Biodiv. Res. Conserv. 65: 23-34, 2022

BRC
www.brc.amu.edu.pl

DOI 10.2478/biorc-2022-0004 Submitted 01.02.2022, Accepted 31.03.2022

Digitization of and online access to data from the natural history collections of Adam Mickiewicz University in Poznań: Assumptions and implementation of the AMUNATCOLL project

Bogdan Jackowiak[1,2]*, Jerzy Błoszyk[3,4], Magdalena Dylewska[5], Maciej M. Nowak[1,6], Piotr Szkudlarz[1], Marcin Lawenda[7] & Norbert Meyer[7]

[1]Department of Systematic and Environmental Botany, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; ORCID: BJ https://orcid.org/0000-0003-1684-7380; MMN https://orcid.org/0000-0003-0005-196X; PSz https://orcid.org/0000-0002-4915-2721
[2]Biodiversity Digitization Laboratory, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
[3]Department of General Zoology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; ORCID: JB https://orcid.org/0000-0002-3615-2263
[4]Natural History Collections, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland;
[5]Project Support Centre, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
[6]Biological Spatial Information Laboratory, Faculty of Biology of Adam Mickiewicz University in Poznań, Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
[7]Poznan Supercomputing and Networking Center, Jana Pawła II 10, 61-139 Poznań, Poland; ORCID: ML https://orcid.org/0000-0003-48443655; NM https://orcid.org/0000-0003-4020-5329

* corresponding author (e-mail: bogjack@amu.edu.pl)

Abstract: This paper describes a project aimed at digitizing and openly sharing the natural history collections (AMUNATCOLL) of the Faculty of Biology at Adam Mickiewicz University in Poznań (Poland). The result of this project is a database (including 2.2 million records) of plant, fungal and animal specimens, which is available online via the AMUNATCOLL portal and on the Global Biodiversity Information Facility website. This article presents selected aspects of the "life cycle" of this project, with a particular focus on its preparatory phase.

Key words: biodiversity collections, data analysis and repositories, digitization project, open access to data, open science cloud, project workflow

1. Introduction

Biological specimens are widely known and long-established in systematic and taxonomic studies (Lane 1996), and they provide a basis for the repeatability or reinterpretation of the preserved aspects of various phenotypes. Each specimen constitutes evidence of biodiversity and a documentation of a particular species that is present in a particular place at a particular time (Hilton *et al*. 2021). Natural History Collections (NHCs) are a part of the foundation of global biological sciences research and education infrastructure. Biodiversity collections are more than shelves of preserved specimens; they also contain tissue and gene samples, parasites that are associated with a specimen and environmental data captured during specimen collection, including audio and video recordings and field notes (Gropp 2018). Collectively, these specimens and associated data facilitate disciplinary and interdisciplinary research in fields such as comparative biology, evolution, ecology, biogeography, public health, conservation biology, environmental change and resource stewardship (Krishtalka & Humphrey 2000; Hebert *et al*. 2003; Suarez & Tsutsui 2004; Hebert & Gregory 2005; Savolainen *et al.* 2005;

Wandeler *et al.* 2007; Borisenko *et al.* 2009; Castillo-Figueroa 2018; Gropp 2018; Pearson & Mast 2019; Yap *et al*. 2022). As such, these specimens can provide key evidence for biodiversity and conservation initiatives (Hilton *et al*. 2021).

Hundreds of millions of plant, fungal and animal specimens that have been deposited in natural history museums have the potential to transform research on global biological change. Due to financial limitations and a lack of understanding of the value of biodiversity collections, these collections remain underused, given their capacity to reveal patterns that are not observable based on other data sources (Meineke *et al*. 2018). However, advances in information technology and an increased willingness to share primary biodiversity data are currently enabling unprecedented access to this information. Combining the presence of the species data with electronic cartography using a variety of algorithms, the task of estimating the niches occupied by species and their areas of distribution becomes feasible at resolutions one to three orders of magnitude higher than was possible only a few years ago. To make it possible to employ this method, limitations such as the lack of high-quality taxonomic determinations, precise geotagging of data and the availability of high-quality and up-to-date taxonomic treatments of the groups must be overcome (Soberón & Peterson 2004). The scientific effectiveness of the digitization of NHCs therefore depends on the participation of taxonomy researchers, geoinformation scientists and computer scientists in the digitization process.

