaboration will be pursued with the existing Distributed European School of Taxonomy (DEST), an

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

initiative run by the Consortium of European Taxonomic Facilities (CETAF). This platform will be used to allocate the catalogue of training resources through which the target audience will be efficiently outreached. As an optimal solution offered to the DiSSCo community and NH training providers, DEST is acting as the physical and online platform showcasing the community the NH training on offer. It also has an interface for the training providers, as well as for the trainees, dedicated to administration. The present catalogue of training activities will feed the DEST platform and contribute to increasing the richness and up-to-dateness of the content, its robustness in terms of data quality and criteria, and its networking role notably through the access to trainers' contact.

To conclude, the authors of this report recommend that the catalogue, methodology and analysis are used as a tool and guideline to develop the training activities suitable for DiSSCo needs, as a structure and content provider for DEST and as a collaborative incentive and set of information to foster the use of the existing rich training offer among the NH institutions and to promote and enlarge training activities throughout the entire community. Specific recommendations have been formulated in the text next to each component of the analysis and are listed in Section 5.

Keywords

DiSSCo (Distributed System of Scientific Collections); training landscape; key training area (KTA); training catalogue; training inventory; gap analysis; skills required for digitisation; Natural History (NH) collections; curation; research; Distributed European School of Taxonomy (DEST).

1. Introduction

Deliverable D2.3, falls under the SYNTHESYS+ project and more specifically under Networking Activity NA2 “Harmonisation of policies, best practices, training and support”. The aim is to define the overall requirements in these areas towards the implementation of the Research Infrastructure (RI) Distributed System of Scientific Collections (DiSSCo). The NA2 Task 2.3 concerns the development of scientific collections-centred training activities. This deliverable is a foundational guidance tool to facilitate the future work to be undertaken within the project DiSSCo Prepare and the subsequent phases of DiSSCo RI construction, in developing a comprehensive catalogue of training activities covering the needs of both the DiSSCo users and service providers.

More specifically, this report presents the first catalogue of existing training activities provided by the T2.3 partners and a few additional SYNTHESYS beneficiaries, followed by an in-depth analysis of the training landscape within the Natural History Collections (NHC) community. The analysis is based on several metrics such as the training topics, the different formats, the target audiences, etc. Additionally, the strengths and limitations of the training have been identified. Finally, a gap analysis has been conducted in relation with the known needs of DiSSCo. Based on the results obtained a set of recommendations have been formulated to best fulfil DiSSCo's training objectives.

The chapters of this document will address in detail the following aspects:

1. The background information and the scope of the task 2.3 and achievements;
2. The data and methods used to undertake the analysis;
3. The description and presentation of the results;
4. An overall discussion including recommendations;
5. Recommendation list;
6. Conclusions and perspectives.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

1.1 Background

1.1.1 The teaching roles of Natural History (NH) Institutions

Since their foundation, NH institutions (museums and botanical gardens) had educational duties towards the general public. This is reflected both by their exhibitions and by a wide variety of educational programs targeting schools, students, and citizens of all ages (Tamjid Mujtaba, 2018). More recently with the concepts of Open Science and notably the principles of Open Education, the role of NH institutions in terms of non-academic teaching has entered a new dimension by offering more and more training material in open access. However, many educational programmes for schools and citizens cannot make sufficiently nor optimally use of fragile or rare NH specimens themselves. By becoming digitally available the specimens and their associated information will in the future constitute a considerable added value for education, especially for remote training. Due to the quite limited number of such training activities in the present analysis, they were not included in this Deliverable.

Many NH institutions support academic courses in partnership with Universities or independently, where the in-depth knowledge on collections of the institutions facilitates general academic work while their staff provides specialized guidance and training in techniques of biodiversity science through Masters and Ph.D. supervision. NH institutions offer education for students in the form of internships or training activities that complement the classical university curricula by providing access to the unique environments of museums and botanic gardens with their rich history and collections.

Another group of training activities is closely related to the need of the experienced museum staff to ensure the preservation of, and access to collections. These training activities aim at exchanging expertise and best practices among the teams of different NH institutions (scientific, curation, collection management, and technical staff), young researchers, professionals from various disciplines, and also citizens.

In our analysis, we focus on the courses, where the participation of experts from NH institutions is crucial. Important is also the tooling-up of the scientific community in their different areas of expertise. The courses can address the needs of different stages in the careers of the staff ranging from early-career investigators, young professionals to advanced more experienced senior staff members. In this report we will refer to them generically as “training activities” to be understood within the scope explained here above. The training activities organized by NH institutions can be

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

roughly divided into four categories: Research-oriented; Collection-oriented; Digital-oriented and Policy-oriented. We give here below the definition of these categories and their scope in the context of this report.

Research-oriented training activities. NH museums and botanical gardens represent geo- and biodiversity repositories playing crucial roles for research in different biological and earth sciences disciplines. The long-term sampling and recording of specimens and specimen data through the centuries allow the reconstruction of natural patterns and processes. The scientific staff (i.e., researchers collecting, sampling, observing, or documenting new specimens in the field to study and answer scientific questions) is, therefore, best placed and qualified for providing the training activities in research disciplines such as taxonomy and nomenclature, identification and classification, evolutionary biology, ecology, systematics, and many more. In our analysis, we call it "specimen-based research". In the last two decades, this kind of training has been expanded by molecular and analytical research, where processing and analysis of big data are crucial. Scientific illustration courses are complementary to this group, representing one of the oldest training courses museums provide to future professionals to illustrate the publications. There has always been a great interest of the wider public for these very detailed and beautiful drawings. The courses have lately become popular with artists and the general public. The discipline evolved to include digital technologies either complementary to hand drawings or in some cases replacing them completely. Modern techniques of photo stacking or photogrammetry also enable sharp in-focus photographic images in all layers of the specimens, which was previously only possible with drawings. These techniques are now all part of such training.

Collection-oriented training activities. NH collections are rich repositories of objects representing the natural world. They are a physical knowledge base about the history of our planet and the evolution of life. They can as such be assimilated to a database containing comprehensive reference information. NH institutions require special buildings and digital infrastructures to properly house, exhibit, and conserve the collections. To make the collections openly accessible - both through physical and digital access - to not only visiting researchers, but also to the public, they also need to invest in appropriate and secure facilities.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

The vast variety of object types in the collections (from animals and plants to minerals or related notes, letters, and books), their sizes (from virus to giant dinosaurs), mode of conservation (e.g., dry or in fluid, thin-sections, or boreholes), vulnerability and great importance for research (e.g., type specimens, extinct species) and for mankind in general (e.g., Stanley's or Darwin's field notes) is challenging and requires experienced staff such as collection managers, curators, IT managers or legal experts, who can keep collections curated, safe, accessible and growing. This defines the importance of NH institutions participating in teaching activities and exchanging their experience in collection management. The training activities in general are structured around the basics of collection maintenance: preventive and remedial conservation, specimen curation, mounting or replicating specimens, handling and transportation, collection techniques, and also museology. There are also training activities on the quality of the storage environment and collection assessment. Indeed, due to digitisation, another group of training activities emerged, around comparison, selection, and use of collection management systems (e.g., Brahms, Botalista, Colhelper, Darwin Core, DINA, Specify, or Pluto). Additional collections are addressed next to the preserved collections such as the construction, the use and management of DNA and tissue banks, living collections, or seed banks.

Digital-oriented training activities. With the development of digitisation of NH specimens and the increasing amount of data becoming available for digital and extended specimen information, a need for new techniques and digital skills has emerged. This is reflected by the training activities around digitisation techniques, standardization and integration of collection data, and citizen science integration. NH collections documentation has evolved from solely handwritten labels and index cards on physical specimens into suites of interconnected over time enriched data (Page 2015, Nelson and Ellis 2018, Lendemer et al. 2019). Physical specimens, from minerals to vertebrates, are now linked to potentially limitless digital resources. This concept of digital enhancement was coined by Webster (2017) as the "extended specimen" and represents the next generation of NH collections (Schindel and Cook 2018). While training activities around digitisation are still closely connected to physical collections, the third group of trainings, digital-oriented trainings, targets a wide field of activities using the data alone: "discovery and registration of data sources; acquisition of data; curation and preservation of data; adding universally unique identifiers; standardising the content with controlled vocabularies; and making data freely and openly available to individuals and machines" (EU BON policy brief 5). Examples are trainings in data mobilization ("BID Capacity

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Enhancement Workshop "Biodiversity Data Mobilization", by GBIF), storage and use ("Long term data storage", by MfN), software for data analysis and processing ("QGIS for distribution data", by RMCA).

Policies-oriented activities. NH collections were never a sealed treasury, but on the contrary, they have always been open to researchers and professionals, provided the rules of handling and use were respected. Institutional internal policies and guidelines regarding handling of specimens and access to collections have been part of collection management training. Big data, open knowledge, looking for a fair and equitable share of biodiversity resources, and an increased interconnectedness in our modern world have brought new concerns about the origin, exchange, and use of genetic resources and data, and new challenges for data holders to apply correct policies and legislation when dealing with physical specimens and digital surrogates. Such training courses provide advice and guidance on how to deal with various international conventions (e.g., CBD, CITES, Nagoya Protocol) or how to apply a correct policy (Open Sciences, GDPR, Intellectual Property Rights (IPR), etc.).

1.1.2 Funding sources

A source of inspiration and funding for organizing training activities are project calls launched by the European Union's framework programmes over the last decades. How to attract external funding and write proposals suitable for NH institutions have also become part of existing training programs. A good example is the *Transnational Access* run by SYNTHESYS. The training programs of SYNTHESYS I to III aimed to implement a set of European competencies to improve performance and promote staff training and mobility within European institutions. Many other EU or nationally funded projects include training or capacity-building activities related directly or indirectly to geo- and biodiversity collections. For example, the training activities and manuals produced within the data mobilisation project EU-BON (Biodiversity Observation Network) can be cited since they contributed substantially to the DarwinCore data standard used by GBIF. Existing large Research Infrastructures have also relevant training programs about data standards, usage and protocols, for example, e-LTER on-site monitoring best practices, or ELIXIR on molecular data analysis pipelines around using the Galaxy platform. In the framework of the implementation of the European Open Science Cloud (EOSC) and European High-performance Computing Joint Undertaking (EuroHPC), numerous trainings are organized both at EU, national, and local levels on using supercomputing facilities, or machine learning algorithms. In terms of enhancement of the capacity-building of researchers, the European

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Network EURAXESS has a vast range of training activities, including issues related to open science, data, Intellectual Property Rights (IPR) management, and Career Development. EU programs like ERASMUS+ develop training programs, wherein our community Biotalent was a successful example. Many of these courses are under an Open Educational license which enables the reuse of the training material by third parties.

Participation in courses can be free or subject to fees. When providing various types of certificates, they may have a different impact on the career of the participants. Due to the limitation of seats, there is sometimes a selection process to be a participant. However, the increase of online courses enables to accommodate more participants and decreases the need for a stricter selection of participants. These aspects are presented in the results sections in more detail. Obviously, the availability of additional funding encourages the organization of new training, after the termination of the funding period the training may still be available on request or as an online static training resource.

1.2 Scope

1.2.1 Participants

The inventory of the training activities was restricted to the twelve institutional and organisational partners participating in the SYNTHESYS+ Task NA2.3 (MNHN, RBINS, MNCN-CSIC, NHMW, UGOT, RBG Kew, RMCA, MfN-SNSB, Meise BG, HNHM, NRM, with GBIF being distributed within some of the other twelve and CETAF providing support and guidance through DEST). The first reason for limiting the inventory to twelve institutions/organisations is that among the 32 partners of the SYNTHESYS+ project, these twelve have the necessary critical mass and excellent expertise to tackle the tasks and challenges to deliver meaningful and exploitable outputs on which further institutions can build in future. The second reason is that gathering data on existing available training activities is a long and arduous task. The timeframe of the Task NA2.3 did not allow enough time for partners to expand the inventory of training activities to the 32 institutions of the SYNTHESYS+ consortium. Discussions with WP2 partners from DiSSCo PREPARE projects are in progress so that the inventory delivered under SYNTHESYS+ constitutes the baseline for further improvement and enlargement under DiSSCo Prepare.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

1.2.2 Coverage and range

Some training activities arise from individual initiatives, for example those connected to specific academic research or software applications. Still there are also institutional training programs offering a selection of courses around specific topics, of diverse length and specificity. Such is the case of the programme offered by the Natural History Museum of London, UK (NHM-London), "The Identification Trainers for the Future" (2014-2018), or the Natural Sciences Museums of Berlin, DE (MfN-Berlin) workshop on "Basic Collection Techniques" consisting of three parts: preventive conservation techniques; documentation, digitisation and labelling; requirements on different collection types (dry and wet). Additionally, training programs can be initiated and funded by global biodiversity initiatives, such as the FishBase training programme of the Royal Museum for Central Africa in Tervuren, BE (RMCA-Tervuren) for African students (cf. present catalogue), or trainings provided in the framework of Global Biodiversity Information Facility (GBIF) as it is the case on BIDs and capacity building programs. Moreover, trainings may be organised on a national level, or being part of collaborative initiatives between different institutions and across countries, like ForBio in Norway.

On the European level the project-based EU COST action for "Mobilising Data, Experts and Policies in Scientific Collections" (COST Mobilise) with limited timeframe and more sustainably, the Distributed European School of Taxonomy (DEST) on long-term basis are major players promoting transfer of knowledge built thanks to the efficient combination of natural science collections and technology. In this context it is certainly worthwhile to aim at having trainings developed within short term projects and that become long term sustainable under DEST.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

2. Methodology

2.1 Catalogue of existing trainings and Key Training Areas (KTAs)

2.1.1 Construction of the catalogue

To collect an aggregated list of existing and available training courses, the NA2.3 SYNTHESYS+ partners were invited to seek, within their institutions and organisations, training resources potentially available for the DiSSCo community (i.e., any training activity that enables individuals/researchers/curators to navigate collections information effectively, as well as optimise digital access and utilisation of NH collections). The query was carried out by reviewing existing documentation, and contacting or interviewing colleagues in charge of training activities. The list of training activities and related data were aggregated in a shared Google spreadsheet along with the titles of training activities, the following related data were collected over two years: *Institution*; *city; targeted audience*; *length (in days)*; *in-take capacity (number of participants)*; *frequency*; *the position of the trainer; language*; *physical, virtual, or hybrid* (i.e., *physical and virtual), and the possibility to be transformed into virtual; fees (if applicable)*; and *certification (if applicable)* (Appendix A). The focus was put on training activities currently provided and taking place at the institution's premises. Training activities under development have not been included in the statistics nor those related to capacity building which are often taking place out of Europe and less directly linked to the collections.

2.1.2 Training activities and classification by KTAs

Each training activity was classified under one or several Key Training Areas (KTA, defined as broad educational categories/domains/areas on which NH trainers are currently concentrating). This classification aimed at organising the listed training activities into a reduced number of main areas to best characterize and analyse the diversity of the catalogue. KTAs were delimited and defined based on the description/content/topic of each training activity, and the professional judgment of partners. Eight main KTAs were defined: "*History of collections*"; "*Basic knowledge related to collections*"; "*Specimen*"; "*Data*"; "*Equipment use*"; "*Policy & legislation*"; "*Training multipliers*"; and "*Text & Media*". To gain a higher level of specificity in the characterization of the training activities, it was decided during a workshop, held in Paris in February 2020 for collecting feedback from Task partners (see Report of the workshop under the Milestone MS26), that the classification should follow a branching pattern. Some of the main KTAs were divided into subcategories of level #1, and some of the subcategories of level #1 were further subdivided into subcategories of level #2. For

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

instance, the main KTA "*Data*" was divided into six subcategories of level #1: "*data curation*"; "*data management*"; "*data standard & interoperability*"; "*data processing/use*"; "*databasing*"; and "*software skills*". Later on, the subcategory "*software skills*" was divided into two subcategories of level #2: "*applying evolutionary models*"; and "*analytic research*" (this example is shown in Figure 1). This branching classification allows each training activity to be sorted following their content and should help future users of the DEST, to streamline their navigation and access to training activities of interest. Branching classification and definitions are given in the following table.

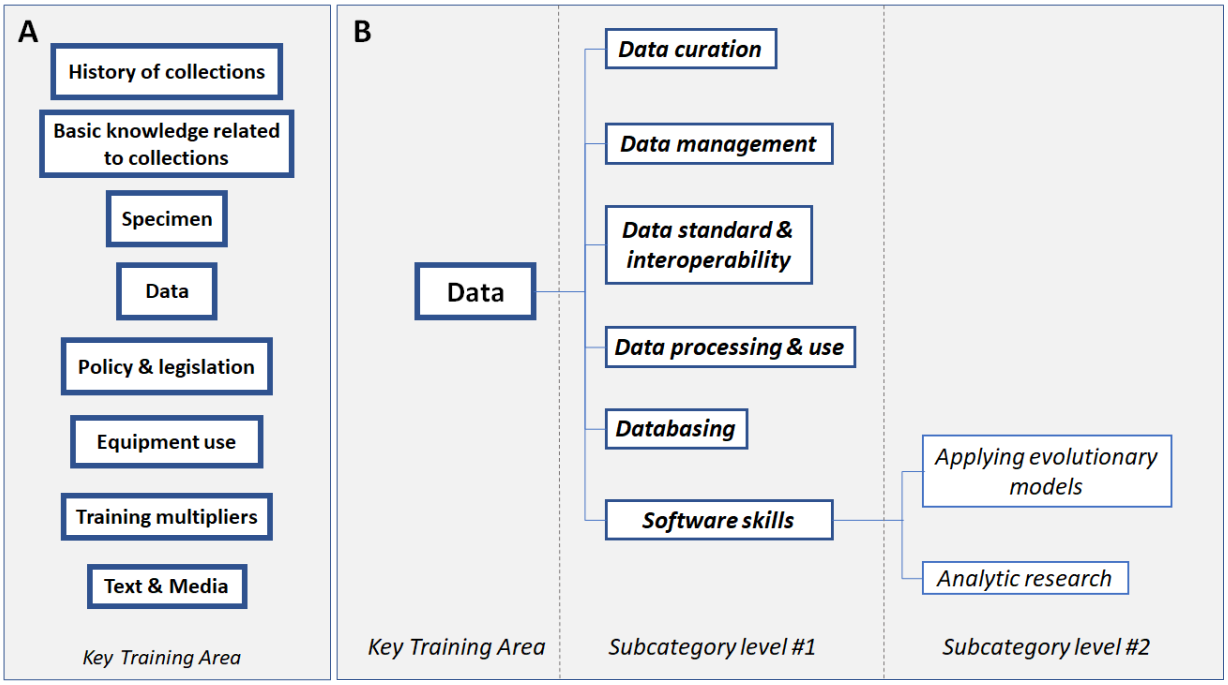


Figure 1. A) The eight Key Training Areas (KTAs) defined for the present study. B) Example of branching classification of the subcategories of levels #1 and #2 within the KTA "*Data*".

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Table 1. Definitions of KTAs and subcategories #1, with self-explanatory subcategories #2

KTA 1: History of collections: it covers history of NH *collection* (e.g., institutional history, decolonisation, collector, expeditions).

KTA 2: Basic knowledge related to collections: All the activities surrounding the documentation and accessibility of NH collections.

KTA 3: Specimens: Physical objects or observations (e.g., photo, sound...), or a direct derivative (e.g., image of specimen), but not the product of analysis. Not necessarily yet curated. Not yet hosted by the institution.

***#1. Museum specimens**: Physical object or observation (e.g., photo, sound), or a direct derivative, (e.g., image), not the product of analysis. Not necessarily yet curated, hosted by the institution, and it must be inventoried.*

#2: Collection/Sampling of specimens.

#2: Field sampling.

#2: Exhibition.

***#1. Specimen curation**: Knowledge required to curate NH collection (e.g., care & maintenance, collection management, or preventive conservation).*

#2: Care & maintenance

#2: Collection management

#2: Preventive conservation

***#1. Digitisation**: Knowledge required to digitize NH collection (e.g., databasing metadata, data quality, imaging, digitisation techniques).*

#2: Databasing metadata.

#2: Data quality.

#2: Imaging.

#2: Digitisation techniques.

***#1. Specimen based research**: All research activities based on NH collection.*

#2: Taxonomy.

#2: Biodiversity.

#2: Evolution.

#2: Ecology.

#2: Conservation.

#2: Nomenclature.

#2: Geodiversity.

#2: Genomic and molecular science.

#2: Data analysis.

#2: Methodology, Mineralogy & Geology.

#2: Methodology Archaeology.

KTA 4: Data: Ability to provide, enrich, and/or use NH *collection* data in a consistent and harmonised manner.

***#1. Data curation**: Knowledge required to curate NH collection (e.g., accessibility, data quality).*

***#1. Data management**: All activities related to managing data as a valuable resource (e.g., storage).*

***#1. Data standards and interoperability**: All activities addressing the ability of systems and services that create, exchange, and consume data to have clear, shared expectations for the contents, context and meaning of that data.*

***#1. Data processing/use**: The carrying out of operations on data, especially by a computer, to retrieve, transform, or classify information (e.g., genomic data).*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

#1. Databasing: *Knowledge and practices required for the addition of data to a database.*

#1. Software skills*: Abilities to efficiently use computer programs and applications.*

#2: Applying evolutionary models.

#2: Analytic research.

KTA 5: *Equipment use*: Ability to run scientific equipment (e.g., DNA sequencing, 2D/3D imaging facilities, remote sensing).

KTA 6: *Policy & legislation*: Covers all areas of legislative practices tied to both internal (e.g., collection strategy, data management, or IT strategy) and external policies (e.g., Convention on Biological Diversity, Nagoya Protocol on ABS, or any Global activity contributing to data alignment initiatives for bio-geo collections).

#1. Policy mandates*: Laws, regulations and conventions that are mandatory*.

#2: Cultural heritage: *cultural goods protection legislation e.g., Council Regulation (EC) No 116/2009 of 18 December 2008 on the export of cultural goods.*

#2: CITES: *Convention on International Trade in Endangered Species of Wild Fauna and Flora.*

#2: ABS: *Nagoya Protocol on Access and Benefit-sharing and national laws of provider countries.*

#1. Implementation for collection: *Policy and protocols established by institutions to comply with all current legislation.*

#2: Curation: *e.g., rules for visitors, loan policy, access policy, shipping policy.*

#2: Research: *policies and procedures addressing the ownership of research material and data, their storage, access and retention beyond the end of the project.*

KTA 7: *Training multipliers*: Use of digitisation shared tools, databases, or platforms.

#1. Training the trainer: *Skills enabling to best educate other adults in their workplace or community.*

#1. Citizen science: *The collection and analysis of data relating to the natural world by members of the general public, typically as part of a collaborative project with professional scientists.*

KTA 8: *Text & Media*: Every description or representation of the world, fictional or otherwise. Skills to describe or define reality, using textual, numeric, or similar formats.

#1. Knowledge transfer: *Sharing or disseminating knowledge.*

#2: Exhibition.

#2: Media.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

2.2 Mapping of existing training to DiSSCo needs

At this early stage of the DiSSCo RI preparatory phase, the training programme needs to assist the community in the further implementation of the different objectives of DiSSCo. For this purpose, NA2.3 partners were invited to focus on the five dimensions that the EU-funded project DiSSCo Prepare considers as overarching for the further construction and operation of the RI, i.e., the Implementation Readiness Levels (IRL): Science, Data, Technological, Organizational, and Financial drivers. The participants were requested to identify, based on the current training landscape, potential gaps toward the development of a robust and efficient DiSSCo RI training programme. These five dimensions correspond to the recognised facets of the challenge ahead to improve DiSSCo's ability to become fully operational.