Limited funding, a declining number of taxonomists and a lack of specialists with skills in digitization impede the development of biodiversity informatics based on NHCs. This claim is particularly true of the thousands of medium and small, yet still very valuable, collections that are spread throughout many regions of the world. Many such collections are considering a transition from analogue specimen data to a digital system, but they are limited by the lack of any descriptions of such a project that have previously been implemented in other science and museum centres. Until recently, the Faculty of Biology at Adam Mickiewicz University in Poznań (FBAMU), which houses very valuable collections of plants, fungi and animals, also faced such problems (Jackowiak *et al*. 2022a). However, activities undertaken in 2015 resulted in the acquisition of significant funding for a project entitled the AMU Nature Collections – Online (AMUNATCOLL), which involved digitizing and sharing the nature data resources of the FBAMU. Separate articles included in this BRC volume present the main concepts and components of the AMUNATCOLL IT system: the metadata structure, the database and the operational procedures (Lawenda *et al*. 2022), the portal and the mobile application and the graphical interface (Nowak *et al*. 2022).

This paper aims to provide an outline of the assumptions, workflow and main results of the AMUNATCOLL project, with a focus on the substantive and organizational problems associated with the process of digitizing NHCs.

2. Project background

The FBAMU preserves NHCs that have significant scientific and didactic value, including unique specimens from many regions of the world. Samples of plants, fungi and animals document studies that have been conducted to investigate the flora, funga and fauna of Poland, thus serving as a foundation for the assessment of regional biodiversity. For several decades, these samples have been dispersed among various research laboratories (Jackowiak *et al*. 2022a). In 2004, a museum-like unit known as "Natural History Collections" was established in the Faculty of Biology. The large rooms and constantly improving facilities of this unity as well as its team of qualified staff members have made it possible to organize and preserve these collections as well as to make them available for stationary scientific research (Błoszyk & Konwerski 2017). The fact that these specimen data were presented in analogue form, however, limited the widespread use of these NHCs and the advancement of biodiversity research. The digitization of these NHCs has become another challenge for the Faculty.

On 8 January 2014, the Polish government implemented the Operational Programme Digital Poland (OPDP 2022). The preeminent objective of the OPDP, which appeared in the European Union's 2014-2020 financial outlook, is to take advantage of the potential of digitization to improve the quality of life in society. The budget of the OPDP was EUR 2,255.6 million and included three priority areas of focus: (1) universal access to high-speed internet, (2) e-government and open government, and (3) e-integration of digitally excluded groups and dissemination of information and communication technologies (ICT). The OPDP assumed that the beneficiaries of these funds, which would be awarded in a competitive manner, could include, among others, scientific institutions.

Action was taken based on the assumption that an opportunity to develop an IT system to collect, analyse and openly share digitized data regarding the natural specimens in the collections had finally emerged. The process associated with this project, from the emergence of the idea to its materialization, involved four key stages: initiation, project preparation, project implementation and system functioning.

3. Project initiation

In 2015, the dean of the FBAMU appointed an initiative team (ITeam), which, over a short period of time, identified the need for the digitization of the NHCs more precisely, determined that it was in line with the university's development strategy and assessed the organizational and financial capacities of the Faculty. The ITeam noted that the OPDP offers potential opportunities for the development of nature collections based on modern IT technologies but stipulates serious requirements for the beneficiaries even at the stage of the funding application. These demands range beyond the standards of projects associated with classical biological research; hence, an interdisciplinary approach and the participation of project management specialists in the proposal preparation and implementation phases were necessary.

4. Preparing the project

In the same year, a team was established to prepare a feasibility study and to submit an application to a funding institution for this type of project. This work, which took several months to complete, focused on the following tasks: (1) identification of the resources based on an inventory of the natural collections; (2) identification of the target group and recognition of the expectations of potential users of the information system to be designed; (3) planning the digitization process; (4) selection of an institution to take responsibility for the IT and technological aspects of the project; and (5) conceptual development of an IT system to ensure the implementation of the presumed objectives.