2.2.1 Assessment of DiSSCo training objectives

The content and goals of each IRL were carefully studied to assess their potential training needs. For each IRL, a summary of objectives and goals was reported in a shared table. For instance, "*Data"* was summarized as "*the capacity to provide and enrich data in a consistent, harmonised manner, compatible with the overall RI"* (DiSSCo PREPARE Grant Agreement). Based on these IRL summaries a subgroup of NA2.3 partners analysed potential training objectives. Each dimension was then divided into training objectives, representing identified DiSSCo training objectives. Thus, "*Data"* was further divided into six training objectives: "*Pre-digitisation curation"; "Standardisation interoperability"; "Data curation"; "Digitisation"; "Databasing";* and "*Data sharing"*. Training objectives were then associated with a proper definition. For example, "*Digitisation"* was defined as: "*All processes that enable converting a specimen/object into a digital format (technicity, standards assessment, protocols)"* (cf. Table 2). A total of 29 training objectives were defined and considered as the main objectives of the DiSSCo training programme (Table 2).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Table 2. Definitions of the 29 DiSSCo training objectives (with a capital letter - S, D, T, O, and F - indicating the respective IRL the training objectives is attached to) defined for the present survey as main training goals of the DiSSCo Training Programme

IRL Science:

***#S-1. Collection/sampling of specimen**: Collecting within a scientific framework.*

***#S-2. Specimen curation:** Knowledge required to curate NS collection.*

***#S-3. Museum specimen-based research:** All research activities based on NS collection.*

#S-4 Taxonomy

#S-5. Ecology

#S-6. Biodiversity

#S-7. Geodiversity

#S-8. Evolution

***#S-9. Data analysis:** Software skills, analytical and technical research, or data manipulation/transformation.*

***#S-10. Publication and outputs:** Publication process and outputs measuring*

***#S-11 Citizens science - training the trainers:** Use of digitisation shared tools, databases, and platforms.*

***#S-12. Citizens science - training citizens:** Use of digitisation, shared tools, databases, and platforms.*

IRL Data

***#D-13. Pre-digitisation curation**: Accepting & accessioning specimens/objects, workflow to get them ready for digitisation (e.g., labels, proper storage room, correct names and metadata), or metadata capture and handling.*

***#D-14. Standardisation interoperability**: Data processing workflows that convert the structure of disparate datasets into a common data format (e.g., data FAIRness (Findable, Accessible, Interoperable and Reusable data), specimens data refinery, or molecular standards and processes…).*

***#D-15. Data curation:** All the activities surrounding the documentation and accessibility of NS collection data.*

***#D-16. Digitisation:** All processes that enable converting a specimen/object into a digital format (technicity, standards assessment, protocols).*

***#D-17. Databasing:** Integration and management of a specimen/object and associated (meta)data in databases.*

***#D-18. Data sharing:** Dissemination through public e-portal.*

IRL Technological

***#T-19. Operating:** Running scientific equipment (e.g., sequencing or 2D/3D imaging facilities).*

***#T-20. Development:** Conceptual design and implementation of software and digital tools (e.g., data architecture, data carpentry).*

***#T-21. IT data management:** All activities surrounding the manipulation, protection, storage, or safety of data.*

***#T-22. e-Monitoring:** Digital management services (e.g., E-service, helpdesk, mediation services): administrators, and users.*

IRL Organisational

***#O-23. Policy:** Covers all areas of legislative practices tied to both internal (e.g., collection strategy, data management, or IT strategy) and external policies (e.g., Convention on Biological Diversity, Nagoya Protocol on ABS, or any Global activity contributing to data alignment initiatives for bio-geo collections).*

***#O-24. Governance:** Legal and decisional expertise related to inter-community partnership activities and organisation (e.g., decision making, governance levels, stakeholders' engagement).*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

#O-25. Management: *Defining strategies and management plans, administering and coordinating projects & activities (incl. Services).*
#O-26. Communication: *Expertise related to institutional, international, public, media, partnership, and internal communication.*

IRL Financial

#F-27 Management: *Management of financial and budgetary strategy: expertise in all accounting, management, treasury, tax and financial communication functions.*
#F-28. Financial awareness: *Drawing up and monitoring project budgets, the balance of payments, payroll, equipment investments.*
#F-29. Basic information: *Provision of financial information on projects.*

2.2.2 Assessment of DiSSCo training needs

A contingency table was built and shared with partners in an Excel Google Drive document, in which every training activity in the catalogue (displayed vertically) can be cross-matched against each training objective of the five dimensions (displayed horizontally). Partners were invited to interview trainers of their respective institution and organisation to collect information and fill the contingency table. The options for filling the form were: YES; NO; PARTLY; or DON'T KNOW. With, for example, *Yes* meaning that the training activity matches the dimension objectives, *Partly* meaning that only a part of the training activity matches the dimension objectives. To receive feedback and additional comments from trainers and partners on the task, a set of specific questions was created, such as:

- *How do you feel about matching existing training activities with the DiSSCo IRLs?*
- *For which IRLs do you find most difficult to match training activities, if any?*
- *Do you think your institution is ready to deliver additional training activities linked to the IRLs?*
- *Do you think your institution will be able to provide additional training activities in the future (linked or not to IRLs)?*
- *Any other comments?*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

2.2.3 Identification of DiSSCo training gaps

Training gaps can be defined as areas that are either not or barely covered in the current NH training landscape, but critically needed to reach defined objectives. Gaps detection thus requires: *(i)* a knowledge of what exists; *(ii)* a knowledge of the objectives to be reached; and *(iii)* both qualitative and quantitative comparisons between these two sources of knowledge. To this end, the proportions of *Yes* and *No* in the contingency table '*training activities and related data x IRL* training objectives' were analysed. The current number of training activities fulfilling each IRL and related training objectives was carefully studied in regard to the identified DiSSCo needs to highlight less represented IRLs training objectives, or areas where the current training programmes do not specifically address or prioritise DiSSCo RI objectives.

2.3 Survey tools

The methodology was implemented using an Excel spreadsheet (shared on Google Drive) named "NA2.3 Catalogue and survey of trainings activities" and composed of 7 spreadsheets:

1. *"Guidelines"*, setting out the purpose of the survey, the content of the document, the rules for filling it and updating it;
2. *"KTAs"*, with the tree structure of the defined KTAs, related subcategories, and associated definitions;
3. *"IRLs definitions"*, with a definition of IRL main focus and a definition of each training objective;
4. *"KTA analysis"*, contingency table containing vertically the catalogue of existing available training activities, and horizontally training features (city, length, support, fees…) and KTAs with related subcategories of level #1 and #2;
5. *"IRL gap analysis"*, a contingency table containing vertically the catalogue of existing available training activities, and horizontally the IRLs and training objectives, as well as an empty field at the right end for collecting individualised comments about each training activity;
6. *"General feedback"*, proposing 4 questions (see above) *plus* an empty field to collect open comments;
7. *"NA2.3 partners"*, listing all partners/institutions and contacts.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

The collection of the data (training activities, related data, KTAs and IRLs training objectives, and feedback) were collected and edited over two years and finalised in March 2021. After the aggregation phase, the catalogue and metadata were cleaned, consolidated through feedback among partners, and standardised according to the rules defined in Appendix A. Once everything was completed, harmonized and cleaned, a series of calculations were performed. All figures of the present document are based on the content of this dataset (Appendix A).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3. Results

3.1 NH training landscape

3.1.1 General description of the catalogue

In total, 91 training activities were collected from the 12 participating institutions and organisations (MNHN, RBINS, MNCN-CSIC, NHMW, UGOT-GGBC, RBGK, RMCA, MfN, SNSB, Meise BG, HNHM, NRM). The five training activities that were still under development at the time of the present work and the three training activities listed as capacity building were removed from the following analyses, thus leading to **the study of 83 training activities.**

The analysed training activities are different in their typology (e.g., format, length, audience, etc.). Most of them (70) are given in a classroom, physically or virtually, by a trainer to a limited number of participants, and range in length from 1 to 20 days, with a median of 4.45 days. Ten training activities consist of a broadcast online video (from 13 to 80 minutes), for which the number of participants or views cannot be estimated. One is a one-hour seminar given virtually every two weeks with no capacity limit: « *Scientific Data and Information Seminar* » offered by MfN. Two are given physically (in a lab or a classroom) and exceed three months: the course « *Zootaxonomy* » offered by HNHM lasts 2 semesters, includes 2 theoretical courses, and 4 seminars (of an hour each) per week, *plus* 1 week of field course; the course « *Training course on FishBase and fish taxonomy* » offered to African researchers by RMCA lasts 3 months. The format of these 2 cases differs from the others as the training provided by the HNHM corresponds to a university course, while the one provided by RMCA is similar to a “residence” at the museum.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.1.2 Classification of the catalogue by KTAs

The graphs below (Fig. 2 A, B, C) are based on 83 training activities of the catalogue and give the numbers of training activities per KTAs (Fig. 2, A) and associated subcategories of level #1 (Fig. 2, B), and subcategories of level #2 (Fig. 2, C). The most represented KTA is "*Specimens*" with 49 training activities (59% of the catalogue), followed by "*Data*" with 24 training activities (29%), "*Policy and legislation*" with 12 training activities (14%), "*Basic knowledge related to collections*" with 5 training activities (6%), and "*Training multipliers*" with 3 training activities (4%). This also highlights the KTAs that are less-well represented, such as: "*Equipment use*" (1%), "*History of collections*" (0%), and "*Text and Media*" (0%).

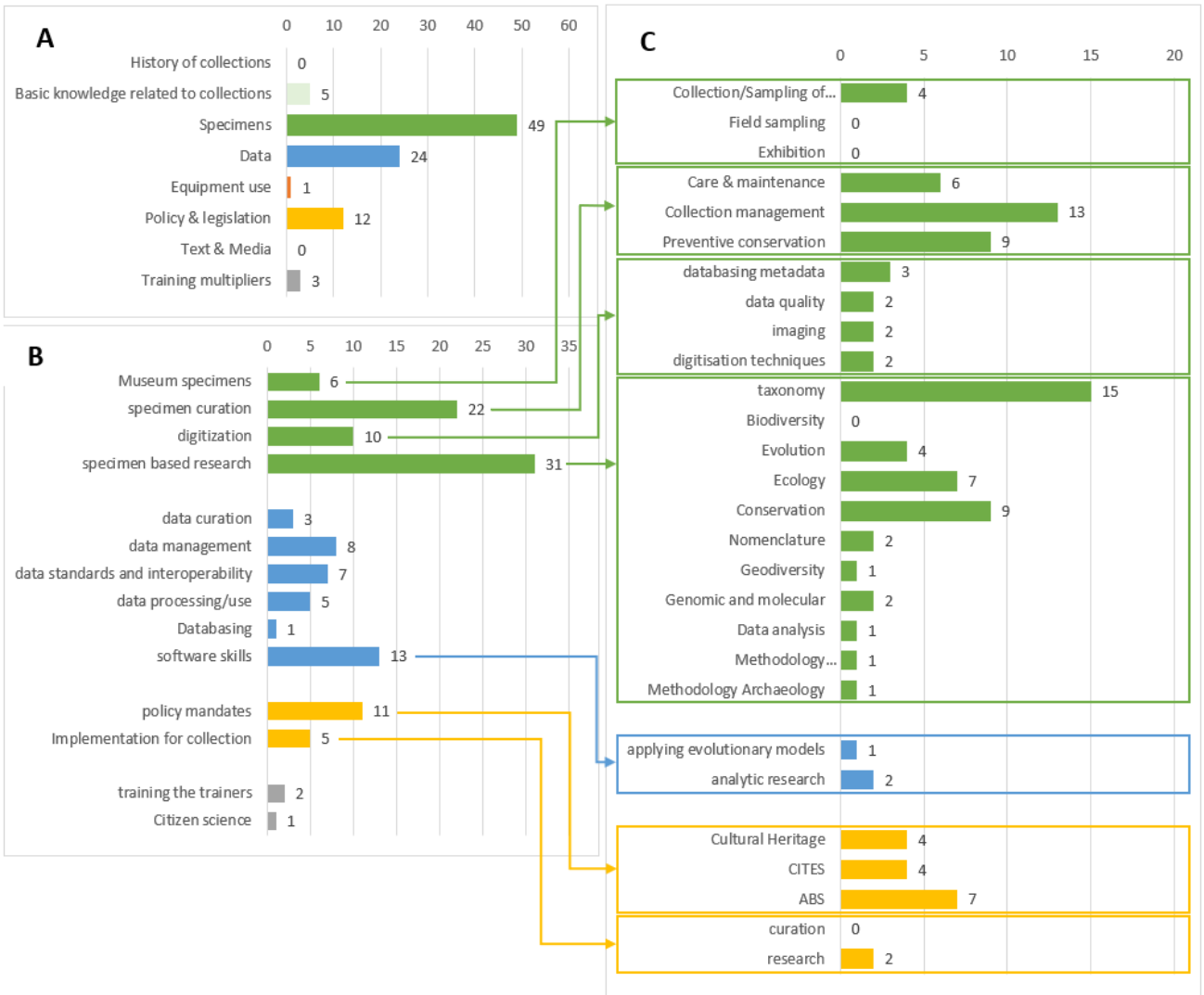


Figure 2. Number of training courses per: (**A**) KTA; (**B**) subcategories #1; and (**C**) subcategories #2. *[Some courses correspond to several KTA, thus, the number of matches per KTA and subcategories does not correspond to the strict number of courses.]*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

The KTA "*Specimens*". Under this KTA, the subcategories of level #1 that are the most represented are: *#1 specimen-based research* (31 training activities), *#1 specimen curation* (22), *#1 digitisation* (10), and *#1 Museum specimens* (6).

- ***#1 Specimen-based research*** addresses highly specialized topics such as ***#2 Taxonomy***, ***#2 Ecology***, or ***#2 Conservation*** and is taught by almost all the institutions of the catalogue.
- ***#1 Specimen curation*** addresses specialized topics such as ***#2 Care & maintenance***, ***#2 Collection management***, and ***#2 Preventive conservation***. Among the 22 training activities of this category, many address current areas of concerns. Berlin (MfN) proposes courses in ***Disaster preparedness***, Madrid (MNCN-CSIC) covers ***Conservation, use and management of DNA and tissue banks***, Wien (NHMW) teaches ***Handling historical and fragile museum specimens and types for scientific research***, Stockholm (NRM) addresses ***Training in identification of minerals/rocks with analytical techniques, organizing systematic mineral collections, and risk management***, London (RGBK) addresses ***Use and Curation of Biocultural (ethnobiological) Collections*.**
- ***#1 Digitisation*** is covered by 10 training activities and address a wide range of skill techniques covering ***#2 Databasing metadata, #2 Digitisation techniques,*** and ***#2 Data quality*** with for instance: ***BID Capacity Enhancement Workshop: Biodiversity Data Mobilization*** by GBIF-France. The subcategory ***#2 Imaging*** is covered among others by Berlin (MfN) ***Digitisation techniques***, Wien (NHMW) ***Scientific illustrations: scientific drawing, multifocal imaging and scanning electron microscopy***, and MNHN Paris ***Scientific illustrations***.

The KTA *"Data"*. The subcategories of level #1 that are the most represented in the KTA *Data* are: *#1 Software skills* (13 training activities), *#1 data management* (8), *#1 data standards and interoperability* (7), *#1 data processing/use* (5), *#1 data curation* (3), and *#1 databasing* (1).

- ***#1 Software skills.*** Among the 13 training activities, many address current needs for data intensive science research, such as for example: ***Integrative taxonomy in the "big data" era*** (MNHN, Paris); ***Phylogenetic reconstructions using Tree analysis using New Technology (TNT)*** (NHMW, Wien); ***QGIS for distribution data*** (RMCA, Tervuren); or ***Python for biologists*** (UGOT, Gothenburg).
- ***#1 Data management*** includes 8 training activities with for example: ***Specify/DINA*** (MfN, Berlin); ***Best practice in plant tissue collection and preservation*** (RGBK, London).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

- ***#1 Data standards and interoperability*** includes 7 training activities, with among others: ***Data standards in natural sciences*** (MfN, Berlin).
- ***#1 Data processing/use*** includes 5 training activities, with among others: ***BID: data mobilisation, data quality, usage of data, project management, niche modelling, policy making*** (GBIF-Meise).
- ***#1 Data curation*** and ***#1 Databasing*** are less well covered with only 3 and 1 training activities respectively. ***#1 Data curation*** being covered only by the GBIF-Paris and GBIF-Meise courses.

The KTA *"Policy and legislation"*. Regarding subcategories within this KTA, the most represented subcategories is *#1 Policy mandates* with eleven courses.; *#1 Implementation for collection* is represented in five.

- ***#1 Policy mandates*** refers to mandatory laws, regulations and conventions. Seven courses specifically address that category, with for example: ***Access and Benefit Sharing*** (MfN Berlin/Munich)**;** ***CITES in Scientific Collections: Identification*** (MNCN-CSIC, Madrid); ***Due Diligence for utilizing genetic resources in the context of the Nagoya-Protocol*** (NHMW, Wien); ***An introducing to "Access and Benefit Sharing" in public research*** (RBINS, Brussels)***.***
- ***#1 Implementation for collection***. A single course addresses specifically this topic: ***CITES in Scientific Collections: Management*** (MNCN-CSIC, Madrid).
- Four training activities cover both ***#1 Policy mandates*** and **#1** ***Implementation for collection***: ***Best practice advice on implementing ABS policy mandates*** (MNCN-CSIC, Madrid); ***Best practice advice on implementing Cultural Heritage policy mandates*** (MNCN-CSIC, Madrid)***; Diploma in Management, Care and Exhibition of Natural Sciences Collections*** (MNCN-CSIC, Madrid); and ***Field work and relevant policy framework (CBD, NP, Fieldwork Planning, TK, and CITES)*** (RGBK, London).

The KTA *"Basic knowledge related to collections"* is poorly represented in the NH training landscape with only 5 training activities dispensed by NHMW Wien in German, on video. This topic is broad and might overlap w3.1.2 with other KTAs or be part of training activities but not considered as a main focus. Collection staff learn much through training on the job, or on courses dealing with particular aspects of collection management or the use of collections and their data in research. These 5 activities dispensed by NHMW Wien provide a broad overview of general knowledge related to NH

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

collections. The lectures give a broad overview of evolution by using museum specimens for explanations. They are intended to increase interest and engage a broader public with NH institutions and collections.

The KTA *"Training multipliers"* is poorly represented in the NH training landscape with only 3 training activities: dispensed by NHMW Wien (‘***Citizen Science’***), Meise BG (‘***Courses and workshops for high and elementary school including aiming at the teachers’***) and RBINS Brussels (‘***Training citizens on the introduction to taxonomy’***).

The KTA *"Equipment use"* is poorly represented in the NH training landscape with only 1 training activity dispensed by MNCN-CSIC Madrid.

The KTA *"Text and media"* is not represented in the current NH training landscape as it is not the main focus of activities, it may have been overlooked by institutions or during the survey. However, this KTA may be aimed at future training activities.

The KTA *"History of collections"* is not represented in the NH training landscape as a topic on its own but is part of existing training activities to provide background knowledge. For instance, it is part of the courses offered by MfN Berlin, a focused overview of the history of NH collections is provided in one session of the ***Training on Natural History Collections***. Other training activities (e.g., Basic Collections Techniques) incorporate aspects of the history of natural science collections, e.g., when providing the background on preservation of specimens in liquids.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.1.3 Analyses of the metadata catalogue

3.1.3.1 Capacity and average length of individual training activities per KTAs and institutions

The KTAs that have the highest capacity (Fig. 3, A) and the largest amount of training time (Fig. 3, B) are *Specimens* (778 people - 329 days), followed by *Data* (512 people - 79 days), and *Policy and legislation* (255 people - 46 days). Interestingly, *Data* has a large number of dedicated days (Fig. 3, B) for a small number of participants (Fig. 3, A), suggesting that *Data* courses are long lasting courses. Some caution has to be taken regarding the following data: the exact number of participants per KTA could not be estimated as some training activities are associated with several KTAs. Besides, this analysis does not consider the frequency of the training course. For instance, some training activities occur twice a year, others take place on demand. Finally, these are rough numbers. For instance, when the collected data corresponded to a range with a min-max number of participants, the maximum number of participants was selected (e.g., if 10-20 participants, we used 20), and when the given information was "group", the number of participants was estimated to 5.

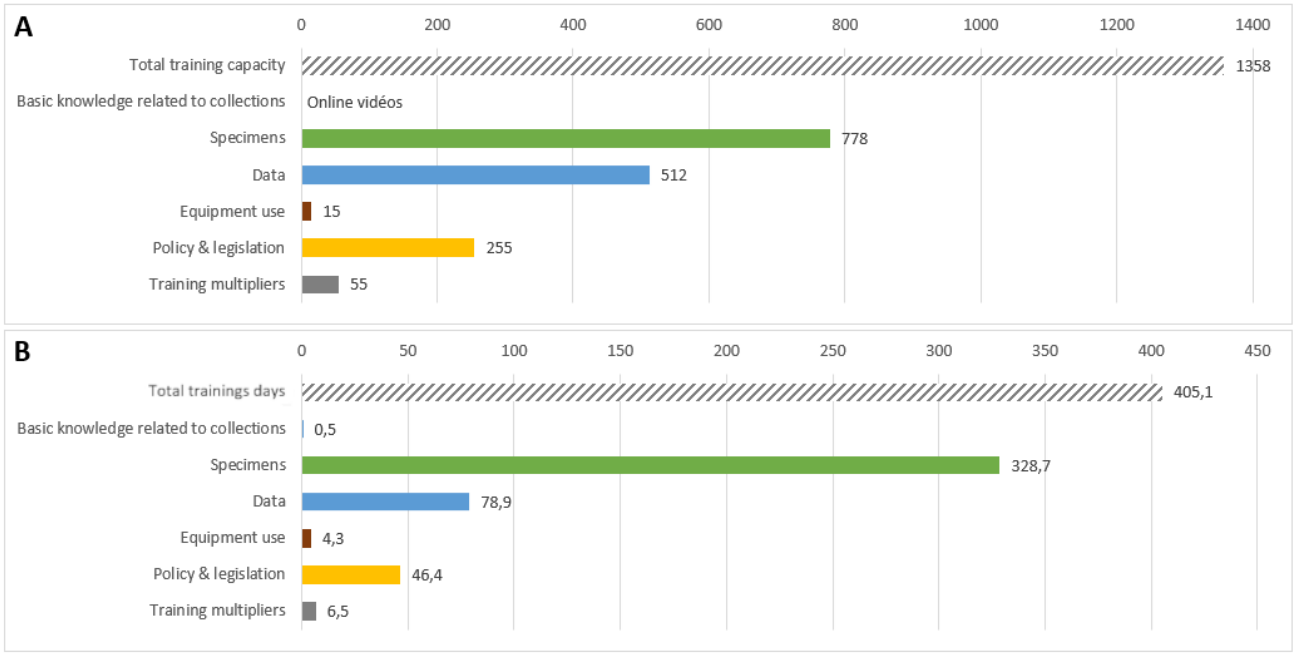


Figure 3. (**A**) Distribution of capacity and (**B**) number of days per KTA.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

In these graphs, the total training capacity *in blue* and total training days *in red* were gathered together (Fig. 4, A) and dispatched among training institutions (Fig. 4, B).