4.1. Identification of resources based on an inventory of the natural collections

The inventory of the NHCs proved to be a very intensive process, despite the fact that these collections had been deposited in a common space since 2004 and systematically catalogued for an analogue system featuring a local database. The task of producing an inventory of natural specimens is laborious because it must reflect not only the richness and diversity of the organisms but also the manner in which they are collected, described, stored and preserved. It was agreed that 54 collections of plants, fungi and animals as well as 12 specialist and iconographic collections would be digitized. It was estimated that these collections would result in a digital database containing approximately two million unique records that would reflect the biodiversity of many places worldwide. A good specimen inventory is crucial to properly plan for the digitization process and to identify the opportunities for data use. During the initial phase of the project, it was already recognized that the task of digitizing these NHCs offered an opportunity to make these resources widely available. It was proposed that the new form of these resources should go beyond their strictly scientific functions, which had previously been the main focus of the collections. This expansion was also in line with the expectations of the potential funder, who was willing to co-finance information systems that would offer public sector information (PSI), as defined by the European Parliament and Council (EU) Directive and the Act of the Parliament of the Republic of Poland (DAPRP 2022).

4.2. Identification of the target group and recognition of the expectations of potential users of the information system to be designed

Biodiversity has been a fundamental subject of scientific research since the term was first defined (Arthur 1916). During the second half of the twentieth century, due to constantly increasing anthropogenic threats, biodiversity became an important topic of discussion in increasingly broad social, economic and political contexts (CBD 1992; MEA 2005).

Assuming that the digital NHC database has an important role to play in this respect, an analysis of its potential users was conducted. As a result, five core target groups were defined, the requirements of these groups were identified and their numbers were estimated.

The first group consists of scientists and PhD students from 43 public research institutions, including 24 universities and 19 research institutes associated with the Polish Academy of Sciences. These individuals are key recipients of the project outcomes. These people will use digital data primarily for the purposes of research and teaching. Some of these people will contribute new scientific information to the database and correct the biodiversity knowledge resources that have been made available. The estimation of the first group was difficult because the number of professionals who use such databases worldwide is unknown. It is difficult to estimate the number of institutions in which they are employed even approximately.

The second group is composed of representatives of nearly three thousand state and local government units at various levels. In Poland, this group includes employees from 23 national parks, 16 landscape parks, 16 provincial offices, 16 marshals' offices, 314 district authority offices and 2,478 city and municipal offices. The representatives of these institutions can use digital biodiversity data to facilitate spatial planning and the development of procedures to support important administrative decisions (e.g., decisions regarding the environmental conditions of investment projects).

The third target group includes state services and officers, i.e., representatives of nearly 720 institutions that use biodiversity data to facilitate their statutory activities. In Poland, this group includes institutions such as the Regional Directorates for Environmental Protection (16), the Provincial Inspectorates for Environmental Protection (16), the Customs Chambers (16), the Regional Directorates of State Forests (16), the Forest Districts (430), the Promotional Forest Complexes (23), the Forest Management and Geodesy Offices and the Forest Gene Bank.

The fourth group comprises educational institutions of all levels and specializations, of which there are approximately 21,500 in Poland. Analysing the needs of teachers and students in terms of access to biodiversity knowledge, it was concluded that the sharing of information resources related to the natural collections and tools for their analysis should have a positive impact on their awareness and activity in the context of nature conservation.

The fifth group includes NGOs and all members of the public with an interest in biodiversity and natural history museums. The numbers of this group are the most difficult to estimate, but this group is also potentially the largest and most diverse in terms of requirements for the database.

The conclusions that can be drawn from this identification of the NHCs and the target groups were important for the digitization process, the construction of the database and the organization of the information system.

4.3. Planning for the digitization of the collection

The digitization of NHCs involves converting specimen data from analogue form to digital signals (Nelson *et al.* 2018). The following digitization techniques are planned to be used for the AMUNATCOLL project: (1) transcribing the text data from sample labels and other sample-related documents into digital records of those labels and documents; (2) translating the physical specimens into digital, two-dimensional images of these specimens (taking images from light microscopes, stereo microscopes, scanning microscopes, specialized scanners and CANON full-matrix cameras); (3) converting analogue photos as well as audio and video recordings to digital recordings; (4) converting textual descriptions of locations into geographical coordinates and creating a digital layer of points to represent specimens; and (5) converting biodiversity data from other sources (e.g., literature, field observations) into a digital format using the same techniques.

It was already necessary to answer several key questions regarding the digitization process during the planning phase. First, who will be involved in the process? Second, what equipment will be available? Third, how much time is needed to digitize specimens with such diverse characteristics? Indeed, without a good understanding of these aspects, it would have been difficult to anticipate the timetable of activities and the costs of the project.

Therefore, it took a great deal of time to test the speed of digitization, taking into account the various parameters of the specimens, equipment and people involved in this work. However, these tests also highlighted the importance of the organization and logistics of a process that seemingly consisted uncomplicated operations. This approach made it easier to discuss the concept of the information system and to ensure that the digital data were collected, analysed and made available to all interested users.

4.4. Development of an IT system concept and selection of a partner IT institution

Many systems on the internet are related to the presentation of biodiversity information. They differ in terms of their purposes and in the scope of functionality that they offer to their users. As a part of the work involved in determining the functional scope of the implemented interfaces, we investigated the experiences of renowned scientific institutions and IT databases, with a particular emphasis on online databases, which are among the tools most frequently employed by scientific users. Detailed results of a comparative analysis of the most important websites and portals of this type are presented in a separate article (Nowak *et al.* 2022). Although the conclusions of this analysis were very informative, the decision was made to proceed with the development of our own IT system. The following arguments were made in support of this solution: (1) none of the operators of which we were aware offered such systems to other users; (2) the excellent Global Biodiversity Information Facility platform makes itself available in terms of space and functionality, but using these resources requires the preparation of advanced IT tools to export the appropriately prepared data; (3) the identification of the target groups indicated that their expectations ranged beyond the functions that were available on the and portals we analysed.

Developing our own information system for the biodiversity documented in the NHCs required the cooperation of biologists and computer scientists specialized in this field. A long-term collaboration between the institution that was responsible for the NHCs and an institution with the appropriate information technology capacity was thus necessary. In the AMUNATCOLL project, the relevant decisions had already been made at the stage when the concept of the future IT system was outlined. Representatives of the Faculty of Biology at Adam Mickiewicz University in Poznań and the Poznan Supercomputing and Networking Centre were included in a team of several people to prepare the project documentation. In addition, a specialist in IT project design was invited to join this team.

4.5. Feasibility study and funding application for project implementation

The feasibility study for the AMUNATCOLL project was prepared between March and November 2017. This document responded to the needs of Adam Mickiewicz University in Poznań, in particular those of the Faculty of Biology, with respect to the digitization of its NHCs. The scope of the study was in line with the requirements for applying for grants from the European Regional Development Fund under the Digital Poland Operational Programme 2014-2020: Priority Axis II – E-administration and open government; Activity 2.3 – Digital accessibility and usability of public sector information; Sub-activity 2.3.1 – Digital accessibility of public sector information from administrative sources and scientific resources.

In accordance with the Intermediate Institution's guidelines, the structure of the Feasibility Study comprises three main parts. Chapters 1-3 present a general description of the project and a summary of the results of the feasibility studies that were conducted. Chapters 4-8 contain detailed analyses of the project; intervention logic; organizational, legal and technical analyses; and the plan for project implementation and operation. Chapters 9-13 address financial, economic, sensitivity and risk analyses.

The feasibility study served as a foundation for the project application, which was prepared in the form stipulated by the Intermediate Institution. Both documents were consistent regarding every aspect of the project.

4.5.1. Project goals, objectives and timetable

The strategic goal of the AMUNATCOLL project was to make the digital data associated with the NHCs of the FBAMU completely open and available online. The following specific objectives were prerequisites for the achievement of this goal: (1) the development of a standardized set of metadata to record the unique values of the FBAMU's NHCs in a manner that would be compatible with international standards; (2) the development of a digital database for the NHCs of the FBAMU; (3) the creation of an information system to ensure interoperability, communication with other databases, the possibility of developing applications and facilitating multidimensional data analysis, including the spatial dimension; (4) the provision of the appropriate hardware and software infrastructure as well as the organizational facilities necessary to guarantee the sustainability and further development of the system within the scope of the resources of the Project Partners; and (5) equipping the project's scientific team with