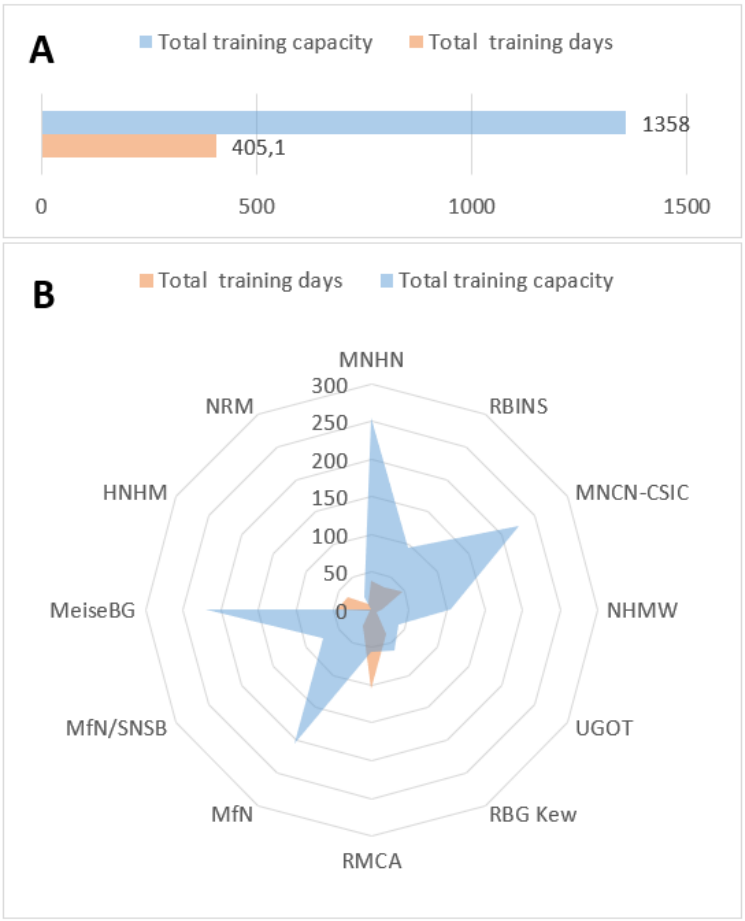


Figure 4. Average length and in-take capacity of individual training activities: **A)** Total number of training days and in-take capacity, **B)** training days and in-take capacity per institution.

The total number of training days is estimated at 405 days for 83 training activities, of which videos represent only 1 day (each video was quantified at 0.1 days).

The total number of training capacity is estimated at 1 358 people for 83 training activities. However, within this calculation the online videos and the unique seminar (11 training activities with unlimited capacity), and 1 course of 35.6 days (HNHM) with missing data for capacity are not represented in the capacity calculation.

Overall, across institutions the training capacity appears quite large compared to the number of seats suggesting broad training impact (high number of venues) (Fig. 4, B). Training activities range from 4 to 100 people (median of 15 people), with for example 7 activities gathering 420 persons, and 64 activities gathering 938 persons. This shows that some training activities target a large number of people at once. Concerning the length, training activities range from 0.5 to 64.5 days (median of 3.5 days, online videos and seminar excluded). Sixty-seven activities last two weeks or less (half day) and only five activities last between 20 to 64.5 days. This shows that generally training activities tend to be short.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.1.2.2 Training format: physical, virtual, and hybrid learning

Three types of support were considered in the present study: (1) **Physical** (face-to-face training); (2) **Virtual** (online classroom and/or video movies); and (3) **Hybrid** (training activities that incorporate elements of both the physical and virtual classroom in the same course). As some courses are delivered through multiple supports, the number of times a support is used does not correspond to the total number of courses. For instance, in the KTA *Specimens* (Fig. 5, A) 41 activities are offered in a physical classroom, of which 39 are strictly physical (not represented in Fig. 5, A) and 2 are offered both in physical and virtual.

Figure 5. Training support, in terms of physical, virtual, and hybrid learning: **A)** Distribution of support modes per KTA; **B)** Distribution of physical and hybrid support modes that can be developed virtually; **C)** Distribution of physical and hybrid support modes per KTA that can be developed virtually.

All KTAs pooled, 55 training activities are strictly physical, 14 are strictly virtual, 9 are strictly hybrid, 4 are offered in physical and virtual (the training can be given either physically or virtually), and 1 is virtual *plus* hybrid (cf. Table 3). The compilation of these data suggests that 77% of the training activities are offered fully or partly in physical (physical and hybrid), 23% are virtual. When studied across KTAs, the proportion of the three modes of support stays roughly the same.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Table 3. Distribution of training courses by type of strict accessibility.

Supports	nb	%	Accessibility	nb	%
physical	**55**	66%	Physical access required	**64**	77%
Hybrid	**9**	11%			
physical / Virtual	**4**	5%	Virtual access available	**19**	23%
Virtual / Hybrid	**1**	1%			
Virtual	**14**	17%			
Total	**83**		Total	**83**	

Based on this first survey, partners were asked whether strictly physical and strictly hybrid training courses (i.e., 64 training courses) could be developed in virtual access (Fig. 5, B and Appendix A). In total, 32 activities (in yellow Fig. 5, B) could be developed into virtual access. **Eleven training activities** (in orange, Fig. 5, B) **cannot be developed into virtual access.** Sixteen could be partly developed into virtual training, and five could 'possibly' be developed into virtual training (i.e., need further investigations).

These 11 training activities include many manual techniques, or require handling specimens, such as for example the followings: *Analysis using non-destructive techniques; Scientific Illustration; Molecular Systematics Practical course for DNA analyses of scientific collections; Plant collection and mounting; Training in curation of palaeontological collections; Creating 3D digital twins of your collection specimens at a budget: Hands-on training on photogrammetry for beginners; DNA analysis of museum specimens (DNA extraction, DNA quality evaluation and data analysis).*

The vast majority of the trainings' target audience must attend on-site.

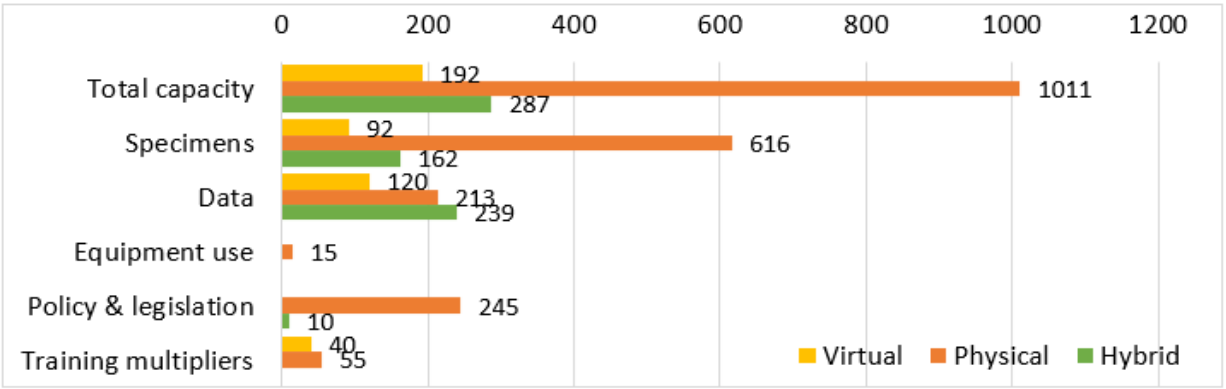


Figure 6. Distribution of support modes (physical, virtual, and hybrid) by KTA with the highest number of participants indicated at the right end of each bar.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Only 14% of the public has access to the training courses virtually (yellow bars in Fig. 6). However, there are disparities in the results depending on the KTAs covered. For "*Specimen*" and "*Equipment use*", there seems to be a clear dominance of physical and hybrid media, which could be explained by the need to handle specimens or equipment. This tendency is reinforced for the "*Policy and legislation*" field, where almost all courses are given strictly in physical form. Only two courses are virtual, but they do not appear in the above graph because they are online videos for which audience capacity cannot be determined. This tendency seems to be reduced for the public who have access to training in the field of "*Data*" and "*Trainings multipliers*", where practices seem to be more on multiple media, with a dominance of training given in hybrid format for "*Data*".

3.1.2.3 Training language

This graph (Fig. 7) refers to the number of times a language is used in a training activity. As some courses are given in several languages, the number of times a language is used does not correspond to the total number of courses. Out of the 83 training activities of the catalogue, 58 can be provided in English, 25 are fully or partially in French, 23 are fully or partially in German, and 13 are fully or partially in Spanish.

This shows that 70% of the training activities are taught in English. Of those, 36 (43%) do provide at least one other language, mostly the one spoken in their country. A total of 25 (30%) training activities are not given at all in English.

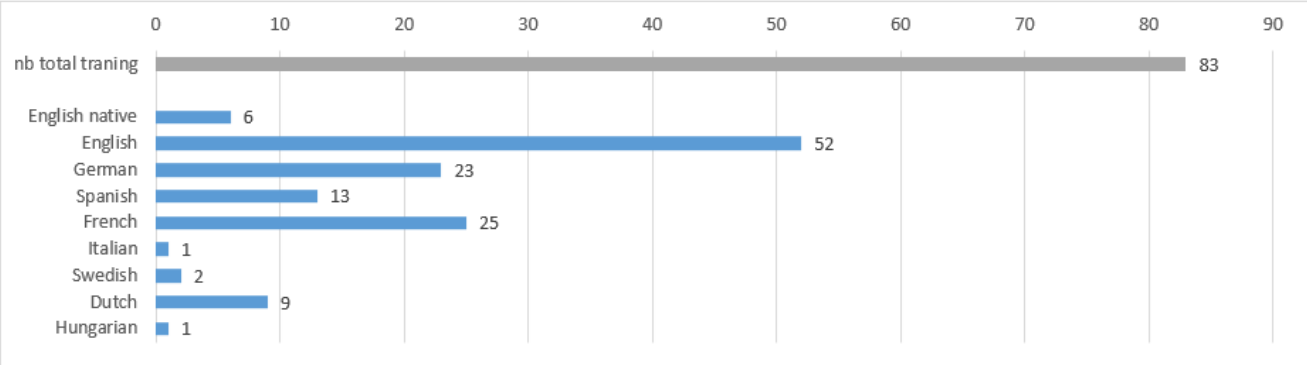


Figure 7. Distribution of the languages used for a total of 83 training activities.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.1.2.4 Training fees and certifications

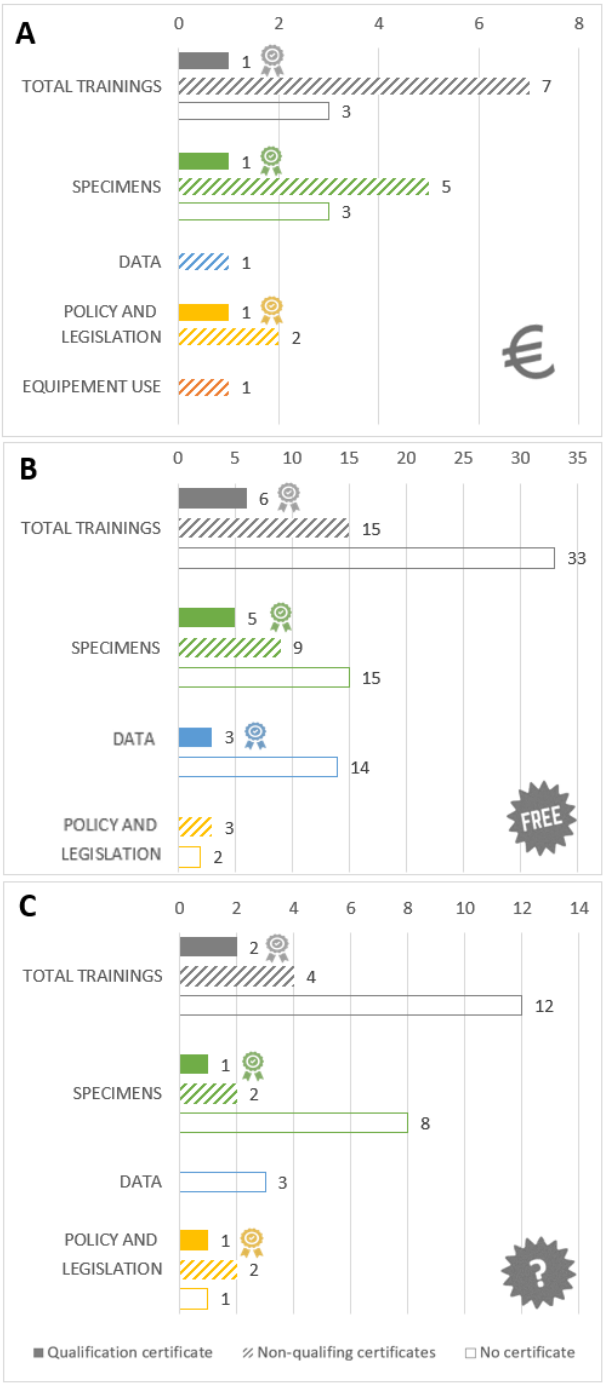


Figure 8. Distribution of certification per KTA: **A)** Training with fees, **B)** training without fees, **C)** Training possibly with fees.

Among the 83 training activities: 11 have fees, 54 are free, and 18 do not provide any information; and 48 do not provide certification, 26 provide a non-qualifying certification, and 9 provide a qualifying certification.

In this section, we presented all the training activities that have fees (Fig. 8, A), all the training activities that are free (Fig. 8, B), and all the training activities for which the feature is not known (Fig. 8, C). Within each of these three categories, we highlighted numbers of training activities offering certification, qualifying or not. Indeed, instinctively one could think that qualifying training courses have fees, and that non-qualifying training courses are free. Among the total training activities with fees (11 courses, Fig. 8, A, *in grey*), only 1 (9%) has a qualifying certification. Among the total training activities that are free (54 courses, Fig. 8, B, *in grey*), 6 (11%) have a qualifying certification. Training activities with fees are not more certifying than free training activities.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.1.2.5 Professional background of trainers and target audience

Professional background of trainers. While education programs in NH institutions are led by two categories of educators: educational staff employed to develop and support visitors, including schools, and scientific staff supervising fellowship programs, often in collaboration with universities, the collected data (Fig. 9) show that most of the training activities involve the scientific staff (mainly scientists, scientific and data managers, curators, collection staff, laboratory technicians).

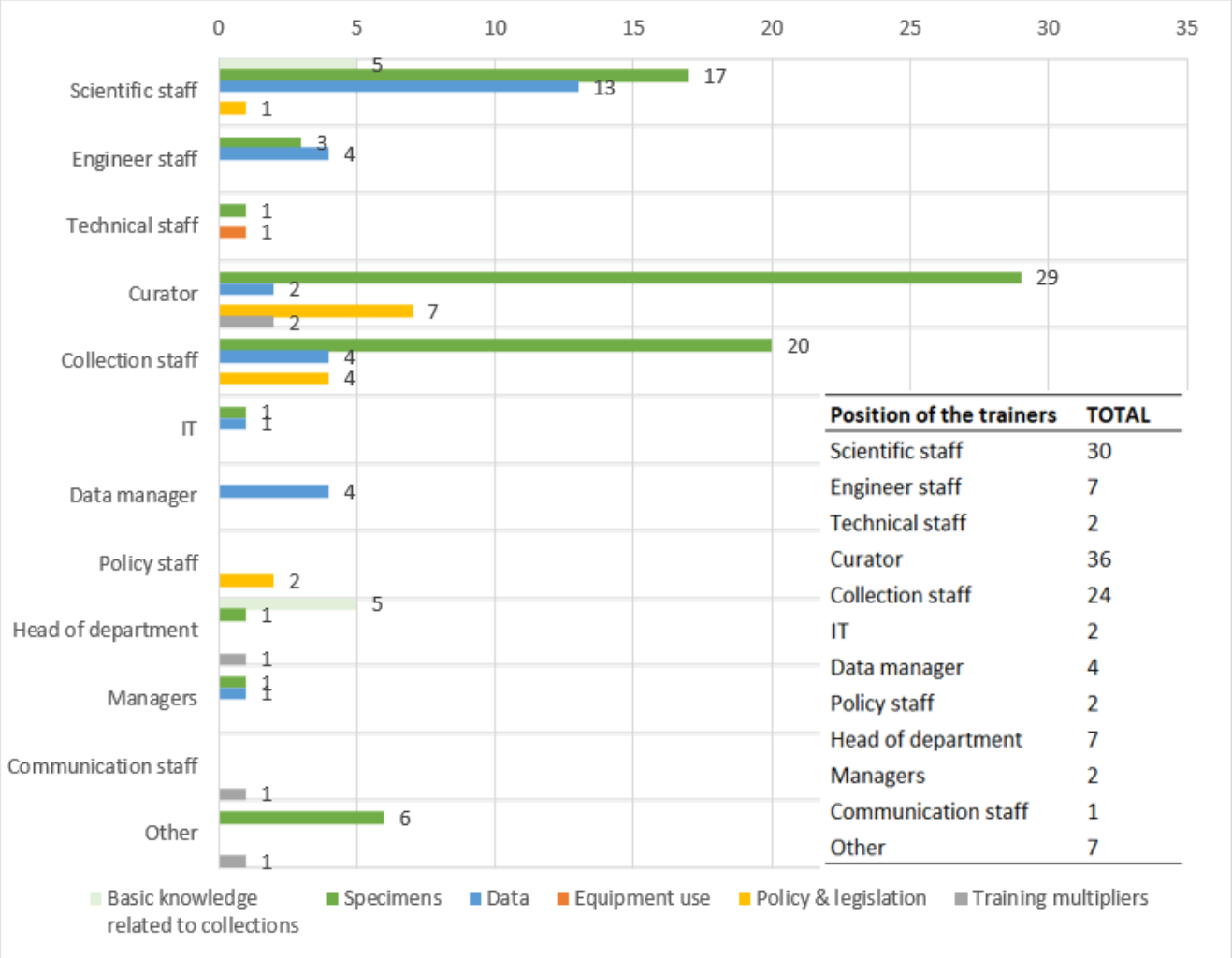


Figure 9. Distribution of the position of trainers per KTA.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

The KTA "*Specimen"* is taught by curators (29 training activities), scientific staff (17), collection staff (20), and others (6) (i.e., "internal involved", "external involved", "input from others", "member of citizen science board NHMW", "Seed Bank partnership coordinators").

The KTA "*Data"* is taught by scientific staff (13), collection staff, engineer staff, data managers, IT and managers (14).

The KTA *"Policy and legislation"* is taught by curators (7), collection staff (4), policy staff (2), and scientists (1).

Target audiences of the training activities. Generally, science-oriented training activities are open to master, PhD students, and early career researchers, and they are slightly longer. Collection- and data-based trainings are open for all including technicians, curators, academics and practical scientists, who want to update their skills.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.2 Analysing the NH training landscape in view of the DiSSCo training objectives

In section 3.1, we have identified and reviewed the existing training activities within SYNTHESYS+ NA2.3 consortium. In this section, we have worked towards the skills that are required to align with DiSSCo objectives, the potential missing topics to be included in the DiSSCo training programme, and the identification of the potential gaps and improvements toward the development of a proactive and efficient DiSSCo training programme.

3.2.1 General gap analysis based on DiSSCo training objectives

The 29 training objectives listed in chapter 2.2.1, representing identified DiSSCo training objectives, were crossed with the 83 training activities of the catalogue in a contingency table, resulting in a total number of 2 407 answers, each coded with YES, NO, PARTLY, or DO NOT KNOW (Table 4). All DiSSCo training objectives are represented in at least one of the 83 training activities. In addition, the survey received feedback from all partners and some trainers (Appendix B), that will be considered and integrated in the recommendations.

Primary results show that, as expected, the current training courses mostly focus on **Science** (84%), and **Data** (63%), less on **Technology** and **Organisation** (25% each), and only a few on the **Financial** (7%) dimension. These last three are potential gaps and need careful examination to decipher how and if they can be addressed/developed in the DiSSCo training program.

In total (i.e., all IRLs together) the contingency table with 83 training activities and 29 training objectives received 69% of NO, 22% of YES, 7% of PARTLY, and 2% of DO NOT KNOW.

This indicates that training activities usually are not multifunctional but specialised in a single or only a few training objectives.

Table 4. Total number and percentage for each IRL match.

Answers received, all IRLs together		
Yes	533	22%
Partly	163	7%
No	1663	69%
Don't know	48	2%
Total nb of answers	**2407**	

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Although PARTLY (7%) answers were not considered in the following analyses, they might be of interest for future investigations. Once DiSSCo's priorities for training to implement its work plan have been resolved, training courses addressing a topic 'PARTLY' may expand that topic to better fit DiSSCo training objectives. If 'PARTLY' answers had been considered as YES, 29% of the current catalogue would already be addressing DiSSCo objectives.

As each IRL contained a different number of training objectives (i.e., S=12, D=6, T=4, O=4, and F=3), the strict number of YES and NO in one IRL cannot be compared to the strict number of YES and NO in another IRL. When considering the proportion (in %) of YES and NO within each IRL taken independently, "*Science"* and "*Data"* were about the same with ca. 30% of YES, and 60-56% of NO*, "Technological"* and "*Organisational"* were also similar with ca. 12% of YES and 80-85% of NO*,* and "*Financial"* had a much lower 4.4% of YES*,* and higher 94% of NO*.*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

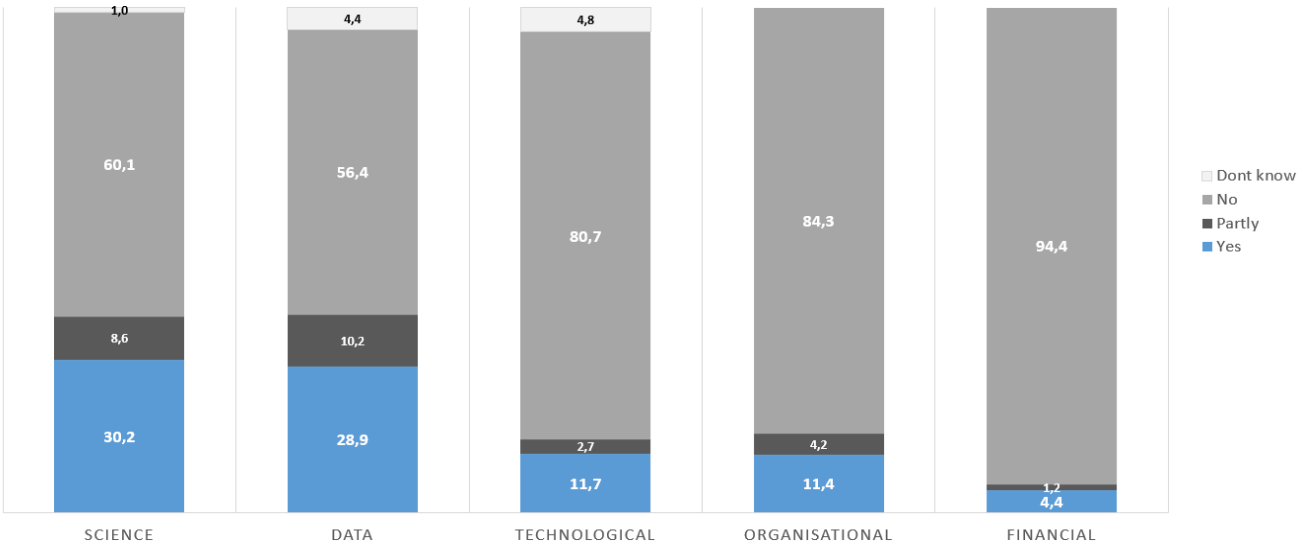


Figure 10. Percentage of responses (YES, PARTLY, DON'T KNOW) obtained for each IRL.

3.2.2 IRL Science

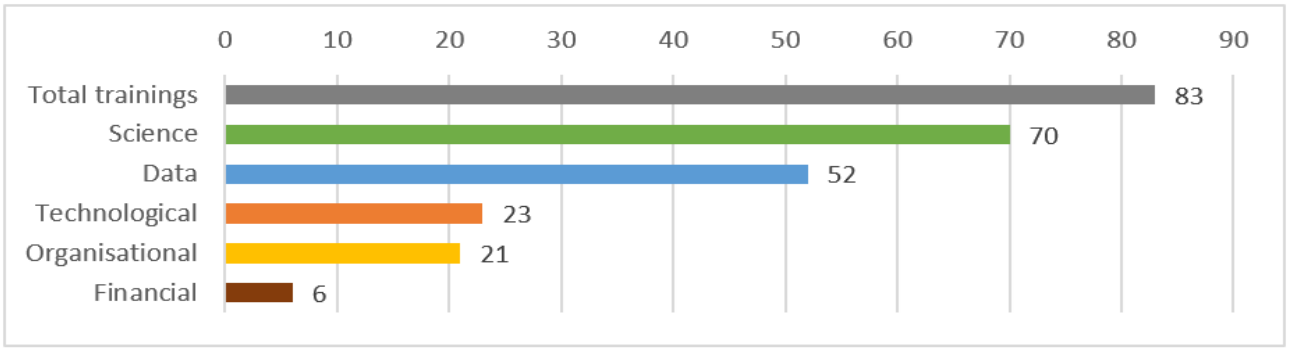


Figure 11. Number of training matching the IRLs.