Table 1. Timetable for key tasks in the AMUNATCOLL project

Tasks	Deadlines for key project tasks								
	Year	1		2				3	
	Month	3	12	22	23	24	35	36	
	Week	12	48	90	94	96	140	143	144
Development and approval of the metadata structure		100%							
Digitization of the botanical and mycological collections (target number of records: 500,000)			30%			70%	100%		
Digitization of the zoological collections (target number of records: 500,000)			30%			70%	100%		
Geotagging of textual descriptions of the specimens (target number of records: 2,000,000)			30%			70%		100%	
Creation of an information system for disseminating and sharing the data: Portal*				100%					
Creation of an information system for disseminating and sharing the data: Mobile application*					100%				
Developing the database repository with the required converters								100%	
Provision of open access to the AMUNATCOLL IT system									100%

* – This task was initiated within 2 weeks of the project's beginning

the competences necessary to digitize other biological resources with the aim of carrying out similar projects using other datasets.

Thirty-six months were allotted to the entire package of goals and objectives. Using the experience developed during the initial phase of the project, a key output indicator of 2 million records, i.e., digitized specimens of algae, plants, fungi and animals, was established. The specific output indicator was set at 960 TB. This size was intended to guarantee secure storage of the digitized text data and space-intensive image data.

For tasks that were crucial to the success of the project, strict deadlines for achieving certain milestones were planned (Table 1). It was assumed that quantitative indicators (e.g., the number of digitized objects) would be monitored systematically and reported on a monthly and quarterly frequency.

Two supporting tasks were also defined: (1) the provision of training in the use of the database to potential users and (2) the dissemination of information regarding the project through promotional activities. The first task was aimed at improving the solutions that had already been introduced during the development of the IT system. The second task was intended to increase the number of persons who used the digitized biodiversity data to fulfil official duties or conduct scientific research.

4.5.2. Equipment

This project featured high requirements regarding the standards for the digitization process and the storage of the results of digitization. The digitization process, in addition to its requirement of procedures to allow many people to work smoothly simultaneously, required top-class devices that could cope with such a heavy load. In addition, a key to success was the correct selection of equipment for the scanned materials, i.e., equipment that could ensure precision and display scientifically important features. Taking these assumptions into account, the following equipment was selected:

- Scanner Zeutschel OS 14000 TT – used for scanning herbarium sheets, this machine allows two sheets to be scanned simultaneously and subsequently divided into two files in the graphics program;
- Light microscope OLYMPUS BX 51 and BX 53, DP74 camera and CS software (cellSens Dimension) – this set of hardware and software allows photos to be taken using a light microscope for solid and unstable microscopic slides. The large resolution of this equipment allows objects smaller than 1 mm to be photographed;
- Stereoscopic microscope OLYMPUS SZX16, DP74 camera and CS (cellSens Dimension) software – this equipment allows pictures of preserved or living animals and plant diasporas up to 3 cm. in size to be taken. Four positions were used mainly for photographing spiders, molluscs, crustaceans, twigs and insects. The available magnification (from 7 to 110 times) allows the details of the morphological structure of arthropods (Arthropoda) and molluscs (Mollusca) to be photographed;
- Scanning microscope Zeiss EVO 40 – this equipment allows pictures of preparations that have been sprayed with gold to be taken. In this way, it is possible to photograph morphological details under high magnification. The Nature Collection has an extensive collection of scanning photos, mainly of mites (Acari);
- CANON full-frame cameras – these cameras were used to take photographs of collected specimens using a shadowless table and lenses of various lengths (ranging from wide angle lenses to telephoto and macro lenses). These cameras are used to photograph large objects (e.g., stuffed birds, fish, corals) and macro objects (e.g., snail shells, insects).

In addition to digitization equipment, project partners also provided numerous ICT systems for the storage and processing of large amounts of material produced by the digitization process. The most important servers are enumerated below:

- Production and development servers – used for hosting the portal and accompanying services;
- Specimen and observation databases – hosted on servers that maintain the databases used by the portal and the mobile application;
- Monitoring server – this server ensures the continuity of the system's operation and the security of the data stored, which are crucial tasks due to the need to access the data and the services offered.