With 9 and 15-17 dedicated training activities respectively (Fig. 12), the survey suggests that **Geodiversity** and **Citizen Science** are not as well represented as some other categories. This result was also highlighted during the MS26 workshop (i.e., one year before the present IRL based analysis), thus suggesting that these training fields do constitute real gaps - at least in this analysis - and will need careful/in depth investigations through the DiSSCo Prepare Project WP2 network activities.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

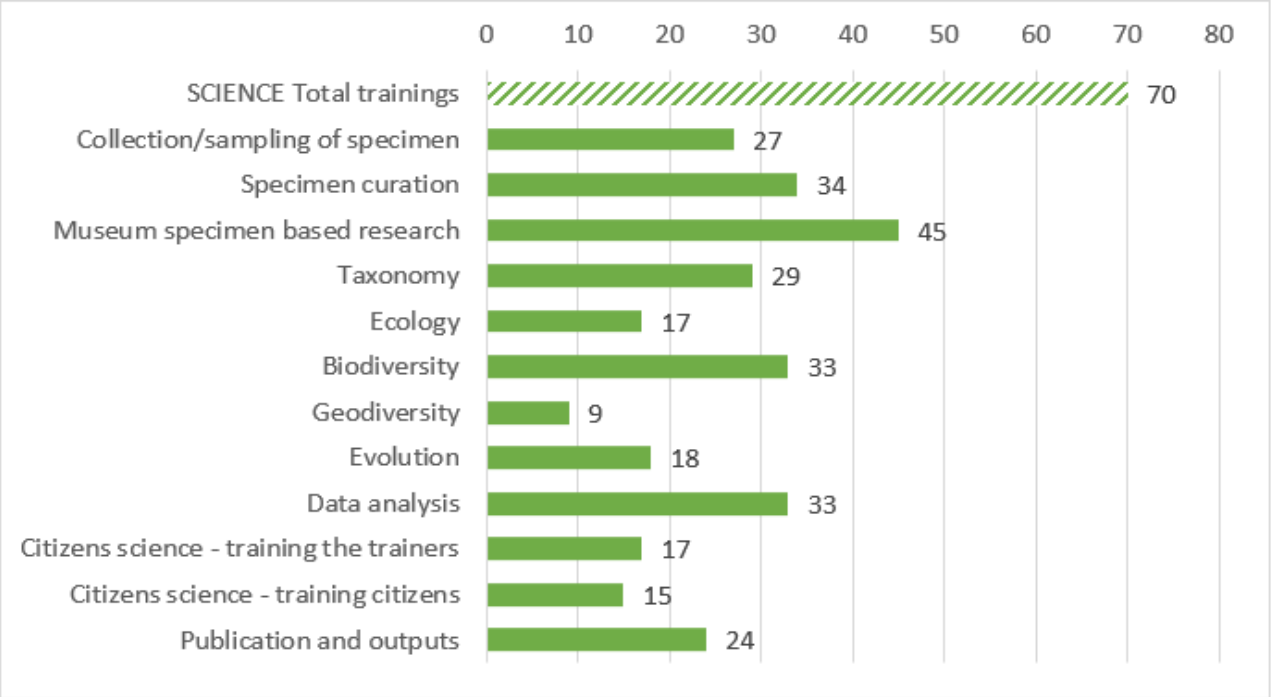


Figure 12. Number of training matching IRL SCIENCE training objectives.

With 17 and 18 dedicated training activities respectively, the highly specialized topics such as Ecology and Evolution, are rather little numbers for NH educational stands. These topics are usually dispensed at University, or during Master and PhD internships, and might not be catalogued here as training activities open to a large EU community. Results here might not indicate real gaps, but rather the fact that these topics are highly specialized, and thus generally taught through internships.

Among the current available training activities, a large number address *Specimen based research* (45), *Specimen curation* (34), *Biodiversity* (33), and *Data analysis* (33), suggesting that these categories are well covered by the current catalogue. Therefore, the question that must be raised is the extent to which these training activities fully address the current DiSSCo training objectives. For the IRL Science, the identified need is to enable more users to embark on data-intensive science research in bio- and geo-diversity.

Under #*Data analysis* some training courses such as: *Integrative taxonomy in the "big data" era*; *Introduction to genomics data; Introduction to Linux and command line; Introduction to the analysis platform GALAXY; Programming with Python for beginners* (MNHN); *Python for biologists;*

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Spatial Analysis in R (UGOT) seems to well address current identified needs. Under *Specimen curation* training activity such as: *Conservation, use and management of DNA and tissue banks* (MNCN-CSIC) might also constitute good training in data-intensive science. However, based on the feedback received from partners (Appendix B) more training activities in big-data management, should be developed.

3.2.3 IRL Data

Across all training objectives together, IRL *Data* and IRL *Science* received a similar proportion of YES (30%) and NO (56%) (Fig. 10). However, if in the IRL *Science* differences were obvious among the coverage of some training objectives (e.g., only 11% for #S-7, geodiversity, versus 55% for #S-3, museum specimen-based research), within the IRL *Data*, every training objective received a more aligned proportion of YES (19%-29%) and NO (55%-61%), thus suggesting a good coverage across the 6 training objectives (Fig. 13). These training objectives are each covered by 19 to 29 training activities, in total 52 out of all 83 training activities cover at least one "*Data*" training objective. Twenty-four of these training activities cover three or more "*Data*" training objectives and can be assumed to have their emphasis on the "*Data*" dimension, while most of the remaining 28 training activities focus on other aspects for which the *Data* dimension is only partially relevant. **Eight training activities** cover all 6 training objectives and therefore can be addressed as comprehensive trainings in the "*Data*" dimension: *Digitisation techniques* (MfN)*; Conservation, use and management of DNA and tissue banks* (MNCN-CSIC)*; Plant Nomenclature; Seed Banking; Use and Curation of Biocultural (ethnobiological) Collections* (RGBK)*; Herbarium training; BID: data mobilisation, data quality, usage of data, project management, Niche modelling, Policy making; Best practices in digitisation, How to set up a digitisation projects (in form of a card game), Usage of GIS for data quality check and errors detection* (GBIF-Meise BG).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

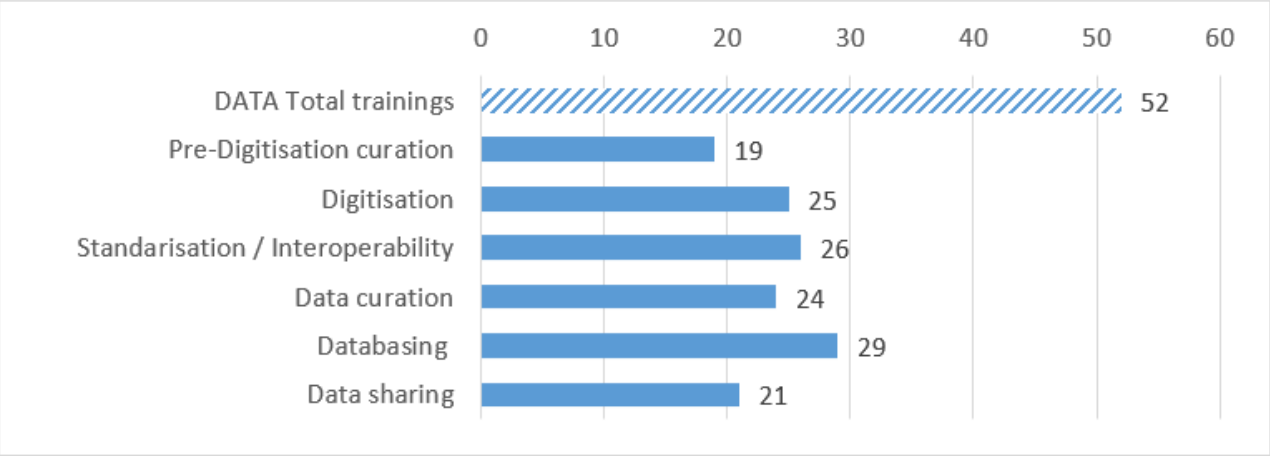


Figure 13. Number of training matching per IRL DATA training objectives.

Within the IRL Data, DiSSCo training objectives are to increase the capacity of the national and institutional nodes to provide and enrich data in a consistent, and harmonised manner, compatible within the overall RI. Specifically, training objectives are intended to improve digital skills and competencies using new mechanisms and tools; to implement best practices for data generation; to improve skills for the enrichment and mobilisation of data at the institutional level; and to develop team working practices.

In the current catalogue of training activities, a handful addresses topics such as: **DNA databasing**; the political/societal/ethical challenges related to the growing **uses of DNA databases;** or **circulation of genomic data at global scale and over time,** and none addresses **biocurator education, an impediment to the expansion of training activities in Data curation, pre-Digitisation curation, Digitisation, and Databasing.**

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.2.4 IRL Technological

All training objectives together, IRL *Technological* received 11.7% YES, 80.7% NO, and 2.7% PARTLY (Fig. 10). The detailed graph (Fig. 14) shows the training objectives *#Operating* and *#IT data management,* relatively well covered with 15 training activities each. This is explained by the various training targeting the new generation of NH data, tools and equipment (*#Operating*), as well as the skills required to manage and transform NH metadata (*#IT data management*). However, if training activities surrounding the manipulation of NH data are well represented, none of the activities of the catalogue target protection, storage, or safety of the data. Reflecting feedback received from IT developers (Appendix B), the training objective *#Development* with only 8 activities could be improved as none of the courses address ELViS or other elements of the DiSSCo data architecture dealing with FAIR digital objects, and data fabrication. The training objective *#e-Monitoring* appears very low with a single activity offered by Meise. This training objective, defined as '*Digital management services (e.g., E-service, helpdesk, mediation services): administrators and users'*, is however targeted in different organisations of DiSSCo (e.g., TDWG, API design, digital platform enabling biodiversity information freely and openly accessible to everyone), and can also be found within scientific communities actively working on web-based taxonomic tools (e.g. World Register of Marine Species; Flanders Marine Institute). Training and skills development in e-monitoring are currently mostly organized through workshops – with dedicated training in the use of the online interface -, and has most probably been overlooked in the current training inventory.

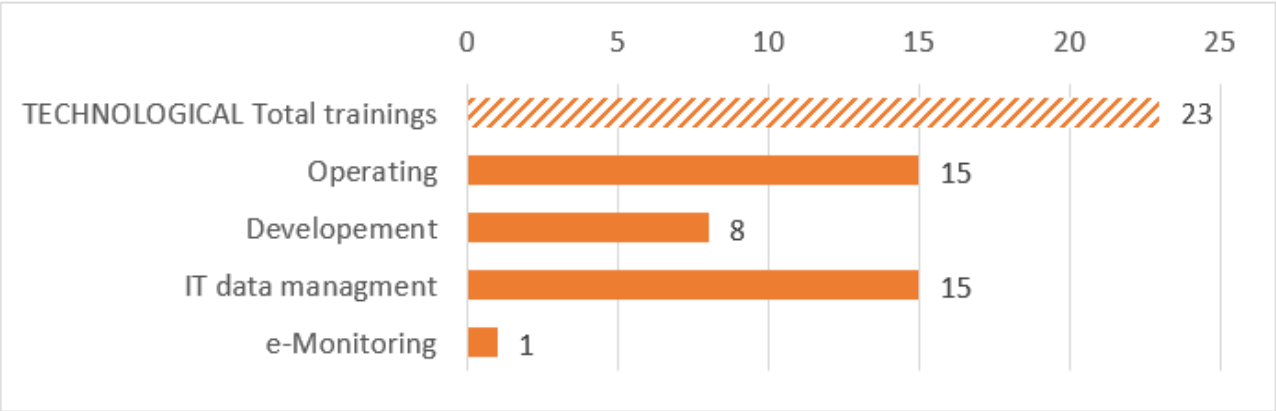


Figure 14. Number of training matching per IRL TECHNOLOGICAL training objectives.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

3.2.5 IRL Organisational

All training objectives together, IRL *Organisational* received 11.4% YES, 84.3% NO, and 4.2% PARTLY (Fig. 10). The 11.4% of YES correspond to 21 training activities addressing one or several of the training objectives *#Policy, #Governance, #Management, #Communication.* Among these 21 activities, 4 address broad topics as they match each of the 5 IRLs and all training objectives of ORGANISATIONAL, and 1 less broad but seems specialised in the field of the IRL *Organisational matching all the training objectives:* ***Due Diligence for utilizing genetic resources in the context of the Nagoya-Protocol*** (Wien). The detailed graph shows that #*Policy* and #*Communication* training objectives are relatively well covered with 15 and 12 training activities, respectively. Inversely, *#Governance, #Management* appears less well covered with 4 and 7 training activities, respectively.

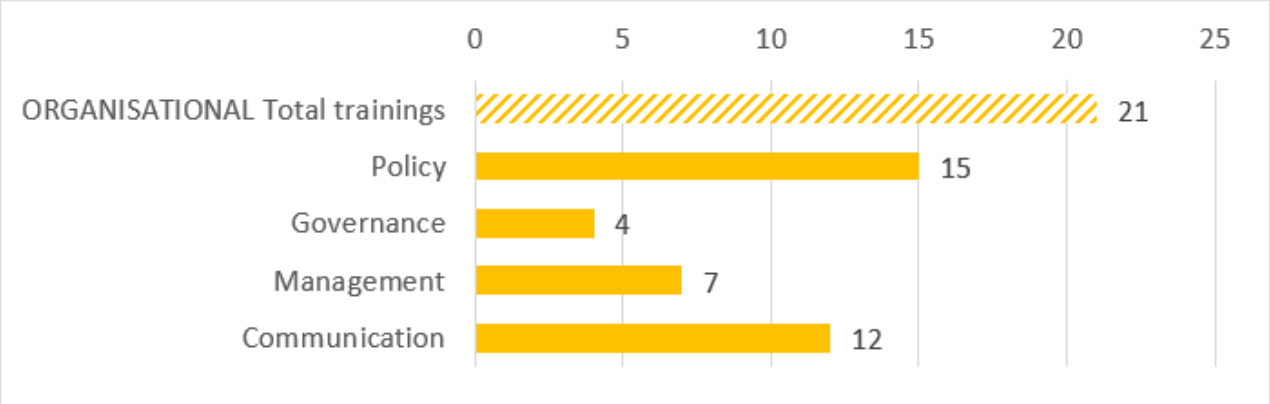


Figure 15. Number of training matching per IRL ORGANISATIONAL training objectives.

When focusing on the 15 training activities addressing #*Policy,* two categories of courses can be observed. A first group, containing 9 training activities, appears highly specialised in governmental policy areas (ABS and the Nagoya Protocol, CITES, Specimen Shipping, Cultural Heritage), while the second group, containing 6 training activities, appear more focused on best practices surrounding NH collection (Botanical collections, Biocultural collection, Digitisation, data management).

If we refer to (Table 5) of national and European policy categories and subjects relating to NH collections collated within the framework of the ICEDIG programme (ICEDIG deliverable 7.1), we can identify several areas that are not covered by the catalogue.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

According to this list of policies, almost all of the courses (14) of the catalogue concern the family "**Collection strategy and management**", and a single course offered by Meise BG seem to deal with "**Data management**", the family "**IT strategy and Management**" and "**Data Strategy and management**" are under-represented.

Table 5. National and European policy categories and subjects - ICEDIG Deliverable 7.1.

Category	No.	Subject
Data Strategy and Management	1	Data & digital media publication
	2	FAIR / Open Data
	3	Freedom of information (FOI)
	4	Intellectual Property Rights (IPR)
	5	Data standards
	6	Personal data
	7	Public Sector Information
IT strategy and management	8	Cloud services & storage
	9	Information risk management
	10	Information security
Collections strategy and management	11	Collections loans management
	12	Protection of sensitive collections data
	13	Collections access & benefit sharing (ABS)
	14	Collections handling
	15	Responsible Research & Innovation (RRI)
	16	Collections scope, storage & care
	17	Digitisation strategy & prioritisation

Communication: Although 12 courses have a correspondence with the training objective #Communication, none of them seems to have communication as their main topic, or as it has been defined under DiSSCo training objectives. This would be communication approached in the services of, for example, collection management, citizen science, etc.). The survey does not make it possible to clearly identify the communication skills that the courses provide.

Management and Governance: 7 courses have a correspondence with the training objective #Management, 4 of which also have a correspondence with #Governance.

As with the results for the #Communication, it appears that none of the courses have management or governance as their main topic. However, the 7 training courses seem to address training themes

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

that are mainly concerned with the management of collection teams (technical and scientific), project management, or assistance with governance and stakeholder engagement (e.g., Nagoya Protocol). The survey does not allow for a clear identification of the management skills that the training courses provide.

3.2.6 IRL Financial

Of the 83 courses offered in the catalogue, only 6 matches with the IRLs Financial (4.4% of YES, 94.4% of No) and none of them is specialised in this field of competences. The training objective #Basic information (5 courses out of 6) provides indications that the courses address some financial aspects, but is not very useful to the analysis. The training objectives #Management and #Financial awareness correspond to the main fields targeted by the survey in the context of the DiSSCo matches as they address highly specialised skills. Only 4 courses address elements of financial project management in the context of the broader themes of Citizen Science (NHMW), digitisation (Part of COST Mobilise training - Meise) and data management (Integrative taxonomy in the "big data" era - Paris, Part of the courses given to GBIF BID - Meise).

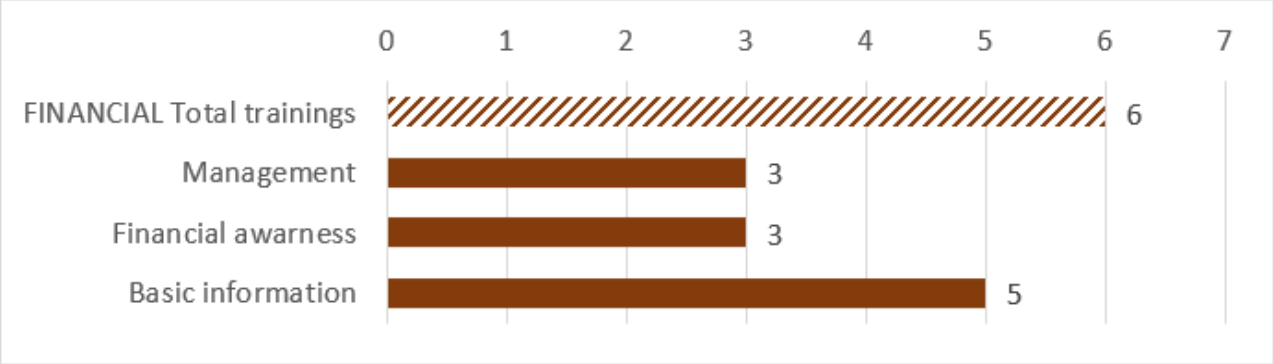


Figure 16. Number of training matching per IRL FINANCIAL training objectives.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4. Discussion and recommendations

In this study (SYNTHESYS+ D2.3), we provide a first catalogue of 83 existing trainings supplied by the SYNTHESYS+ NA2.3 partners, with the related key detailed information (*title, institution, city, target audience, length, in-take capacity, frequency, position of the trainer, language, format, fees and certification*) to be uploaded on the DEST, where applicable. A robust and efficient strategy was to sort the existing training activities by broad educational domains (KTA), and analyse all gathered data in order to characterize the diversity, with strengths and limits, of the current NH training landscape. This thorough analysis enabled identifying - for a sampled list of NH training activities - what are the most represented training areas, what is the training capacity in terms of space, trainer time, and support staff and skill resources, and what are the current practices for training delivery and creation of material.

The benefit outputs of this method ("KTA analysis") are threefold. **First**, once implemented online, notably on the DEST platform, such a searchable and up-to-date catalogue will benefit all types of users looking for a training activity in the field of Natural and Earth sciences, matching their skills, availability, and mobility. **Second**, formatted as it is, the catalogue can be extended by specifically seeking additional training activities in a wider community, including that of DiSSCo RI (i.e., 120+ Institutions, 21 participating countries) and in partnerships with Universities and stakeholders. It can also be updated as training activities evolve. **Third**, through the branching classification by harmonized KTAs, against which every training activity can be cross-matched following their main topics, this method can be implemented as a standardized workflow for NH training landscape assessment. Finally, associated with an in-depth assessment of the current needs of the DiSSCo RI ("*IRL gap analysis*"), the proposed workflow can contribute to the construction of an efficient and proactive 'DiSSCo training programme'. Results presented here show that our analytical pipeline provides a conceptual framework and baseline data to advance investigations on NH training activities.

Recommendation #1: DiSSCo is encouraged to promote the use of the "*KTA analysis*" as a standardized workflow for NH training landscape assessment in the view of DiSSCo training needs. The training inventory must be implemented on the DEST to help users navigating the NH training programme, and extended to a wider community, including that of DiSSCo RI, Universities and stakeholders.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

In the following section, we first discuss results from the "*KTA analysis*" and provide some recommendations on how to better support existing training resources. We specifically present: (1) the main characteristics of the current training landscape; (2) the various criteria (available training material and training practices) required for the development of new training activities; and (3) how the DEST can be used as an efficient resource provider and training dissemination platform.

In a second section, we discuss results of the "*IRL gap analysis*" and provide primary recommendations to the construction of a robust and up-to-date DiSSCo training program. Through the in-depth study of DiSSCo IRLs content and objectives, and based on what we know from the current NH training landscape, we identify potential DiSSCo training gaps in: collection-, policy-, research-, digital-, and dissemination-oriented training activities. We finalize the second section by addressing questions to DiSSCo governance on the boundaries of the DiSSCo training programme.

4.1 Current NH training landscape

4.1.1 Main characteristics of the current training landscape

The Key Training Areas that are the most covered by the catalogue are those targeting "*Specimens", "Data",* and "*Policy & Legislation".* As a consequence, these three KTAs are those for which the number of training days is highest (394 days), and of which the capacity to welcome participants is the largest (1288 people).

In terms of training formats, languages, and qualifying certifications, 77% of the courses are exclusively given on-site, 70% are taught in English, and a very small proportion offer qualifying certification, and have registration fees. No correlation is made between fees and certification, as we observe that paying training activities do not systematically provide a qualifying certification and *vice versa*. There also seems to be no correlation between the charging of fees and the audience of the training. The reasons for charging fees to some trainings are diverse and not well known, and although we may assume that these can be driven by economic factors and/or governmental policies, further investigation is needed to identify the motivations behind the fees applied to some trainings.

As far as the professional background of the trainers is concerned, it appears to be quite diverse among scientific staff. Indeed, although education related to collections in NH institutions can be led by two categories of educators (i.e., educational staff and scientific staff) most of the training activities of the catalogue involve scientific staff, and mainly scientists, curators, and collection managers.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Regarding the target audience of NH training activities, they gather, on the one hand, masters, Ph.D. students, and early career researchers, and on the other hand technicians, curators, engineers, and junior scientists.

Five KTAs were less well represented in the catalogue but nonetheless considered by NA2.3 partners as key areas serving important constituencies. Among them, the following three KTAs "*Equipment use", "Training multipliers"* and "*Text and Media"* are relatively new training topics. They may require new training expertise, new agendas, or update of existing training activities. The two next KTAs "*History of collections"* and "*Basic knowledge related to collection",* were added during the last months of the present analysis. The KTA "*History of collections"* needs further inventory effort, while the KTA "*Basic knowledge related to collection"* needs a more grounded definition.

Recommendation #2: Less well-covered KTAs such as "*Equipment use", "Training multipliers"* "*Text and Media*" and "*History of collections"* need to be investigated more in-depth by DiSSCo Prepare Project partners in order to best explore these new areas specificities and requirements.

Recommendation #3: Further investigations on the motivations from institutions for providing trainings with qualifying certifications and for charging fees are required to set up a coherent and relevant assessment of needs and possibly a business plan.

4.1.2 On the development of new training activities, training material and training practices

For the 83 training activities collated in the catalogue - which represent only a fraction of the existing available NH training activities across Europe - the time dedicated to NH training corresponds to a full year (405 days) and reaches close to 1 400 participants (1 358 people), excluding video views. Except for the few institutions currently developing new training activities (e.g., BG Meise and RBINS), general comments from partners suggest that trainers are already fully committed. Before developing new training activities, and to avoid having overloaded staff unable to fulfil an effective training programme responding to DiSSCo training needs, partners and trainers are requesting beforehand that some support be provided on the following tasks: support or reduce administrative tasks (e.g., reception of requests, booking of training spaces and classrooms, fee administration, e-tool for admin tasks); more space in premises (many European NH institutions

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

are located in historic cities (London, Paris, Wien, Berlin) where space is becoming a problem); funds for more training material (e.g., computers, free access to paying software, free shared space for videos); and service for English translation and dissemination.

Recommendation #4: DiSSCo should foresee a lighter work load for trainers through the development of digital tools simplifying administrative and repetitive tasks.

Recommendation #5: DiSSCo should take into consideration the growing lack of space in many European NH institutions located in historic cities and work towards a strategic plan to improve the physical and working/training conditions.

Recommendation #6: DiSSCo should promote open access to training software and develop financial and sustainable strategies to invest in IT material tailored and adequate for training.

Among the 64 training courses of the catalogue that are not delivered virtually, half (32) could be developed for virtual access depending on the time, funds, and the support the trainers may receive from the community (DiSSCo), while 11 cannot be given virtually. The limits to virtual development can be summarized in three main factors: Firstly, both students and trainers do not want training that is fully online. They want to experience and be physically in the spot to study the collections, they want to meet the professionals to improve and enrich their career. Secondly, many training activities concern the development of practical skills. For instance, learning, PCR techniques, or other manual tasks cannot be done online. Finally, a third of the training activities studied are currently not provided in English, which might be an obstacle for English non-speaking-trainers.

Recommendation #7: DiSSCo should promote on-site trainings which cannot be done virtually (13% of the catalogue), and propose English translation services.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4.1.3 The DEST NH trainings platform - a contribution to DiSSCo

The Distributed European School of Taxonomy (DEST) offers education and training opportunities to students and professionals from any nationality studying, working or interested in the field of taxonomy, biodiversity, geodiversity, and conservation. After its creation under the EDIT Network of Excellence (2006-2011) which was a test case with "Expert in training" and "Modern taxonomy" programmes as well as "Summer schools" training 185 students, the DEST has been managed and supported until 2017 by RBINS offering 65 courses, including training-on-the-job for **361** students. These courses were exclusively created for DEST. Training on-the-job are individual trainings or in very small groups. DEST continued gaining recognition within Europe and beyond with the offer of training possibilities growing. From 2018, the DEST has been "adopted" by the CETAF, the network of the community, in order to guarantee its long-term sustainability. DEST aims to be a platform for EU-based training courses relevant to natural history collections, collection management, taxonomy, and systematics. It can advertise courses from CETAF members and others, as well as support the organisation of courses by experts, CETAF members, and other organisations. DEST has operated at a low level in recent years but is on the course of re-establishing itself as a training platform. At the moment it is mainly a website (https://cetaf.org/dest/) hosting courses by CETAF members and others, still, through close collaboration with the NA2.3 partners following the present analysis of the trainings, the platform will benefit from a guide and structure for the further implementation and management of the trainings and their visibility.

DEST is currently led by Hugo de Boer, from NHM UiO, and managed by Maarten Van Steenberge from RBINS, in collaboration with a DEST committee. This committee consists of the aforementioned director and manager, and representatives of the RBINS (OD Taxonomy and Phylogeny- T. Backeljau, C. Paleco), Naturalis (T. Sijbrandij), and CETAF (A. Casino).

Identified gaps in DEST in relation to Synthesys+ D2.3 and DiSSCo training and capacity-building needs are:

1. Most courses are aimed at data users, and the course portfolio for data suppliers and professionals needs to be developed;
2. DEST needs to expand the website to include online registration and payment;
3. DEST needs to create a streamlined online form for submission of new courses for their announcement on the website;

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4. DEST does not aim to develop its own online teaching platform, but could adopt existing or budding solutions;
5. DEST operates on 0.25 FTE (full-time equivalent), and needs to have at least a full-time project manager to operate successfully in the current landscape of activity.

Recommendation #8: A collaboration between DiSSCo and the DEST must be established in order to develop the DiSSCo training strategy based on the current practices and KTAs needs identified but also to support the DEST as a resource provider and dissemination platform.

This "*KTA analysis*" showed that significant effort is being made by NH institutions to provide education and training in all fields of their activities (e.g., collection, curation, research, dissemination, expertise, and policy) and to align with science progress. Many strengths were identified in the course of our work, but also several limits. In the next section, we analyse these strengths and limits in the light of DiSSCo training objectives and provide recommendations for the development of a DiSSCo training programme and strategy.

4.2. Preliminary assessment for the DiSSCo training programme

The landscape of NH collections is changing next to physical objects, almost all collections now have potentially limitless digital data linked to physical specimens creating an "extended specimen" (Webster 2017) - the adaptation to this change is at the heart of DiSSCo training objectives. This linkage requires the creation of new infrastructures (servers, intra-, and inter-institutional platforms, e-tools, and e-services), new workflows, new institutional rules adapting new concepts for digital storage, standardisation, persistent identifiers, collection and data policies to define the access and use to these "next-generation" collections. In addition, digitisation on demand is becoming more widespread and needs adjustments from the collection management teams through prioritization strategies focusing on the digitisation of specimens or collections that are of greater interest to researchers or communities.

However, current impediments to further progress can be non-exhaustively characterised as follows: Lack of personnel with proper qualifications, underfunding and space limitations, and poor scientific recognition accorded to database developers and managers. Bottlenecks exist, such as the

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

huge amount of data collection still to be databased or still to be recorded (from handwritten labels and catalogues to interoperable digital data). The knowledge to curate NH collections is not evenly distributed among NH institutions, and may not cover all types of taxa/collections. Moreover, databases are still not always publicly available and accessible for research.

4.2.1 Collection-oriented trainings

The "IRL g*ap analysis*" suggested that under the new DiSSCo RI objectives, NH collections must be in full and constant alignment with the ever-changing needs of its users. Training objectives must take account of the accelerated transformation of NH data, technology, and access, as well as prepare users to adapt to the current socio-/climato- economic climate.

Training strengths. The present catalogue offers many collection-oriented training activities, involving all the main categories of related personnel (curators, researchers and collection staff), and providing a wide variety of approaches and methods for collection management: *from accepting, accessioning and handling a specimen/object to sharing its metadata through public e-portals.* Training in pre-digitisation curation and metadata capture are generously available in the catalogue, thus fulfilling already some of the DiSSCo training objectives. In the catalogue, are included as well various training to convert disparate datasets into a standardized data format. Indeed, regarding the IRL "*Data"*, of which training objectives are to provide and enrich data in a consistent and harmonized manner, two thirds of the catalogue cover that aspect, with nine courses having comprehensive content in this respect. This represents a solid foundation for remedying harmonized access to collection data. In addition, some training activities go beyond DiSSCo training objectives by addressing current areas of concern such as disaster preparedness, integrated pest management, collection movement, or handling of old collections is of importance and valuable for junior curators and collection staff.

Based on the "*KTA analysis*", our results show that the training material is more and more available in digital form, with two thirds of all training activities being available in English, both these elements are increasing the availability and capacity of the trainings for potential attendees. These training activities, if combined in one platform like e.g., the DEST, could counteract the current unevenly distributed knowledge on NH collections curation.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #9: Specimen curation courses should be disseminated among curators and collection staff through social media, virtual meetings, and physical visits. This could be organised once per year, but need human resources to coordinate effectively.

Recommendation #10: DiSSCo must promote increased exchange among institutions to share experience in care and maintenance, collection management, and preventive conservation.

Limiting factors & gaps. While the IRL "*Data"* is covered by a wide variety of training activities, the number of training courses dedicated only to this objective might be insufficient for the present needs. Indeed, only a small number of trainings are dealing with "data quality", "imaging", "digitisation techniques", "databasing" and "data curation", while few trainings are dedicated to "field sampling" and "collections for exhibition". Missing topics such as "field sampling" and "exhibition" can be explained by the fact that the former is often part of the university or summer school curricula, while the later can be attributed to a certain incapacity or impossibility to continuously update users on collection contents. Missing training for exhibition could also be attributed to the fact that this task often is performed by specialised workforce (e.g., taxidermists, who have not been interviewed in the present analysis. etc).

There are delays in mobilising NH collection data though widely accessible sites such as GBIF and the majority of NH collections are undigitised (Paton et al. 2020; Hardisty et al. 2020). Gathering basic metadata will help identify priorities for digitisation to serve research and Hardisty et al. (2020) suggest that DiSSCo institutions should consider quickly creating basic level inventories to facilitate this. Training will be needed to ensure NH institutions can effectively use the developing *Minimum Information about a Digital Specimen* (MIDS) standard (Hardisty et al. 2020).

Recommendation #11: DiSSCo should support a specific training in implementation of MIDS and promote training in CSAT (Collections Self-Assessment Tool), the online tool developed by Synthesys III to assess the management of their virtual and new physical collections.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #12: After the collection description part of ELViS is completed, the training should be put in place to provide guidance to ELViS users on describing their collections, implementation of MIDS/MICS minimum information standards.

Partners are highlighting the need for training in DNA collection databasing and circulation of genomic data, also noting the need to address the political/societal/ethical challenges related to the growing use of DNA data. Institutions frequently do not have curators focusing in this relatively new and rapidly developing area, citing the greater need for 'biocuration' (Howe et al. 2008). This topic must also be linked with policy-oriented trainings (e.g., Nagoya and Digitalised Sequences Information, DSI).

Recommendation #13: Training in curation of DNA collection and tissue samples and relevant data standards for circulation and use of genomic data should be developed perhaps in conjunction with Global Genome Biodiversity Network (GGBN), and/or following trainings and training materials on various software for collection management provided by e-Recolnat experience.

Recommendation #14: DiSSCo must promote scientific recognition to database managers through inclusion in scientific projects and more citation and acknowledgement in research papers.

Recommendation #15: Growing databases and computerisation of collections require training the staff in fundamentals of informatics, data management planning, preparedness of collection to computer disaster.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4.2.2 Policy-oriented trainings

Under the strategies put in place for the harmonisation of policies across DiSSCo RI facilities, DiSSCo training objectives must cover all areas of legislative practices linked to both internal (e.g., collection strategy, data management, or IT strategy) and external policies (e.g., Convention on Biological Diversity, Nagoya Protocol on ABS, or any Global activity contributing to data alignment initiatives for bio-geo collections).

Training strengths. Concerning the strategy and management of collections, two topics, ABS and CITES, are currently well covered by NH institutions and accessible in multiple languages (En, Fr, Ge, Span). Concerning data management policy, courses from the catalogue are mainly provided by GBIF. Besides these highly specialised courses, many other courses with a broader topic address curation and management legislation (e.g., plant health legislation).

Limiting factors & gaps. Six areas need further training improvement and development to align with DiSSCo objectives. Indeed, the actual catalogue lacks in-depth training on the various key policy areas: (1) the implementation of the laws at institutional level (institutional policy) are practically non-existent in almost all policy areas affecting DiSSCo (ICEDIG Deliverable 7.1). The major gap detected is that there is no training in the catalogue that addresses differences in national legislation and different implementation of the European legislation, this is critical, especially cultural heritage and conservation landscape, flora and fauna legislation because it affects permits to collect or excavate in the field, loans between institutions, and new incomes. It could also affect the harmonisation of policies across DiSSCo institutions; (2) cultural goods protection legislation (cultural heritage) that is poorly included; (3) General Data Protections Regulation (GDPR on personal data management, regarding data of people access collections) that usually is only mentioned in some courses; (4) Intellectual Property Rights (IPR), mainly regarding 2D and 3D digitisation), also only mentioned; (5) Legislation on Animal welfare, health and safety at work that is also necessary to avoid risks when handling specimens in labs and in the field, and while working in collections' stores; and (6) Legislation on exportation and importation of biological material (fresh, freeze, and fluid preserved), also required for a correct shipping, related with collections loan management.

In the Post-Covid era, it could be a problem that currently most of the training is face-to-face, although almost all could be adapted to virtual.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #16: A thorough investigation into the curricula of each specific legislation course, to compare and detect if there are differences in timing, contents and approach is required to more clearly identify gaps and repetition.

Recommendation #17: National legislations could affect harmonisation of policies across institutions, e.g., national legislation on landscape conservation and cultural heritage that affect policies on field collecting, loans and incomes at national level. DiSSCo consortium should see the mentioned gap as an opportunity to analyse these differences, synthesize them and organize their own course, with trainers from all over Europe. The introduction to these analyses could be included in a Workshop. A working group should prepare some tasks to be developed in the framework of a next DiSSCo project.

Recommendation #18: In the context of DiSSCo RI's objectives of access to data and large-scale digitisation, it would be necessary to develop trainings fulfilling ICEDIG D 7.1 areas in the "*Data strategy and management*", especially on Intellectual Property Rights, data standards, and General Data Protections Regulation subjects, as well as in and "*IT strategy and management*" (cloud services and storage, information risk management and information security).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4.2.3 Research-oriented trainings

The "IRL Gap analysis" suggested that under the new DiSSCo RI objectives, NH institutions must enable identified users to embark on data-intensive research.

Training strengths. Generally, there is good coverage of basic skills to manage the data related to NH collections or basic digitisation of the collections to support research activities. For example, several courses deal with taxonomy, ecology, conservation, and nomenclature.

Limiting factors & gaps. The gaps in training largely reflect the need to adopt new technologies to maximise the impact of NH collection data in research on global challenges. New approaches to using NH collections have been enabled by increased digitisation, new genomic approaches to their study and the increase in computer power, data standards and protocols (Paton et al. 2020; Baker et al. 2020). Creation of 'extended' or 'digital' specimens, linking collection data to other classes of data such as genomic and environmental data (Webster 2017; Lendemer et al. 2020; Hardisty et al. 2020) has made the auditable information of the traditional NH specimen, representing the occurrence of an organism in space and time, relevant and usable by a broad range of research (Hebberling et al 2019; 2021). Several courses covering digitisation exist and research methodologies and tools for analysis of data are broadly covered in academic degree courses. DiSSCo and NH institutions have a vital role in supporting the infrastructure to service this research. This infrastructure depends on data standards and protocols being developed adhered to by data providers and data aggregators. Given the importance to the use of NH collections and linkage to other data in addressing global challenges, training in biocuration and use and provision of collection data using FAIR principals is vital to DiSSCo's aims.

Recommendation #19: Training to support the implementation of the extended or digital specimen through appropriate data management and curation using community standards should be developed. Training should also cover methodologies to allow the production of FAIR data, promoting the use of NH collection data as 'big data' to address today's societal challenges.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #20: High efficiency skills needed to adapt to fast evolving technological development (e.g., -omics, photogrammetry, 3D laser scanning, and computed tomography...) and growing demand.

Recommendation #21: DiSSCo must sustain permanent dialogue with its trainers to understand scientific and technological challenges in data intensive domains.

The "*IRL gap analysis*" highlighted a rather low proportion of training activities in Geodiversity. This could be due to sampling biases in our inventory. Among the 12 participating Institutions, some are Botanical Gardens, and do not have an Earth Science Department. This area is also a strength of University departments and Geological infrastructures and is probably better covered by them. This trend could also be due to less need to consult actual specimens. Another aspect to consider is the relatively recent use of the term 'geodiversity'. This term was first used in 1993 (Wiedenbein, 1993; see Gray, 2008) after the international adoption of the CBD in 1992, and gained equal weight to biodiversity conservation only recently. Geodiversity is now recognized as essential to species and habitat protection. In alignment with H2020 research objectives, geodiversity education in the NH training landscape needs to be put at the same level of importance as biodiversity and receive proper educational focus. Protection of geodiversity and geological sites, policy development and best practices in soil and landscape conservation has not been inventoried in depth in the present survey. Educational activities might be found or developed in collaboration with the Conservation of Nature (IUCN) that has completed a best practice guideline for geo-conservation and management in protected areas (Tormey, 2019).

Recommendation #22: DiSSCo must place Geodiversity education at the same level of importance as Biodiversity education and promote development of proper educational focus in policy development and best practices to Earth Science.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4.2.4 Digital-oriented trainings

Within the construction phase of the DiSSCo RI, training objectives must accompany the RI in its technological implementations of existing e-infrastructures and the establishment of its data architecture as well as the technical specifications of its future services. Digital training objectives are numerous and broad, from reinforcing basic software skills and equipment use, supporting competencies in protection, storage, and safety of NH and personal data, to developing training in data architecture and data carpentry.

Strengths. Modern research in the fields of NH science and conservation is unthinkable without the use of modern technologies (e.g., 2D/3D technology with advanced equipment allowing digital imaging, X-ray computed tomography, remote sensing, tracking and monitoring tools, computational resources, and next generation sequencing). If we consult Facilities on ELViS platform, we see that almost all European institutions (of the SYNTHESYS consortium) own many advanced and modern labs, digitizing equipment and software to manipulate the data obtained. Often there are also manuals and short training sessions that could be provided to visitors who wish to use the equipment. Institutions engaged in mass digitisation and digitisation-on-demand have the appropriate expertise, equipment and experience in using external resources and are ready to share them. NH institutions also have experience in creating shared facilities, such as JEMU (Joint Experimental Molecular Unit) or joint digitisation team of RBINS and RMCA in Belgium. They are usually trained in photography, scanning, or genetic techniques and are perfectly placed to deliver high quality training.

Limiting factors & gaps. Despite all these advantages, techniques and use of equipment are perceived as "difficult" as they require new skills and full focus. In order not to reinvent the wheel, it is important to exchange experience between more advanced users and less experienced learners. Some SYNTHESYS partners are further in digitizing their collections and can provide valuable support to those just getting started in choosing the right technique and setting up efficient workflows.

In the ICEDIG Deliverable 8.1, Recommendation 79 was already formulated as "*DiSSCo should organise a training curriculum for its member institutions covering: i) technological aspects, such as features and operation of equipment and software; …iii) efficient digitisation workflows in various situations, including quality management…"*. Combining it with Recommendation 21 ("*DiSSCo should promote re-use and/or cost-sharing of digitisation equipment across institutions and projects where possible, particularly for smaller collections."*), we would make the following recommendations:

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #23: Organise a joint training curriculum on digitisation workflows for different types of collections (cf. ICEDIG D8.1, Recomm. #79).

Recommendation #24: Create a catalogue on equipment and existing expertise so that users can search by different categories (e.g., digitisation of alcohol/dry/thin section specimen, non-destructive analysis of specimen or bioacoustics technology, etc.) and find available manuals, hard or software recommendations, trainings, videos, best practices, etc.

Recommendation #25: Make a list of institutional and out-sourced facilities, with advantages and disadvantages, financial information etc., in order to help others, often smaller institutions, to overcome a threshold (cf. ICEDIG D8.1, Recomm. #21).

4.2.5 Dissemination-oriented trainings

NH collections have been enriched through time by expert amateurs (who are more numerous and administratively less overload than researchers), and this trend is expanding with growing access to the internet (e.g., INPN Espèces, Herbonautes), improvements of well-being and life expectancy, and increased interest of citizens to NH sciences. Contribution of these expert amateurs - at different stages of NH data processing, from the collect, to the sorting, mounting, and pre-digitisation preparation of vouchers - has made it possible to create large datasets.

Training strengths. Three institutions reported to have train-the-trainer activities for citizen scientists and teachers to make from them multipliers. It is likely that other museums also have specific trainings for teachers, implemented on a regular basis, which are not specifically identified as a training course and therefore do not show up in this analysis. Furthermore, a museum is an educational tool *per se* and therefore the dissemination of knowledge is its core activity. As these multipliers are often emotionally attached to the specific institutions, the local language is of great importance. The eagerness to contribute to the construction of historical collections, and the recognition received from NH institutions within an EU-wide network both constitute major *stimuli*.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Limiting factors & gaps. There is a small proportion of trainings in Citizen Science (grouped under the KTA "*Training multipliers*"). This could be due to several factors: (1) it is a relatively recent field of research and practices; (2) it is time consuming with dedicated time for educational/communication phases as well as data validation phases; (3) it requires the development of web applications and other digital tools; and (4) hands-on training activities are often given in the field or in the lab (at small spatial scale) in the native language and in many cases happen informally rather than as registered courses, excluding larger audiences from the EU community.

Recommendation #26: Building awareness of the implemented work with volunteers, which has frequently been developed over a long period in NH institutions.

Recommendation #27: Offer training for volunteers on a regular basis.

Recommendation #28: Develop training activities for citizen scientists in different fields of research.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

4.3 DiSSCo training needs beyond the NH landscape boundaries

What should be included in a future DiSSCo training programme?

The systematic analysis of the content and goals of each IRL led to the fruitful aggregation of potential DiSSCo training objectives (cf. Section 2.2.1). Within the IRLs "*Science*", "*Data*", and parts of "*Technological*" and "*Organisational*", training objectives corresponded to regular, necessary improvements of a current NH training programme. The work thus consisted of assessing to what extent existing activities could be improved to fully meet DiSSCo training needs.

Within the IRLs "*Technological*", "*Organisational*" and "*Financial*", some training objectives corresponded to highly specialized tasks standing beyond NH expertise. Questions were thus raised as to whether these objectives should or not be included in the NH training programme. For instance, in the IRL **Technological**, the training objective "*Development*" target the conceptual design and implementation of software and digital tool such as ELViS (e.g., data architecture, data carpentry), and the training objective "*IT data management*" unite all activities around the manipulation, protection, storage, and safety of data. These tasks are so specialised that NH Institutions usually refer to, or collaborate with, private sector and external providers. Similarly, in the IRL **Organisational**, besides the training objective "*Policy*", three additional training objectives were defined: the training objective "*Governance"* that refers to the skills required for legal and inter-community decision making (e.g., decision making at governance levels, stakeholders' engagement); "*Management*" that refers to defining strategies and management plans, administering and coordinating projects & activities; and "*Communication*" that corresponds to the expertise related to institutional, international, public, media, partnership, and internal communication. Although these three last fields are not the traditional core business of NH institutions, the national, European and international policy context has led them to develop the necessary expertise, particularly in the application of regulations and in the political integration of research programmes and infrastructures. However, it appears that this expertise and competence is not translated into training activity within NH institutions. Complementary to the IRL **Organisational**, the competencies targeted in the IRL **Financial**, aim to respond to the needs for a robust financial framework for calculating costs and contributions in order to guarantee financial commitments at the national and European levels. Across the European NH community, there appears to be a need for strategic financial and budgetary management skills. Such skills may help build and deliver research/collection/technological projects and secure financial commitments at the national and European levels. Overall, the training

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

catalogue has very limited relevance to IRL **Financial**. This field of competence seems, for the moment, relatively remote from the training landscape of NH institutions.

Recommendation #29: In order to meet the communication needs of the scientific and institutional community, it seems important to broaden the survey within the DiSSCo institutions in order to identify specialised communication training on several levels (scientific communication, institutional communication, media/press/social network communication support, etc.).

Recommendation #30: In order to ensure the efficient running of DiSSCo RI and to support institutions in their involvement in European programmes, it also seems relevant to develop training for the integration of stakeholders in European programmes, in particular DiSSCo.