Furthermore, the following data storage systems were established and made available based on the requirements of AMUNATCOLL:

- Data buffer – a system based on Seafile software that allows data to be saved and stored following the digitization process but prior to actual importation to the database (120 TB);
- Disk matrix – space on matrix disks allows source data and processed data to be processed and stored alongside their copies safely manner following importation to the database. This solution is composed of two Ceph-based storages containing 360 TB of space each (local and geographic copy) as well as 120 TB of tape system.

In addition to computer devices, the most significant aspect of the geotagging process was the development

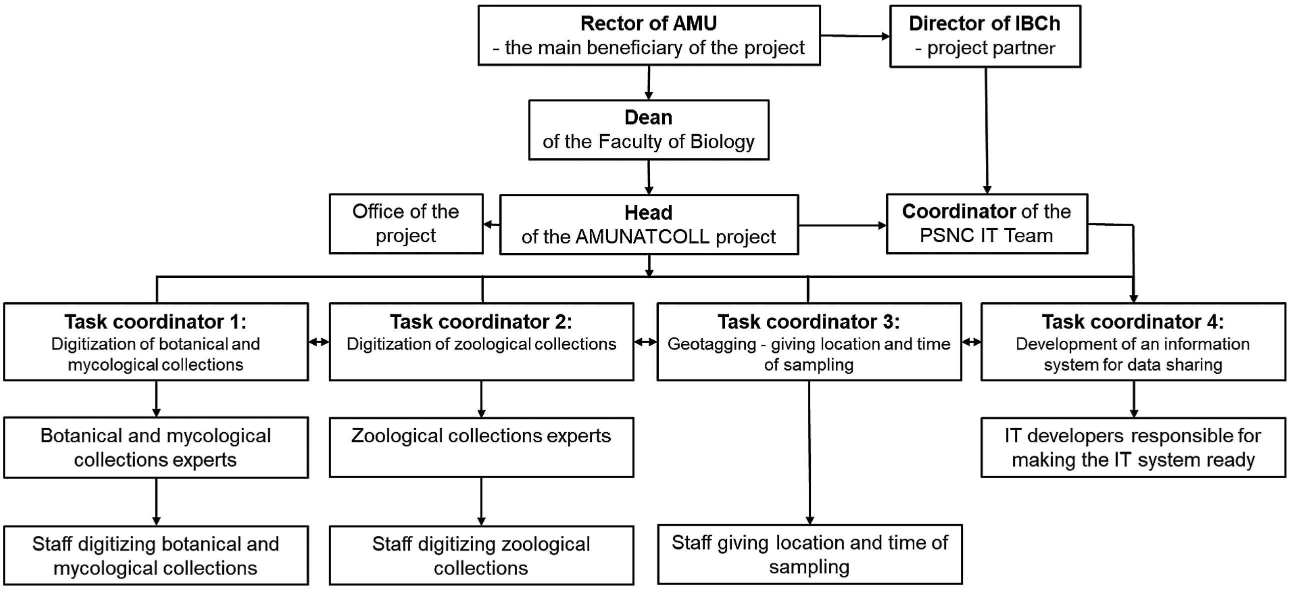

Fig. 1. Functional structure of the contractors who were involved in the AMUNATCOLL project

of a resource based on geographic information used to support this task (Nowak *et al*. 2021).

4.5.3. Staff and management

The organization of a sufficiently large, functional and efficiently managed team of contractors was a prerequisite for achieving the project's goals and satisfying the ambitious qualitative indicators.

The main beneficiary of the project was AMU, whose rector represented the university to the funding institution, made strategic decisions and was formally responsible for the fulfilment of the obligations stipulated in the agreement (Fig. 1). In this regard, the rector cooperated with the director of a partner institution, in this case the Institute of Bioorganic Chemistry of the Polish Academy of Sciences (IBCh), whose structure includes the PSNC. The dean of the FBAMU, which houses the NHCs that were to be digitized, also participated in strategic decision-making.