Recommendation #31: An added value of the services provided by the DiSSCo community (by its members or by the RI directly) could be the offering of training activities for the management of calls for projects intended for the scientific community.

Recommendation #32: As the Financial field of expertise is traditionally far from the NH training environment, it would be useful to deepen the survey specifically targeting the IRL Financial to get a better overview of what currently exists and what are the specific needs.

Recommendation #33: Discussions must be established among the various actors of DiSSCo in order to delimit the level of specificity that must be reached, in terms of training objectives, within the Financial, Organizational and some of the Technological dimensions.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

5. List of recommendations

Recommendation #1: DiSSCo is encouraged to promote the use of the "*KTA analysis*" as a standardized workflow for NH training landscape assessment in the view of DiSSCo training needs. The training inventory must be implemented on the DEST to help users navigating the NH training programme, and extended to a wider community, including that of DiSSCo RI, Universities and stakeholders.

Recommendation #2: Less well-covered KTAs such as "*Equipment use*", "*Training multipliers*" "*Text and Media*" and "*History of collections*" need to be investigated more in-depth by DiSSCo Prepare Project partners in order to best explore these new areas specificities and requirements.

Recommendation #3: Further investigations on the motivations from institutions for providing trainings with qualifying certifications and for charging fees are required to set up a coherent and relevant assessment of needs and possibly a business plan.

Recommendation #4: DiSSCo should foresee a lighter work load for trainers through the development of digital tools simplifying administrative and repetitive tasks.

Recommendation #5: DiSSCo should take into consideration the growing lack of space in many European NH institutions located in historic cities and work towards a strategic plan to improve the physical and working/training conditions.

Recommendation #6: DiSSCo should promote open access to training software and develop financial and sustainable strategies to invest in IT material tailored and adequate for training.

Recommendation #7: DiSSCo should promote on-site trainings which cannot be done virtually (13% of the catalogue), and propose English translation services.

Recommendation #8: A collaboration between DiSSCo and the DEST must be established in order to develop the DiSSCo training strategy based on the current practices and KTAs needs identified but also to support the DEST as a resource provider and dissemination platform.

Recommendation #9: Specimen curation courses should be disseminated among curators and collection staff through social media, virtual meetings, and physical visits. This could be organised once per year, but need human resources to coordinate effectively.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #10: DiSSCo must promote increased exchange among institutions to share experience in care and maintenance, collection management, and preventive conservation.

Recommendation #11: DiSSCo should support a specific training in implementation of MIDS and promote training in CSAT (Collections Self-Assessment Tool), the online tool developed by Synthesys III to assess the management of their virtual and new physical collections.

Recommendation #12: After the collection description part of ELViS is completed, the training should be put in place to provide guidance to ELViS users on describing their collections, implementation of MIDS/MICS minimum information standards.

Recommendation #13: Training in curation of DNA collection and tissue samples and relevant data standards for circulation and use of genomic data should be developed perhaps in conjunction with Global Genome Biodiversity Network (GGBN), and/or following trainings and training materials on various software for collection management provided by e-Recolnat experience.

Recommendation #14: DiSSCo must promote scientific recognition to database managers through inclusion in scientific projects and more citation and acknowledgement in research papers.

Recommendation #15: Growing databases and computerisation of collections require training the staff in fundamentals of informatics, data management planning, preparedness of collection to computer disaster.

Recommendation #16: A thorough investigation into the curricula of each specific legislation course, to compare and detect if there are differences in timing, contents and approach is required to more clearly identify gaps and repetition.

Recommendation #17: National legislations could affect harmonisation of policies across institutions, e.g., national legislation on landscape conservation and cultural heritage that affect policies on field collecting, loans and incomes at national level. DiSSCo consortium should see the mentioned gap as an opportunity to analyse these differences, synthesize them and organize their own course, with trainers from all over Europe. The introduction to these analyses could be included in a Workshop. A working group should prepare some tasks to be developed in the framework of a next DiSSCo project.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #18: In the context of DiSSCo RI's objectives of access to data and large-scale digitisation, it would be necessary to develop trainings fulfilling ICEDIG D 7.1 areas in the "*Data strategy and management*", especially on Intellectual Property Rights, data standards, and General Data Protections Regulation subjects, as well as in and "*IT strategy and management*" (cloud services and storage, information risk management and information security).

Recommendation #19: Training to support the implementation of the extended or digital specimen through appropriate data management and curation using community standards should be developed. Training should also cover methodologies to allow the production of FAIR data, promoting the use of NH collection data as 'big data' to address today's societal challenges.

Recommendation #20: High efficiency skills needed to adapt to fast evolving technological development (e.g., -omics, photogrammetry, 3D laser scanning, and computed tomography...) and growing demand.

Recommendation #21: DiSSCo must sustain permanent dialogue with its trainers to understand scientific and technological challenges in data intensive domains.

Recommendation #22: DiSSCo must place Geodiversity education at the same level of importance as Biodiversity education and promote development of proper educational focus in policy development and best practices to Earth Science.

Recommendation #23: Organise a joint training curriculum on digitisation workflows for different types of collections (cf. ICEDIG D8.1, Recomm. #79).

Recommendation #24: Create a catalogue on equipment and existing expertise so that users can search by different categories (e.g., digitisation of alcohol/dry/thin section specimen, non-destructive analysis of specimen or bioacoustics technology, etc.) and find available manuals, hard or software recommendations, trainings, videos, best practices, etc.

Recommendation #25: Make a list of institutional and out-sourced facilities, with advantages and disadvantages, financial information etc., in order to help others, often smaller institutions, to overcome a threshold (cf. ICEDIG D8.1, Recomm. #21).

Recommendation #26: Building awareness of the implemented work with volunteers, which has frequently been developed over a long period in NH institutions.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Recommendation #27: Offer training for volunteers on a regular basis.

Recommendation #28: Develop training activities for citizen scientists in different fields of research.

Recommendation #29: In order to meet the communication needs of the scientific and institutional community, it seems important to broaden the survey within the DiSSCo institutions in order to identify specialised communication training on several levels (scientific communication, institutional communication, media/press/social network communication support, etc.).

Recommendation #30: In order to ensure the efficient running of DiSSCo RI and to support institutions in their involvement in European programmes, it also seems relevant to develop training for the integration of stakeholders in European programmes, in particular DiSSCo.

Recommendation #31: An added value of the services provided by the DiSSCo community (by its members or by the RI directly) could be the offering of training activities for the management of calls for projects intended for the scientific community.

Recommendation #32: As the Financial field of expertise is traditionally far from the NH training environment, it would be useful to deepen the survey specifically targeting the IRL Financial to get a better overview of what currently exists and what are the specific needs.

Recommendation #33: Discussions must be established among the various actors of DiSSCo in order to delimit the level of specificity that must be reached, in terms of training objectives, within the Financial, Organizational and some of the Technological dimensions.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

6. Conclusion and perspectives

After a thorough and detailed work taking as basis the various training activities provided by the 12 partners, analysing their major features, and comparing them towards the DiSSCo needs and requirements, a robust methodology has been developed. Such a novel and very targeted methodology provides a baseline for further developments with a subset of pivotal trainings but more importantly sets up the process for future training activities to be inventoried, analysed and described. The definitions and categories proposed have helped to identify the various components and topics of the training activities offered by NH museums in relation to their hosted collections enabling the NH institutions involved to have a common understanding on what the scope, reach and impact of their trainings are actually covered.

This first set of detailed information gives the backbone of the state-of-the-art of the training offer proposed by the NH community. From this state-of-the-art, a detailed adequation and projection towards the Implementation Readiness Levels of the future DiSSCo RI has highlighted both gaps but also those trainings matching the envisaged DiSSCo needs. The proposed methodology, results of analysis and recommendations are for use and implementation by the DPP WP2 HUMAN RESOURCES, TRAINING & USERS SUPPORT, and in particular T2.1 which aims at developing a Training Strategy for DiSSCo. The recommendations will pave the way for the future capacity building and tool-up activities to be developed in support to the provision and use of the services the RI will put in place when in full operation.

The involvement from the very beginning of this task of the representatives of the DEST platform has equally contributed to ensure the implementation of the trainings in full compliance with current practices and also future needs. The definition of the criteria used in the methodology identification and its further implementation as well as the set of recommendations extracted from the work carried out provides DEST with a robust structure and a valuable information on the various content types that can perfectly constitute an anchoring element to support the platform and to foster its promotion, enlarge its use and enhance its content by and for the entire scientific community.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

7. References

Bakker, F. T., Antonelli, A., Clarke, J., Cook, J. A., Edwards, S. V., Ericson, P. G., … Källersjö, M. (2020). The Global Museum: Natural history collections and the future of evolutionary biology and public education. PeerJ, 8, e8225. https://doi.org/10.7717/peerj.8225

Haston, E. M., & Hardisty, A. (2020). An Introduction to the Minimum Information about a Digital Specimen (MIDS) Digitisation Standard. Biodiversity Information Science and Standards, 4. e59214 https://doi.org/10.3897/biss.4.59214

Hardisty, A., Saarenmaa, H., Casino, A., Dillen, M., Görddérz, K., Groom, Q., ... & Willemse, L. (2020). Conceptual design blueprint for the DiSSCo digitisation infrastructure-DELIVERABLE D8. 1. Research Ideas and Outcomes, 6. https://doi.org/10.3897/rio.6.e54280

Heberling, J. M., Prather, L. A., & Tonsor, S. J. (2019). The changing uses of herbarium data in an era of global change: An overview using automated content analysis. BioScience, 69(10), 812– 822. https://doi.org/10.1093/biosci/biz094

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences, 118(6). https://doi.org/10.1073/pnas.2018093118

Lendemer, J., Thiers, B., Monfils, A. K., Zaspel, J., Ellwood, E. R., Bentley, A., . . . Aime, M. C. (2019). The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. BioScience, 70(1), 23-30. doi:10.1093/biosci/biz140

Nelson, G., & Ellis, S. (2019). The history and impact of digitisation and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1763), 20170391. doi:doi:10.1098/rstb.2017.0391

Page, L. M., MacFadden, B. J., Fortes, J. A., Soltis, P. S., & Riccardi, G. (2015). Digitisation of Biodiversity Collections Reveals Biggest Data on Biodiversity. BioScience, 65(9), 841-842. doi:10.1093/biosci/biv104

Paton, A., Antonelli, A., Carine, M., Forzza, R. C., Davies, N., Demissew, S., ... & Dickie, J. (2020). Plant and fungal collections: Current status, future perspectives. Plants, People, Planet, 2(5), 499-514. https://doi.org/10.1002/ppp3.10141

Schindel, D. E., & Cook, J. A. (2018). The next generation of natural history collections. PLoS Biology, 16(7), e2006125.

Synthesys NA2.3 MS 26 Report https://dissco.teamwork.com/#/files/8095854
Tamjid Mujtaba, Martin Lawrence, Mary Oliver & Michael J. Reiss (2018) Learning and engagement through natural history museums, Studies in Science Education, 54:1, 41-67, DOI: 10.1080/03057267.2018.1442820

Tormey, D. (2019). New approaches to communication and education through geoheritage. International Journal of Geoheritage and Parks, 7(4), 192-198.

Webster, M. S. (2017). The extended specimen: emerging frontiers in collections-based ornithological research: CRC Press.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

8. Glossary of terms and abbreviations

ABS: Access and Benefit-Sharing. See *Nagoya Protocol*
BID: Capacity Enhancement Support programme: this programme aims to address specific capacity needs identified by GBIF Participants by facilitating collaboration at regional and global levels (https://www.gbif.org/programme/82219/capacity-enhancement-support-programme).
Biotalent: is a blended e-learning biodiversity training programme. Biodiversity and its protection are tightly linked to a societal change which can only be achieved by a strong investment in environmental education (http://biotalent.myspecies.info/).
BRAHMS: is a database software for managing natural history collections, botanic gardens, seed banks, field surveys, taxonomic research and biogeographic study (https://herbaria.plants.ox.ac.uk/bol/).
BOTALISTA: Management software package designed to cover all the business processes of a botanical institute (living collections, Index seminum, seed bank, ex situ and in situ conservation projects, herbaria, Floras, etc.) (https://botalista.community/).
CBD: Convention of Biological Diversity is an international agreement for the protection of biological diversity.
CETAF: Consortium of European Taxonomic Facilities (https://cetaf.org/).
CITES: Convention on International Trade in Endangered Species is an international agreement for the protection of endangered species (https://cites.org/eng/disc/what.php).
COLHELPER: is a e-service of the Muséum national d'Histoire naturelle (France) for Collection management and intended for the scientific community (http://colhelper.mnhn.fr/).
COST: The European Cooperation in Science and Technology is a funding organisation for the creation of research networks, called COST Actions (https://www.cost.eu/who-we-are/about-cost/).
COST Mobilise: is a COST action aiming to foster a cooperative network in Europe to support excellent research activities, and facilitate knowledge and technology transfer around natural science collections (https://www.cost.eu/actions/CA17106/#tabs|Name:overview).
DwC: DarwinCore data standard is a data standard that offers a stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources. Originally developed by the Biodiversity Information Standards (TDWG) community, Darwin Core is 'an evolving community-developed biodiversity data standard (https://www.gbif.org/standards).
DEST: Distributed European School of Taxonomy (https://cetaf.org/dest/).
DINA: ***DI****gital Information system for* ***NA****tural history data* has its roots in a Swedish initiative to replace a heterogeneous collection of unsustainable in-house databases with a modern, web-based national collection management system (https://www.dina-project.net/wiki/Welcome_to_DINA#).
DiSSCo: Distributes System of Systematic Collections (https://www.dissco.eu/).
DiSSCo PREPARE: (https://www.dissco.eu/dissco-prepare-work-programme/).
DSI: Digitalized Sequences Information (See https://www.cbd.int/dsi-gr/).
EDIT: EU funded project European Distributed Institute of Taxonomy (www.e-taxonomy.eu, 2006 -2011)
ELIXIR: is an intergovernmental organisation that brings together life science resources from across Europe. These resources include databases, software tools, training materials, cloud storage and supercomputers (https://elixir-europe.org/about-us).
e-LTER: Integrated European Long-Term Ecosystem & Socio-Ecological Research Infrastructure(https://www.lter-europe.net/elter)
ELViS: an integrated European Loans and Visits System (https://elvis.dissco.eu/welcome).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

EMMRI: The Executive Masters in Management of Research Infrastructures is an international management development programme tailored to the specific needs of Research Infrastructures. This Programme is designed to give experienced science professionals the skills and knowledge to take on greater managerial responsibilities.
EOSC: The European Open Science Cloud is an environment for hosting and processing research data to support EU science (https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en).
European Network EURAXESS: Researchers in Motion is a pan-European initiative delivering information and support services to professional researchers (https://euraxess.ec.europa.eu/).
EuroHPC: The European High Performance Computing (https://eurohpc-ju.europa.eu/)
EU BON: Building the European Biodiversity Observation Network. The FP7 project (2012-2017) aimed to improve access to integrated biodiversity data from earth observation and in situ realms in order to provide relevant information for evidence-based decision making at local, national, and global level (http://www.eubon.eu/).
FAIR: Findable, Accessible, Interoperable, Reusable (data). This requires data standards and protocols to maximise the use of data (https://doi.org/10.1038%2Fsdata.2016.18).
FishBase: a Global Information System on Fishes (https://www.fishbase.se/home.htm).
Galaxy platform: an open source, web-based platform for data intensive biomedical research (https://usegalaxy.org/).
GBIF: Global Biodiversity Information Facility. An international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth (https://www.gbif.org/).
GDPR: General Data Protection Regulation 2016/679 is a regulation in EU law on data protection and privacy in the European Union and the European Economic Area (https://eur-lex.europa.eu/eli/reg/2016/679/oj).
HNHM: Hungarian Natural History Museum (Budapest, Hungary) (http://www.nhmus.hu/en/english_home).
ICEDIG: Innovation and consolidation for large scale digitisation of natural heritage (https://icedig.eu/).
IPR: Intellectual Property Rights.
JEMU: Joint Experimental Molecular Unit. Integrated research infrastructure funded by the Belgian Science Policy and supported by the Royal Belgian Institute of Natural Sciences (RBINS, Brussels) and the Royal Museum for Central Africa (RMCA, Tervuren) (http://jemu.myspecies.info/).
KTA: Key Training Areas.
Meise BG: Meise Botanic Garden (Meise, Belgium) (https://www.plantentuinmeise.be/en/home/).
MIDS: Minimum Information about a Digital Specimen (see https://www.tdwg.org/community/cd/mids/).
MNCN-CSIC: National Museum of Natural Sciences - Spanish National Research Council (Madrid, Spain) (https://www.mncn.csic.es/es).
MfN: Museum of Natural Sciences (Berlin, Germany) (https://www.museumfuernaturkunde.berlin/en).
MNHN: National Museum of Natural History (Paris, France) (https://www.mnhn.fr/en).
Nagoya Protocol: The Nagoya Protocol on Access and Benefit-Sharing is an addition ("protocol") to the Convention of Biological Diversity regulating the utilization of genetic resources and traditional knowledge.
NHM UiO: Natural History Museum University of Oslo (Oslo, Norway) (https://www.nhm.uio.no/english/).
NHMW: Natural History Museum Wien (Wien, Austria) (https://www.nhm-wien.ac.at/en).
NRM: The Swedish Museum of Natural History (Stockholm,Sweden) (https://www.nrm.se/en/16.html).
OER: Open Educational Resources (https://creativecommons.org/about/program-areas/education-oer/).
RBG Kew: Royal Botanic Garden Kew (London, United Kingdom) (https://www.kew.org/).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

RBINS: Royal Belgian Institute of Natural Sciences (Brussels, Belgium) (https://www.naturalsciences.be).
RECOLNAT: Research Infrastructure (https://www.recolnat.org/).
RMCA: Royal Museum for Central Africa (Tervuren, Belgium) (https://www.africamuseum.be/en).
SNSB: Staatliche Naturwissenschaftliche Sammlungen Bayerns (Munich, Germany) (https://www.snsb.mwn.de/index.php/en).
SPECIFY: Specify software processes the data associated with specimens in biological research collections. The Specify Collections Consortium represents approximately 195 collections (https://www.specifysoftware.org/) (See also https://biodiversity.ku.edu/botany/collections/specify).
PLUTO: Plant Variety Database. Common name UPOV (http://www.upov.int/pluto/en/).
UGOT: University of Gothenburg (Gothenburg, Sweden) (https://www.gu.se/en).
WIPO: World Intellectual Property Organization is the global forum for intellectual property (IP) services, policy, information and cooperation; a self-funding agency of the United Nations, with 193 member states (https://www.wipo.int/portal/en/index.html).
Workshop "Biodiversity Data Mobilization": this workshop is part of the capacity enhancement efforts associated to the Biodiversity Information for Development (BID) programme led by GBIF (https://www.gbif.org/event/82794/first-bid-capacity-enhancement-workshop-biodiversity-data-mobilization).

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

D2.3 APPENDIX

Appendix A: Survey tool - Overview document of the survey tool, associated datasets and analysis methods and rules

1. Survey component

1.1 List of data type and data standardisation rules

Data training criteria	Filling rules	**Standardisation rules** (Post-survey)
Institution	Institutional acronyms list	---
City	City list	---
Title of the training	Free field	---
Main Category	KTA list	---
sub category 1	KTA subcategories #1 list	---
sub category 2	KTA subcategories #2 list	---
target audience	Free field	Non-standardised data
Length (days)	Free field	To compare data across trainings, time length of each training
Trainings capacity/n°of possible participants	Free field	To align the data across trainings, were considered the followings: (1) When the given information was a range (min-max), the maximum number of participants was selected (e.g., if 10-20 participants, we used 20), (2) When the given information was a "group", the number of participants was estimated to 5, (3) For online videos and missing data, the capacity is zero
Training developed in YYYY	Free field	Non-standardised data
Frequency of the training	Free field	Non-standardised data
Approach chosen for the training	Free field	Non-standardised data
Position of trainer in the institution	Free field	The information filled in by the partners has been grouped by type of position according to: Scientific staff, Engineer staff, Technical staff, Curator, Collection staff, IT, Data manager, Policy staff, Head of department, Managers, Communication staff, Other (non classify text : « internal and external trainers involved », input from others, member of citizen science board, Seed Bank partnership coordinators).
Training fees (if applicable)	Free field	The data were classified into three types: (1) yes: fees applicable, (2) N/A : no fees, (3) possibly: information is missing or needs to be further investigated
Related credits/Certificate	Free field	Based on the information provided by the partners, we have classified the courses into three levels of certification: (1) no certification, (2) certificates given but not qualifying, e.g. training attendance certificate, (3) Qualifying certificates offered, e.g. when the training provides university credits.
Virtual, physical or hybrid	List of support : (1) Physical (face-to-face training), (2) Virtual (online training with a trainer and videos), (3) Hybrid (training that are partly physical and partly virtual).	---
Can be developped into virtual	Free field	The partners' information has been classified into 4 different levels : (1) yes, the courses can be developed in virtual access, (2) Possibly, the partners indicated that adaptation could be foreseen, (3) No, the training cannot be delivered virtually, (4) Partly, a part of the training can be developed in physics, but not entirely.
Language (s) of the training	Language list	---
Comments on language	Free field	---
Remarks	Free field	---
Contact details	Free field	---

1.2 KTAs and subcategories list

Key Training Area (KTA)	Subcategories 1	Subcategories 2	Note
History of collections	---	---	e.g. institutional history, decolonisation, collector, expeditions
Basic knowledge related to collections	**Biology**	---	
Specimens A specimen is the physical object or observation (photo, sound,...), or a direct derivative, eg image of specimen, not the product of analysis. Not necessarily yet curated	**Museum specimens** A specimen is the physical object or observation: photo sound; or a direct derivative, eg image , not the product of analysis. Not necessarily yet curated, hosted by the institution, and it must be inventoried	Collection/Sampling of specimen Field sampling Exhibition	
	specimen curation	Care & maintenance Collection management Preventive conservation	
	digitization	databasing metadata data quality imaging digitisation techniques	
	specimen based research	taxonomy Biodiversity Evolution Ecology Conservation Nomenclature Geodiversity Genomic and molecular Data analysis Methodology Mineralogy/Geology Methodology Archaeology	
Data	**data curation**	---	e.g. data quality
	data management	---	e.g. storage
	data standards and interoperability	---	
	data processing/use	---	e.g. genomics data
	Databasing	---	
	software skills	applying evolutionary models analytic research	
Equipment use	---	---	
Policy & legislation	**policy mandates**	Cultural Heritage CITES ABS	
	Implementation for collection	curation research	
Training multipliers	**training the trainers**	---	
	Citizen science	---	training citizens
Text & Media	**knowledge transfer**	exhibitions media	

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

1.3 IRLs list and definitions

IRLs DiSSCo Prepare	Matching IRLs to training	
WPs and definition	**Sub-categories**	**Definition**
SCIENCE User needs & socioeconomic impact (WP1) Human Resources, Training & Users Support (WP2) DiSSCo, like any RI, needs to be in full and constant alignment with the needs of its identified user base (see above). This requires the existence of a practical scientific evaluation framework, which allows the RI to inform decisions around its future scientific programme based on the ever-changing needs of its user base.	Collection/sampling of specimen	Collecting within a scientific framework.
	Specimen curation	Knowledge required to curate NSC.
	Museum specimen based research	All research activities based on NSC.
	Taxonomy	
	Ecology	
	Biodiversity	
	Geodiversity	
	Evolution	
	Data analysis	Software skills, analytical and technical research, or data manipulation/transformation.
	Publication and outputs	Publication process and outputs measuring
	Citizens science - training the trainers	Use of digitization shared tools, databasis, or platforms.
	Citizens science - training citizens	Use of digitization shared tools, databasis, ot platforms.
DATA Capacity enhancement (WP3) As a distributed data infrastructure, DiSSCo relies on the capacity of the national and institutional nodes to provide and enrich data in a consistent, harmonised manner, compatible within the overall RI.	Pre-digitisation curation	Accepting & accessioning specimens/objects, workflow to get them ready for digitisation (e.g., labels, propoer storage room, correct names and metadata), or metadata capture and handling.
	Standarisation interoperability	Data processing workflows that convert the structure of disparate datasets into a common data format (e.g., data FAIRness, specimens data refinery, or molecular standards and processes...).
	Data curation	All the activities surrounding the documentation and accessibility of NSC data.
	Digitisation	All processes that enable converting a specimen/object into a digital format (technicity, standards assessment, protocols).
	Databasing	Integration and management of a specimen/object and associated (meta)data in databases.
	Data sharing	Dissemination through public e-portal.
TECHNOLOGICAL Common Resources and Standards (WP5) Technical Architecture & Services provision (WP6) Predominantly operating as a data infrastructure, DiSSCo needs to fully understand the current limitations of technological implementations of existing e-infrastructures and subsequently establish its data architecture as well as the technical specifications of its future services	Operating	Running scientific equipment (e.g., sequencing or 2D/3D imaging facilities).
	Developement	Conceptual design and implementation of software and digital tool (e.g., data architecture, data carpentry).
	IT data managment	All activities surrounding the manipulation, protection, storage, or safety of data.
	e-Monitoring	Digital management services (eg. E-service, helpdesk, mediation services): adminitrators and users.
ORGANISATIONAL Governance, Policy & Legal frameworks (WP7) To proceed with the set-up of its new legal entity, DiSSCo needs to further set out its overall organisational (governance and management) structures, strategic and operational plan, the involvement of funders and infrastructure siting as well as proceed with harmonisation of policies across its facilities.	Policy	Covers all areas of legislative practices tied to both internal (e.g., collection strategy, data management, or IT strategy) and external policies (e.g., Convention on Biological Diversity, Nagoya Protocol on ABS, or any Global activity contributing to data alignment initiatives for bio-geo collections).
	Governance	Legal and decisional expertise related to inter-community partnership activities and organisation (e.g., decision making, governance levels, stakeholders engagement).
	Management	Defining strategies and management plans, administering and coordinating projects & activities (incl. Services).
	Communication	Expertise related to institutional, international, public, media, partnership, and internal communication.
FINANCIAL Business Framework (WP4) Before DiSSCo fully embarks on implementation activities, it is necessary to have a robust financial framework in place that enables accurate calculations of costs and contributions. This framework should deliver the trust to secure financial commitments both at a national and European level.	Management	Management of financial and budgetary strategy: expertise in all accounting, management, treasury, tax and financial communication functions.
	Financial awarness	Drawing up and monitoring project budgets, balance of payments, payroll, equipment investments.
	Basic information	Provision of financial information on projects.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

2. Data and figures

2.1 List of the 83 existing trainings

Institution	Title of the training	Main Category
NHMW	Fatales Design, Evolutionsvorlesung Uni Konstanz	Basic knowledge related to collections
NHMW	Grundprinzipien der Evolution, Teil 1 / principles of evolution, part 1	Basic knowledge related to collections
NHMW	Grundprinzipien der Evolution, Teil 2 / principles of evolution, part 2	Basic knowledge related to collections
NHMW	Grundprinzipien der Evolution, Teil 3 / principles of evolution, part 3	Basic knowledge related to collections
NHMW	Insektensterben / decline of insect species	Basic knowledge related to collections
NHMW	Due Diligence for utilizing genetic resources in the context of the Nagoya-Protocol	Policy & legislation
NHMW	e-Tutorial 1: DSP-Methode zur metrischen Geschlechtsbestimmung / DSP method of (anthropological) sex estimation	Specimens
NHMW	e-Tutorial 2: Beckenmerkmale / human pelvic features	Specimens
NHMW	Handling historical and fragile museum specimens and types for scientific research	Specimens
NHMW	Meteorite identification and classification (Mineralogy)	Specimens
NHMW	Molecular Systematics Practical course for DNA analyses of scientific collections	Specimens
NHMW	Nomenclature and taxonomy	Specimens
NHMW	Scientific illustrations: scientific drawing, multifocus imaging and scanning electron microscopy	Specimens
NHMW	Textile archaeology workshop (Archeology)	Specimens
NHMW	Phylogenetic reconstructions using TNT	Specimens / Data
NHMW	Recognition of meteorite impact structures on Earth – Shock metamorphism in quartz (Mineralogy / Geology)	Specimens / Data
NHMW	Citizen Science	Training multipliers
MfN	Data management	Data
MfN	data standards in natural sciences	Data
MfN	long term data storage	Data
MfN	Specify/DINA	Data
MfN	Scientific Data and Information Seminar (series of seminar sessions, currently internal)	Data
MfN	Basic collection techniques	Specimens
MfN	Digitization techniques	Specimens
MfN	Disaster preparedness	Specimens
MfN	Integrated Pest Management	Specimens
MfN	Moving collections	Specimens
MfN	Natural History Collections	Specimens
MfN/SNSB	Access and Benefit Sharing	Policy & legislation
MfN/SNSB	Shipping Workshop	Policy & legislation
MfN/SNSB	Care of wet collections	Specimens
MNCN-CSIC	Analysis using non-destructive techniques	Equipment use
MNCN-CSIC	Best practice advice on implementing ABS policy mandates	Policy & legislation
MNCN-CSIC	Best practice advice on implementing Cultural Heritage policy mandates	Policy & legislation
MNCN-CSIC	CITES in Scientific Collections: Identification	Policy & legislation
MNCN-CSIC	CITES in Scientific Collections: Management	Policy & legislation
MNCN-CSIC	Techniques in zoological research: introduction to museology	Policy & legislation
MNCN-CSIC	Basic techniques on care and maintenance of natural history collections with historical specimens	Specimens
MNCN-CSIC	Conservation, use and management of DNA and tissue banks	Specimens
MNCN-CSIC	Introduction to basic techniques for DNA analyses of scientific collections	Specimens
MNCN-CSIC	Diploma in Management, Care and Exhibition of Natural Sciences Collections	Specimens / Policy & Legislation
MNHN	Introduction to Linux and command line	Data
MNHN	Introduction to genomics data	Data
MNHN	Introduction to the analysis plateform GALAXY	Data
MNHN	Programing with Python for beginners	Data
MNHN	Scientific Illustration	Specimens
MNHN	Training in Scientific Imaging and Segmentation	Specimens
MNHN	BID Capacity Enhancement Workshop: Biodiversity Data Mobilization	Specimens / Data
MNHN	BID Capacity Enhancement Workshop: Biodiversity Data Use	Specimens / Data
MNHN	Integrative taxonomy in the "big data" era	Specimens / Data
MNHN	Training in Morphometry and morphological analysis	Specimens / Data

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

MeiseBG	Part of the courses given to GBIF BID : data mobilisation, data quality, usage of data, project management, Niche modelling,Policy making	Data
MeiseBG	Diatoms Taxonomic workshops	Specimens
MeiseBG	Seed banks managment course	Specimens
MeiseBG	Herbarium training	Specimens
MeiseBG	Part of COST Mobilise training , Best practices in digitization, How to set up a digitization projects (in form of a card game), Usage of GIS for data quality check and errors detection	Specimens
MeiseBG	Botanical illustrations course	Specimens
MeiseBG	Courses and workshops for high and elementary school including aiming at the teachers	Training multipliers
RBG Kew	Field work and relevant policy frame work (CBD, NP, Fieldwork Planning, TK, and CITES)	Policy & legislation
RBG Kew	Plant Nomenclature	Specimens
RBG Kew	Seed Banking	Specimens
RBG Kew	Tropical Plant ID	Specimens
RBG Kew	Best practice in plant tissue collection and preservation	Specimens / Data
RBG Kew	Use and Curation of Biocultural (ethnobiological) Collections	Specimens / Policy & legislation
RBINS	An introducing to "Access and Benefit Sharing" in public research	Policy & legislation
RBINS	Curation and Management of Natural Sciences Collections	Specimens
RBINS	DNA analysis of museum specimens (DNA extraction, DNA quality evaluation and data analysis)	Specimens
RBINS	Creating 3D digital twins of your collection specimens at a budget: Hands-on training on photogrammetry for beginners	Specimens / Data
RBINS	Virtual Collections: How to begin? Getting started with the introduction course to 2D+ and 3D techniques	Specimens / Data
RBINS	Training citizens on the introduction to taxonomy	Training multipliers
RMCA	PostGIS and WMS protocoles (GeoServer)	Data
RMCA	QGIS advanced (raster treatment and watershed detection)	Data
RMCA	QGIS for distribution data	Data
RMCA	Training Conservation of museum zoology collections	Specimens
RMCA	Training course on FishBase and fish taxonomy	Specimens
RMCA	Training on wood biology	Specimens
NRM	Plant collection and mounting	Specimens
NRM	Training in curation of palaeontological collections	Specimens
NRM	Training in identification of minerals/rocks with analytical techniques, organizing systematic mineral collections, and risk management	Specimens
NRM	Training in zoological nomenclature with focus on status of type secimens	Specimens
UGOT	Python for biologists	Data
UGOT	Spatial Analysis in R	Data
HNHM	Zootaxonomy	Specimens

List of the 8 formations not included in the analyses:

MeiseBG	Capacity building in plant taxonomy (Part of Flore de'Afrique Centrale)	Specimens
MeiseBG	Capacity building in fundal diversity and culturing	Specimens
MeiseBG	Capacity building in tropical diatom taxonomy	Specimens
MeiseBG	Plant identification course	Specimens
MeiseBG	Tropical seaweed course	Specimens
MeiseBG	Invasive species	Specimens
MeiseBG	Future EOSC staff competences needs trainings within the implementation of EOSC in Belgium	Data
MeiseBG	Future courses in framework of the implementation of DiSSCo in Flanders	Specimen / Data

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

2.2 Subsample screenshots of datasets

2.2.1 Catalogue of NH existing trainings

Institution	City	Title of the training	KTA- Main Category	KTA - subcategory #1	KTA - subcategory #2
HNHM	Budapest	Zootaxonomy	Specimens	Specimen based research	Evolution / taxonomy
MeiseBG	Meise	Diatoms Taxonomic workshops	Specimens	Specimen based research	Ecology / taxonomy
MeiseBG	Meise	Seed banks managment course	Specimens	Specimen based research / Specimen curation	Ecology / taxonomy / Collection management
MeiseBG	Meise	Courses and workshops for high and elementary school including aiming at the teachers	Training multipliers	specimen based research / training the trainers	Ecology / taxonomy
MeiseBG	Meise	Herbarium training	Specimens	Specimen curation	Collection management
MeiseBG	Meise	Part of the courses given to GBIF BID : data mobilisation, data quality, usage of data, project management, Niche	Data	data curation / data management / data processing/use	
MeiseBG	Meise	Part of COST Mobilise training , Best practices in digitization, How to set up a digitization projects (in form of a card	Specimens	digitization	
MeiseBG	Meise	Botanical illustrations course	Specimens	digitization	
MfN	Berlin	Basic collection techniques	Specimens	Specimen curation	Care and maintenance / Preventive conserv
MfN	Berlin	Data management	Data	Data management / Data standards and Interoperability	
MfN	Berlin	data standards in natural sciences	Data	Data standards and interoperability	
MfN	Berlin	Digitization techniques	Specimens	Digitization	Databasing metadata / Imaging / Digitizatio
MfN	Berlin	Disaster preparedness	Specimens	Specimen curation	Preventive conservation
MfN	Berlin	Integrated Pest Management	Specimens	Specimen curation	Preventive conservation
MfN	Berlin	long term data storage	Data	Data management / Data standards and Interoperability	
MfN	Berlin	Moving collections	Specimens	Specimen curation	Care and maintenance / Preventive conserv
MfN	Berlin	Natural History Collections	Specimens	Specimen curation	
MfN	Berlin	Specify/DINA	Data	Data management / databasing	
MfN	Berlin	Scientific Data and Information Seminar (series of seminar sessions, currently internal)	Data	Data management / Data standards and Interoperability	
MfN/SNSB	Berlin/Munich	Access and Benefit Sharing	Policy & legislation	policy mandates	ABS
MfN/SNSB	Berlin/Munich	Care of wet collections	Specimens	Specimen curation	
MfN/SNSB	Berlin/Munich	Shipping Workshop	Policy & legislation	policy mandates	
MNCN-CSIC	Madrid	Analysis using non-destructive techniques	Equipment use	---	
MNCN-CSIC	Madrid	Basic techniques on care and maintenance of natural history collections with historical specimens	Specimens	Specimen curation	Care & maintenance
MNCN-CSIC	Madrid	Best practice advice on implementing ABS policy mandates	Policy & legislation	Policy mandates /policy implementation	ABS
MNCN-CSIC	Madrid	Best practice advice on implementing Cultural Heritage policy mandates	Policy & legislation	Policy mandates /policy implementation	Cultural Heritage
MNCN-CSIC	Madrid	CITES in Scientific Collections: Identification	Policy & legislation	Policy mandates	CITES
MNCN-CSIC	Madrid	CITES in Scientific Collections: Management	Policy & legislation	Policy implementation	CITES
MNCN-CSIC	Madrid	Conservation, use and management of DNA and tissue banks	Specimens	Specimen curation	Collection management
MNCN-CSIC	Madrid	Diploma in Management, Care and Exhibition of Natural Sciences Collections	Specimens / Policy & Legislation	Specimen curation, Digitisation / Policy mandates, policy impleme	Care & maintenance, Collection management
MNCN-CSIC	Madrid	Introduction to basic techniques for DNA analyses of scientific collections	Specimens	Specimen based research	Genomic and molecular
MNCN-CSIC	Madrid	Techniques in zoological research: introduction to museology	Policy & legislation	Policy mandates	Cultural Heritage, ABS
MNHN	Paris	BID Capacity Enhancement Workshop: Biodiversity Data Mobilization	Specimens / Data	Digitization / Specimen based research / Data curation / Data m	Databasing metadata/data quality/digitisa
MNHN	Paris	BID Capacity Enhancement Workshop: Biodiversity Data Use	Specimens / Data	Digitization / Specimen based research / Data curation / Data m	Databasing metadata/data quality/digitisa
MNHN	Paris	Integrative taxonomy in the "big data" era	Specimens / Data	Specimen based research / Software skills	Taxonomy / Evolution

2.2.2 IRLs and training objectives cross-matches

IRLs vs Training

Instructions :
y : yes
p : Partly (In this case, please add comments)
dk : Do not know
n : No

		SCIENCE											
		Collection/sampling of specimen	Specimen curation	Museum specimen based research	Taxonomy	Ecology	Biodiversity	Geodiversity	Evolution	Data analysis	Citizens science - training the trainers	Citizens science - training citizens	Publication and outputs
HNHM	Zootaxonomy	n	n	n	y	n	y	n	y	n	n	n	n
MeiseBG	Herbarium training	y	y	y	y	n	n	n	n	y	y	y	n
MeiseBG	Part of the courses given to GBIF BID : data mobilisation, data quality, usage of data, project management, Niche modelling,Policy making	y	y	y	n	n	n	n	n	y	y	y	y
MeiseBG	Part of COST Mobilise training , Best practices in digitization, How to set up a digitization projects (in form of a card game), Usage of GIS for data quality check and errors detection	y	y	y	n	n	n	n	n	y	y	y	y
MeiseBG	Seed banks managment course	y	y	y	y	y	y	n	y	y	n	n	y
MeiseBG	Diatoms Taxonomic workshops	Y	Y	Y	Y	Y	Y	N	Y	Y	n	n	y
MeiseBG	Courses and workshops for high and elementary school including aiming at the teachers	y	n	y	y	y	y	n	y	n	y	y	n
MeiseBG	Botanical illustrations course	Y	Y	Y	Y	n	y	y	n	n	y	y	y
MfN	Basic collection techniques	p	y	n	n	n	n	n	n	n	n	n	n
MfN	Digitization techniques	n	p	n	n	n	n	n	n	n	n	n	n
MfN	Disaster preparedness	n	n	n	n	n	n	n	n	n	n	n	n
MfN	Integrated Pest Management	n	y	n	n	n	n	n	n	n	n	n	n
MfN	Moving collections	n	p	n	n	n	n	n	n	n	n	n	n
MfN	Natural History Collections	p	p	p	n	n	n	n	n	n	p	n	n
MfN	Specify/DINA	n	n	n	n	n	p	p	n	n	n	n	n
MfN	Scientific Data and Information Seminar (series of seminar sessions, currently internal)	n	n	n	n	n	n	n	n	p	n	n	y
MfN	Data management	n	n	n	n	n	n	n	n	p	n	n	dk
MfN	data standards in natural sciences	n	n	n	n	n	n	n	n	p	n	n	dk
MfN	long term data storage	n	n	n	n	n	n	n	n	p	n	n	dk
MfN/SNSB	Access and Benefit Sharing	p	p	n	n	n	n	n	n	n	n	n	n
MfN/SNSB	Care of wet collections	n	y	n	n	n	n	n	n	n	n	n	n
MfN/SNSB	Shipping Workshop	n	p	n	n	n	n	n	n	n	n	n	n
MNCN-CSIC	Conservation, use and management of DNA and tissue banks	p	y	y	p	p	p	n	p	n	n	n	n
MNCN-CSIC	Best practice advice on implementing ABS policy mandates	y	n	p	p	n	p	n	p	n	y	p	y
MNCN-CSIC	Best practice advice on implementing Cultural Heritage policy mandates	y	n	p	n	n	p	p	n	n	y	p	p
MNCN-CSIC	Techniques in zoological research: introduction to museology	y	y	y	y	y	y	n	y	y	p	n	y
MNCN-CSIC	Introduction to basic techniques for DNA analyses of scientific collections	n	n	n	n	n	n	n	n	n	n	n	n

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

2.3 Datasets: training catalogue metadata
[Note: see the standardisation method for each data type in section "1.1"]

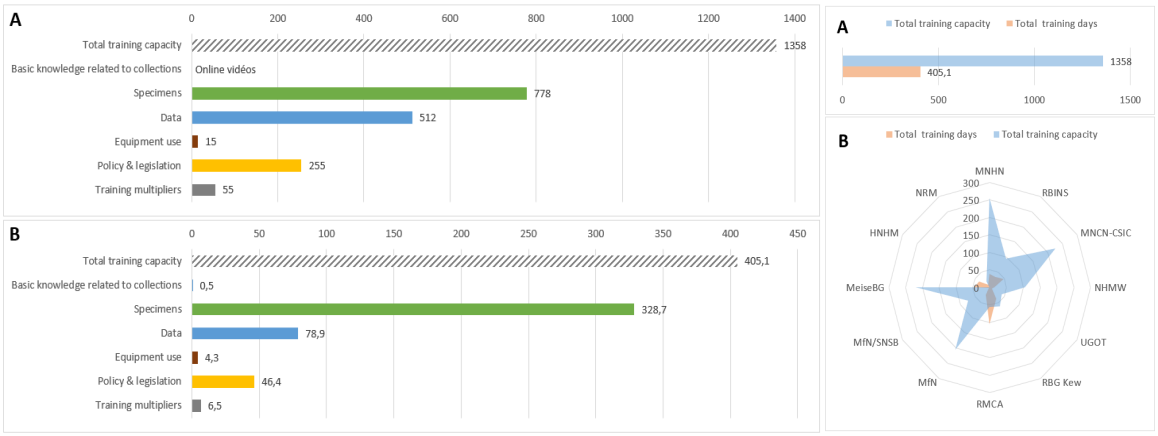


Figure 3. **A)** Distribution of capacity and **B)** number of days per KTA.
Figure 4. Average length and in-take capacity of individual training activities: **A)** Total number of training days and in-take capacity, **B)** training days and in-take capacity per institution.

Dataset Figure 3:

Dataset : Training capacity by KTA		History of collections	Basic knowledge related to collections	Specimens	Data	Equipment use	Policy & legislation	Text & Media	Training multipliers	Total training capacity
MNHN	Paris	0	0	196	240	0	0	0	0	256
RBINS	Brussels	0	0	56	21	0	15	0	25	96
MNCN-CSIC	Madrid	0	0	100	0	15	140	0	0	225
NHMW	Wien	0	0	70	12	0	20	0	15	105
UGOT	Gothenburg	0	0	0	40	0	0	0	0	40
RBG Kew	London	0	0	51	4	0	20	0	0	61
RMCA	Tervuren, Belgium	0	0	40	15	0	0	0	0	55
MfN	Berlin	0	0	125	80	0	0	0	0	205
MfN/SNSB	Berlin/Munich	0	0	15	0	0	60	0	0	75
MeiseBG	Meise	0	0	105	100	0	0	0	15	220
HNHM	Budapest	0	0	0	0	0	0	0	0	0
NRM	Stockholm	0	0	20	0	0	0	0	0	20
	Total training capacity	0	0	778	512	15	255	0	55	1358

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Dataset Figure 4:

Dataset : Training days corresponding to KTAs		History of collections	Basic knowledge related to collections	Specimens	Data	Equipment use	Policy & legislation	Text & Media	Training multipliers	Total number of training days
MNHN	Paris	0	0	35,8	26,8	0	0	0	0	39,8
RBINS	Brussels	0	0	31	6	0	2	0	0,5	33,5
MNCN-CSIC	Madrid	0	0	33,4	0	4,3	31,6	0	0	47,9
NHMW	Wien	0	0,5	11,3	3	0	1,8	0	1	14,6
UGOT	Gothenburg	0	0	0	7	0	0	0	0	7
RBG Kew	London	0	0	32	3	0	9	0	0	37
RMCA	Tervuren, Belgium	0	0	84,5	20	0	0	0	0	104,5
MfN	Berlin	0	0	17,1	8,1	0	0	0	0	25,2
MfN/SNSB	Berlin/Munich	0	0	1	0	0	2	0	0	3
MeiseBG	Meise	0	0	37	5	0	0	0	5	47
HNHM	Budapest	0	0	35,6	0	0	0	0	0	35,6
NRM	Stockholm	0	0	10	0	0	0	0	0	10
	Total training days	0	0,5	328,7	78,9	4,3	46,4	0	6,5	405,1

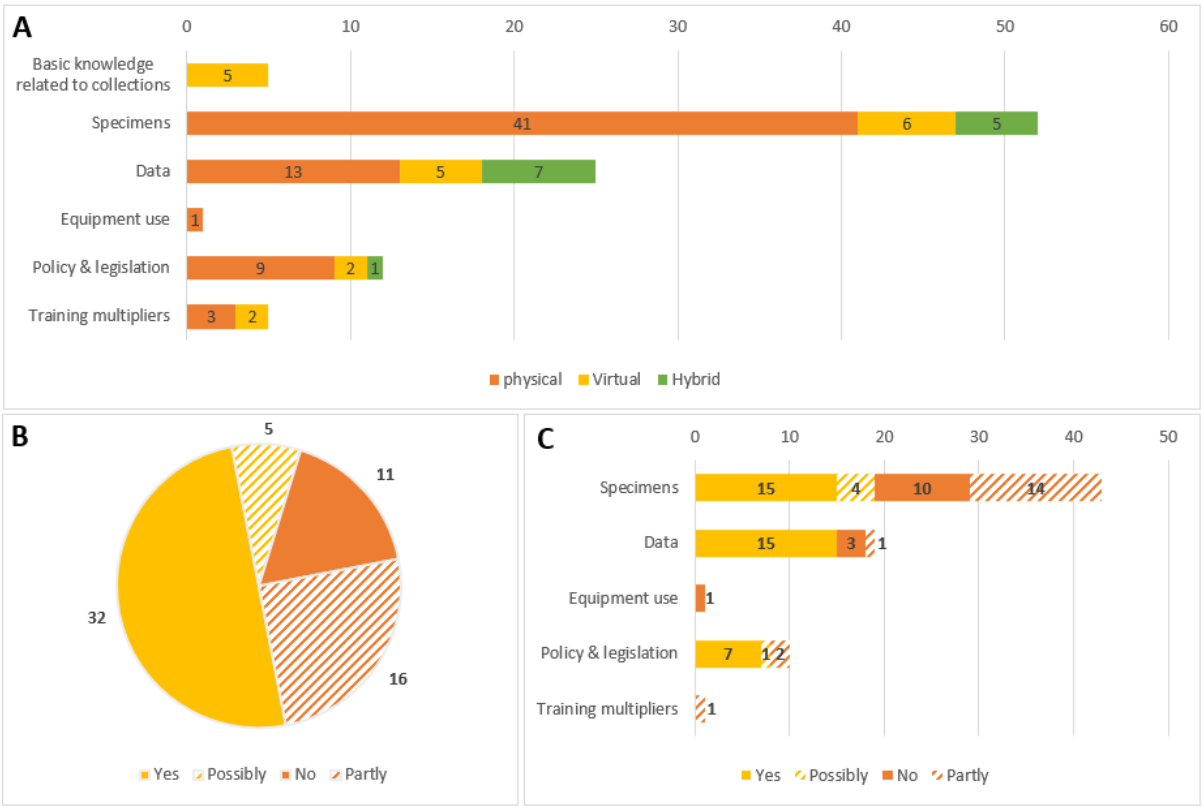


Figure 5. Training support, in terms of physical, virtual, and hybrid learning: **A)** Distribution of support modes per KTA; **B)** Distribution of physical and hybrid support modes that can be developed virtually; **C)** Distribution of physical and hybrid support modes per KTA that can be developed virtually.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

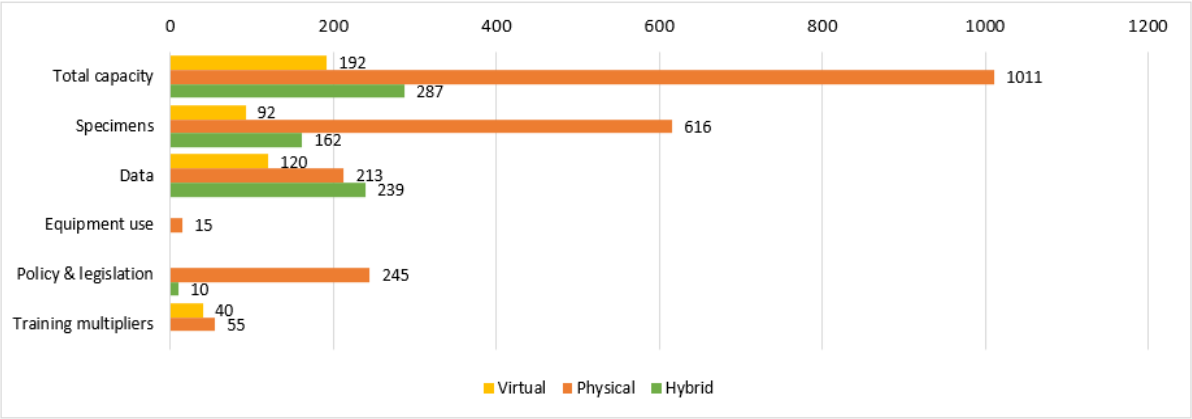


Figure 6. Distribution of support modes (physical, virtual, and hybrid) by KTA with the highest number of participants indicated at the right end of each bar.