Nearly 90 people, all of whom had clearly defined functions and tasks, were involved in the AMUNATCOLL project (Figs. 1-2). Sixty-four percent of the staff

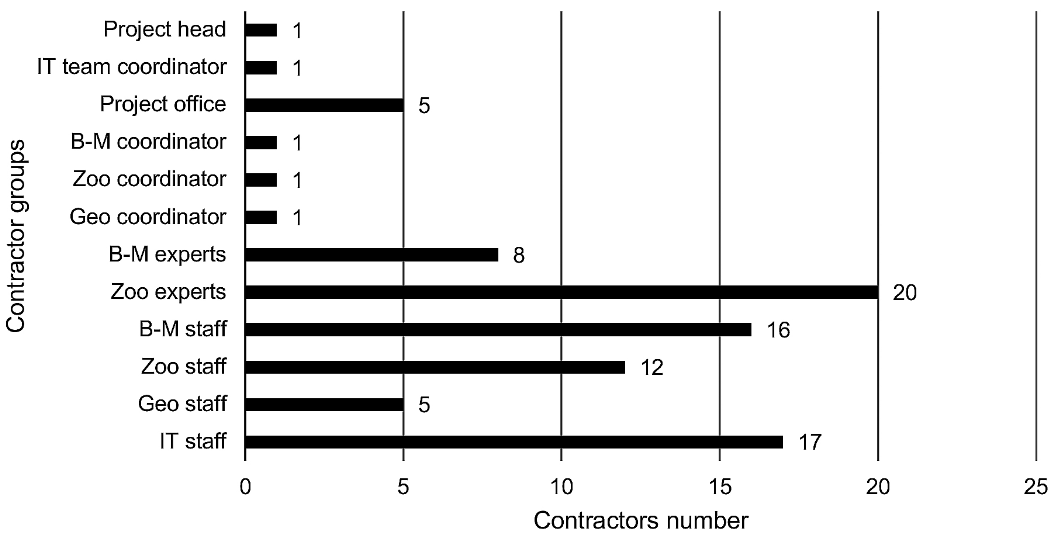

Fig. 2. Proportions of the contractors who were involved in the AMUNATCOLL project
Explanations: B-M – Botanical-Mycological Collections, Zoo – Zoological Collections, Geo – Geotagging

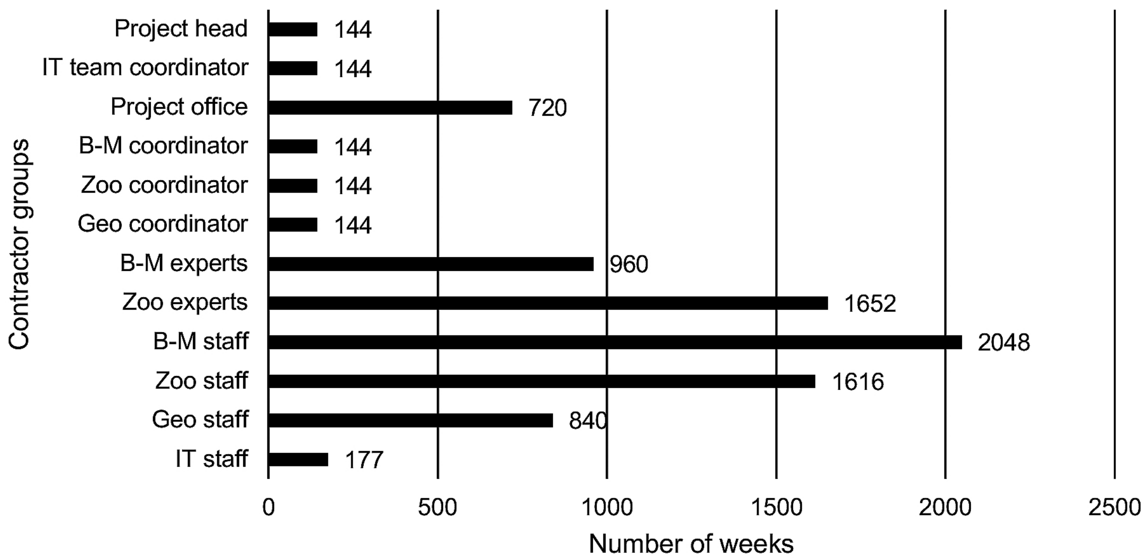

Fig. 3. Length contractor groups' involvement in the AMUNATCOLL project in weeks
Explanations: B-M – Botanical-Mycological Collections, Zoo – Zoological Collections, Geo – Geotagging