Dataset figure 5 and 6:

	TT training	History of collections	Basic knowledge related to collections	Specimens	Data	Equipment use	Policy & legislation	Text & Media	Training multipliers
Support									
physical	**59**	0	0	41	13	1	9	0	3
Virtual	**19**	0	5	6	5	0	2	0	2
Hybrid	**10**	0	0	5	7	0	1	0	0
Physical to Virtual									
Yes	**32**	0	0	15	15	0	7	0	0
Possibly	**5**	0	0	4	0	0	1	0	0
No	**11**	0	0	10	3	1	0	0	0
Partly	**16**	0	0	14	1	0	2	0	1

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

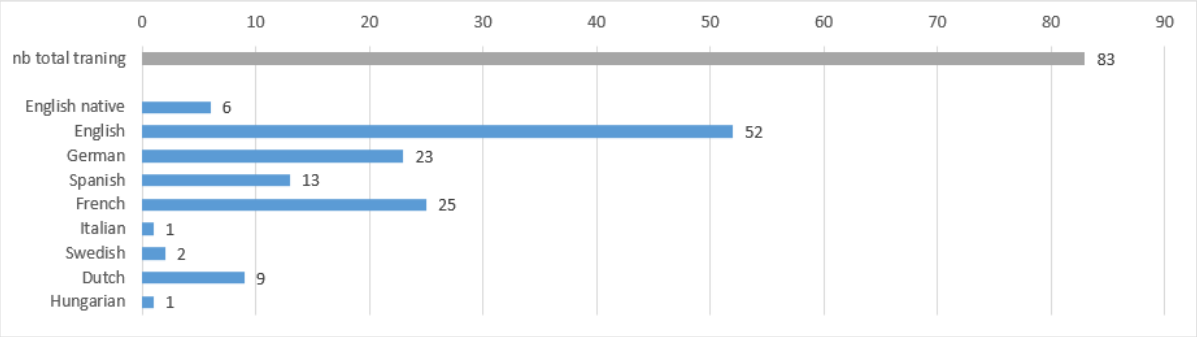


Figure 7. Distribution of the languages used for a total of 83 training.

Dataset figure 7:

		English native	English	German	Spanish	French	Italian	Swedish	Dutch	Hungarian
MNHN	Paris	0	6	0	2	8	0	0	0	0
RBINS	Brussels	0	6	0	0	4	0	0	4	0
MNCN-CSIC	Madrid	0	3	0	10	0	0	0	0	0
NHMW	Wien	0	10	13	0	1	1	0	0	0
UGOT	Gothenburg	0	2	0	0	0	0	0	0	0
RBG Kew	London	6		0	0	0	0	0	0	0
RMCA	Tervuren, Belgium	0	6	0	0	6	0	0	0	0
MfN	Berlin	0	10	7	0	0	0	0	0	0
MfN/SNSB	Berlin/Munich	0	2	3	0	0	0	0	0	0
MeiseBG	Meise	0	5	0	1	6	0	0	5	0
HNHM	Budapest	0	0	0	0	0	0	0	0	1
NRM	Stockholm	0	2	0	0	0	0	2	0	0
	Total language occurencies	6	52	23	13	25	1	2	9	1

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project

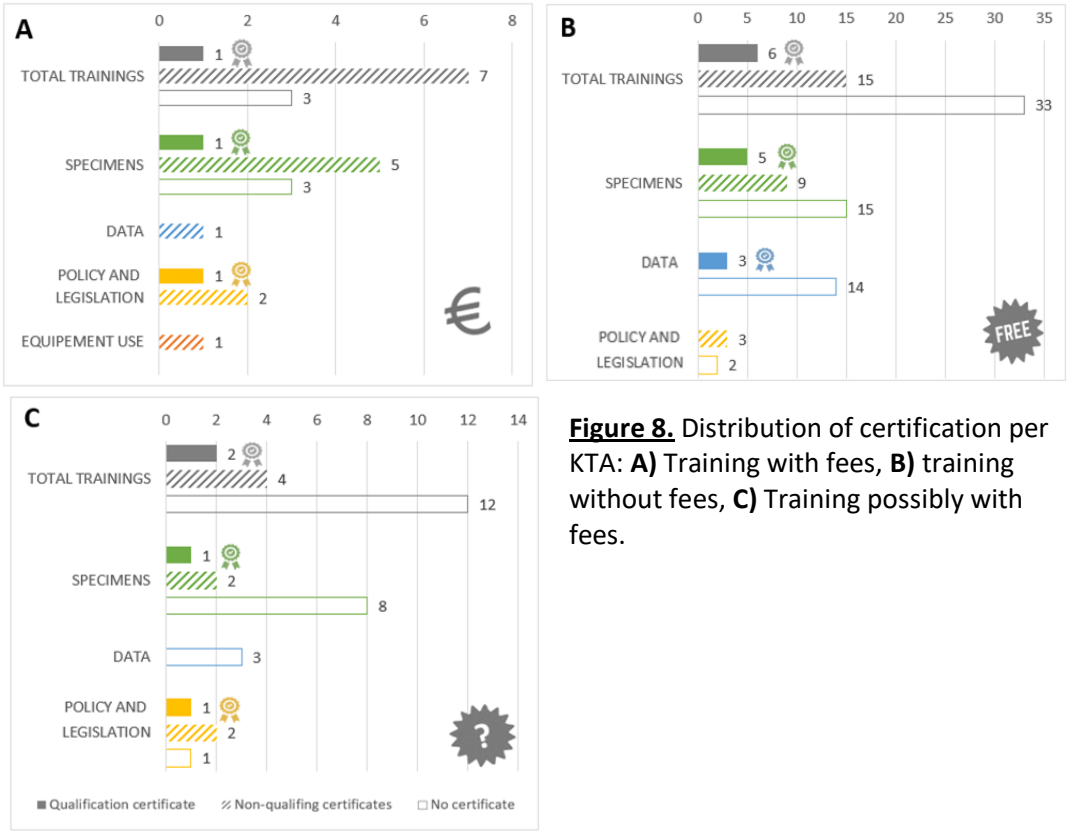


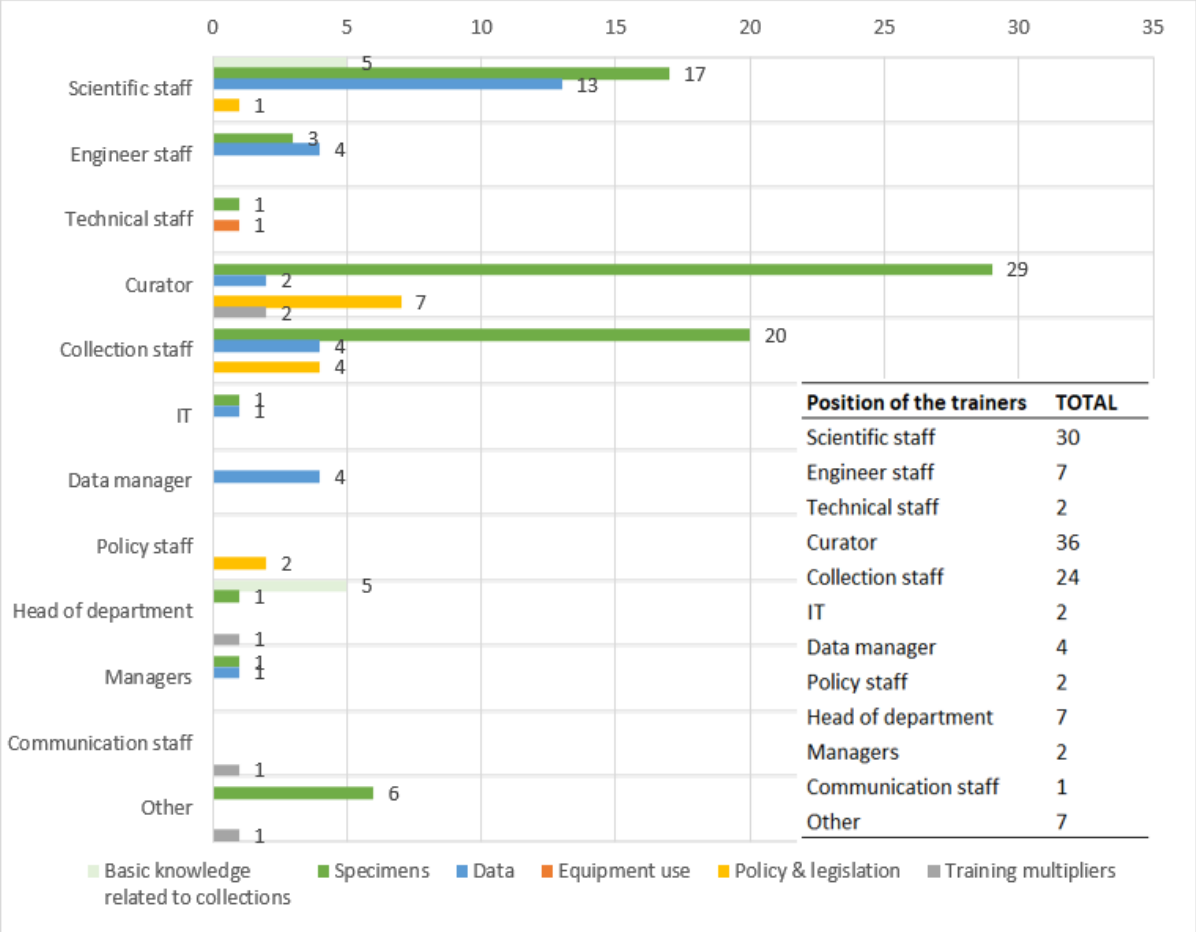
Figure 8. Distribution of certification per KTA: **A)** Training with fees, **B)** training without fees, **C)** Training possibly with fees.

Dataset figure 8:

	Fees involved	Total trainings	History of collections	Basic knowledge related to collections	Specimens	Data	Equipment use	Policy & legislation	Text & Media	Training multipliers
Applicable fees for trainings with qualification certification	Yes	1	0	0	1	0	0	1	0	0
	Possibly	2	0	0	1	0	0	1	0	0
	N/A	6	0	0	5	3	0	0	0	0
Applicable fees for trainings with non-qualification certification	Yes	7	0	0	5	1	1	2	0	0
	Possibly	4	0	0	2	0	0	2	0	0
	N/A	15	0	0	9	3	0	3	0	0
Applicable fees for trainings with no certification	Yes	3	0	0	3	0	0	0	0	0
	Possibly	12	0	0	8	3	0	1	0	2
	N/A	33	0	5	15	14	0	2	0	1
Total training with…										
…qualification certification		9	0	0	7	3	0	2	0	0
…non-qualification certification		26	0	0	16	4	1	7	0	0
…no certification		48	0	5	26	17	0	3	0	3
…fees		11	0	0	9	1	1	3	0	0
…possibly fees		18	0	0	11	3	0	4	0	2
…no fees (N/A)		54	0	5	29	20	0	5	0	1
Total		83	0	5	49	24	1	12	0	3

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+ Synthesis of Systematic Resources a DiSSCo project



Position of the trainers	TOTAL
Scientific staff	30
Engineer staff	7
Technical staff	2
Curator	36
Collection staff	24
IT	2
Data manager	4
Policy staff	2
Head of department	7
Managers	2
Communication staff	1
Other	7

Figure 9. Distribution of the position of trainers per KTA.

Dataset figure 9:

Trainer positions	TT training	History of collections	Basic knowledge related to collections	Specimens	Data	Equipment use	Policy & legislation	Text & Media	Training multipliers
Scientific staff	30	0	5	17	13	0	1	0	0
Engineer staff	7	0	0	3	4	0	0	0	0
Technical staff	2	0	0	1	0	1	0	0	0
Curator	36	0	0	29	2	0	7	0	2
Collection staff	24	0	0	20	4	0	4	0	0
IT	2	0	0	1	1	0	0	0	0
Data manager	4	0	0	0	4	0	0	0	0
Policy staff	2	0	0	0	0	0	2	0	0
Head of department	7	0	5	1	0	0	0	0	1
Managers	2	0	0	1	1	0	0	0	0
Communication staff	1	0	0	0	0	0	0	0	1
Other	7	0	0	6	0	0	0	0	1

2.4 Datasets: IRLs and training objectives

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

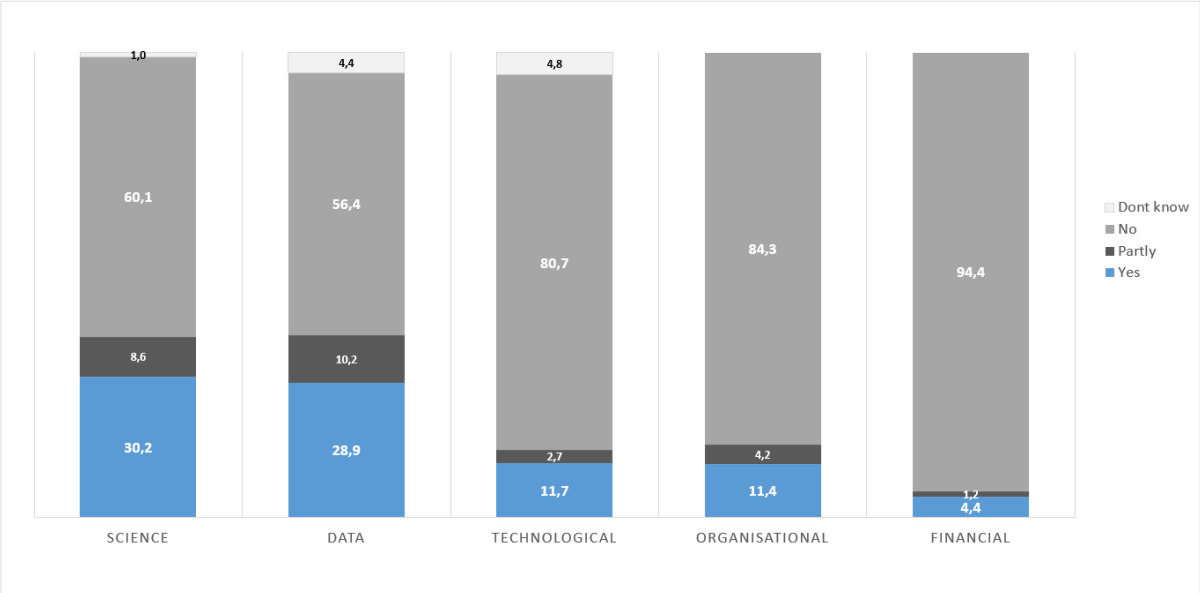


Figure 10. Percentage of responses (YES, PARTLY, DON'T KNOW) obtained for each IRL.

Dataset figure 10:

	Total	SCIENCE	DATA	TECHNOLOGICAL	ORGANISATIONAL	FINANCIAL
Yes	533	301	144	39	38	11
Partly	163	86	51	9	14	3
No	1663	599	281	268	280	235
Dont know	48	10	22	16	0	0
Total	2407	996	498	332	332	249

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

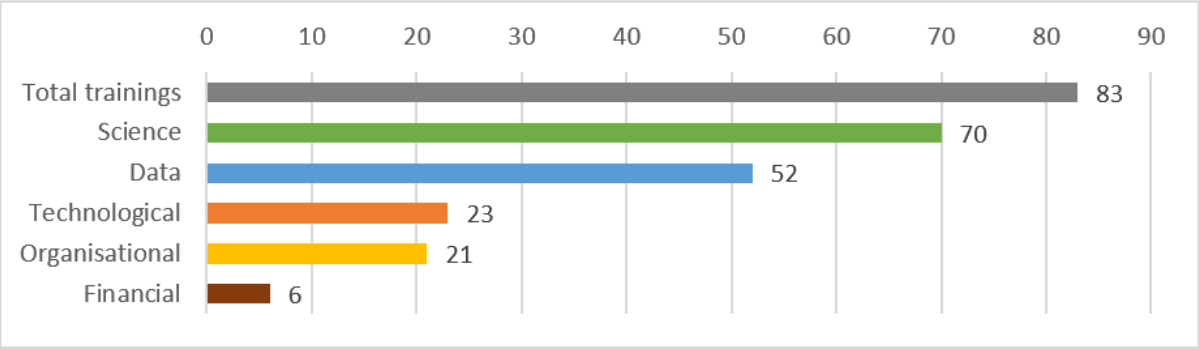


Figure 11. Number of training matching the IRLs.

Dataset figure 11:

		Nb total trainings	SCIENCE	DATA	TECHNOLOGICAL	ORGANISATIONAL	FINANCIAL
MNHN	Paris	10	10	6	7	0	1
RBINS	Brussels	6	4	3	3	2	0
MNCN-CSIC	Madrid	10	10	7	3	4	0
NHMW	Wien	17	17	7	0	6	1
UGOT	Gothenburg	2	2	2	0	0	0
RBG Kew	London	6	6	6	1	4	1
RMCA	Tervuren, Belgium	6	5	6	2	0	1
MfN	Berlin	11	3	5	5	0	0
MfN/SNSB	Berlin/Munich	3	1	0	0	2	0
MeiseBG	Meise	7	7	6	2	3	2
HNHM	Budapest	1	1	0	0	0	0
NRM	Stockholm	4	4	4	0	0	0
	Nb total trainings	**83**	**70**	**52**	**23**	**21**	**6**
	Percentage of match	**100%**	**84%**	**63%**	**28%**	**25%**	**7%**

Appendix B: List of the partner feedback from survey

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Compilation and analysis of partner responses to the questions:

How do you feel about matching existing training activities with the DiSSCo IRLs?
For which IRLs do you find most difficulties matching training activities, if any?

'No particular difficulty in matching the listed training courses to DiSSCo IRLs' training objectives. The proposed categorisation was clear enough to allow correspondences to be identified. However, this categorisation remains rather broad and does not allow for precise matches.
NA2.3 Partners encountered some difficulties in matching training objectives when the content of the courses was not detailed enough (i.e., only training titles were accessible thus giving poor view of training content).
It would have been pertinent to not only focus on DiSSCo IRLs, but also on training areas offered under other infrastructures or organisations (such as Universities or GBIF).
The training objectives the most difficult for partners to match were the one classified under TECHNOLOGICAL, ORGANISATIONAL and FINANCIAL IRLs.

Compilation and analysis of partner responses to the question:

Do you think your institution will be able to provide additional training activities in the future (linked or not to IRLs)?

Most partners report that they have the expertise and skills to align with DiSSCo's training objectives. However, it is noted that there is a lack of funding, availability of staff, and equipment (e.g., dedicated workstations), as well as sometimes a lack of dedicated space (e.g., training room) to develop new training courses.
Many institutions offer a catalogue of training courses that cannot currently be integrated into the DiSSCo catalogue (in-house courses, university courses, national education policies, citizen sciences, etc.) due to the above-mentioned factors, and also because they do not have the administrative (e.g., workflow for managing registration and associated costs) and/or regulatory (e.g., application of the national education tariff) support to expand the target audience internationally.
Some institutions could increase their training audience by adapting the existing one. Some institutions are already developing new training courses, notably through the development of virtual support.

Compilation and analysis of partner responses to « Other comments »:

Enhanced training provision

In order to expand the training catalogue and fill identified gaps, it was suggested that the survey could also include other training sources such as the GBIF, EMMRI, WIPO catalogues.
In order to align with changing needs, and to gain training experience from each other and reinforce training complementarities among institutions providing similar training topics, it would be beneficial to promote the development of trainer exchange programmes among partner institutions. This could be beneficial in building and strengthening a European training strategy in the DiSSCo community.
In order to support partners in increasing their training catalogue and their international outreach, it would be useful to set up translation/interpretation services by the DiSSCo RI.

Improving access to training opportunities

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

It would be helpful to set up training grants programmes to provide funds for travel and accommodation expenses of both trainers and trainees.
This would enhance the value of the training offer and facilitate the transmission of knowledge on a European and international scale, thus strengthening the skills of the scientific community and reinforcing the challenges of the DiSSCo RI.

Development of virtual training provision

In the context of the challenges of opening up exchanges and transferring skills on an international scale and in the light of the health situation limiting the physical movement of people, the importance of developing virtual access to training (digital training platform, virtual workshops, etc.) was highlighted.
It was also added that some institutions already benefit from dissemination channels (e.g., online video-sharing platform) which constitute good educational resources (vulgarisation video, courses, conferences) with a wide audience (general public, specialised public).
It would also be useful to enhance the existing offer. A tool shared by the DiSSCo community for putting educational videos online would make it possible to centralise the offer of each member of the IR and make it accessible to the greatest number.

Promoting new career paths

Comments of IT developers from DiSSCo highlighted the need to promote and develop IT specific trainings, in data carpentry and data architecture adapted to bio- and geo- diversity sciences, and training activities in big-data management.

SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.