were employees and experts who were directly involved in digitizing the collections. The time involvement of this group exceeded 71% of the total (Fig. 3). It was necessary to recruit a larger group of experts for the more diverse zoological collections than for the botanical and mycological collections. The role of these experts was not only to engage in technical activities but also to make substantive assessments of the digitized specimens. An important role in the digitizing process was played by a five-person team that was responsible for supplying the locations and times when the samples were gathered. The coordinators were entrusted with overseeing the work of these three task groups. One of the most important tasks accomplished by these coordinators was also the provision of necessary information to the group of developers to enable them to create an IT system for storing and sharing the data regarding the digitized collections. A group of a dozen IT specialists included developers who were responsible for (1) developing the metadata structure; (2) implementing the database repository with the required converters; (3) preparing the portal and the mobile application; (4) integrating the AMUNATCOLL and GBIF systems; (5) supplying the maps and specimen locations; and (6) conducting internal security audits of the produced software.

The project was supervised by a project lead appointed by the AMU Rector. This individual was primarily responsible for (1) achieving the project's goals and indicators as well as developing and updating any plans for the project, (2) receiving a group of tasks, (3) maintaining the project's documentation, (4) reporting on the project's status and progress; (5) promoting the project and dealing with the media, and (6) organizing legal and administrative consultations. He completed these tasks in cooperation with the coordinator of the IT group and representatives of the PSNC as well as with the assistance of the project office.

5. Project implementation

The implementation of the project consisted of the completion of tasks by a team of contractors with clearly defined roles, who used precisely planned digitization equipment and constructed an information system. During the first few months of the project, the initial three teams focused on preparing the collection and the contractors for digitization as well as purchasing the optical and IT equipment that were necessary for the process to proceed properly (Fig. 4).

The coordinators of these teams, working alongside a group of IT specialists, focused on developing metadata specifications that would comply with international standards while taking into account the specifics of the project based on the expectations of the target groups. The steps that were necessary to achieve full readiness for digitization included (1) agreeing on the rules for transferring the data to the geotagging team and (2) preparing the IT tools to support the process and enable the periodic storage of the data. A key element of the digitization process is the validation of and development of a quality control system for the work of the participants throughout the process.

Once the appropriate level and pace of digitization was achieved, it was possible to direct a team

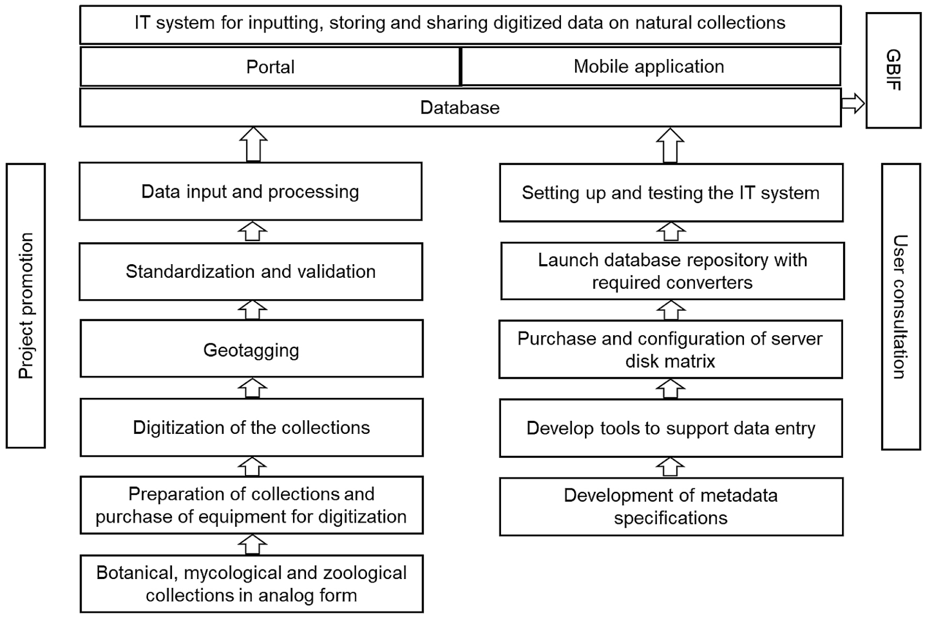

Fig. 4. Diagram of key activities in the AMUNATCOLL project

of IT specialists and coordinators to work on the database structure and the basic interfaces, namely, the portal and the mobile application. Based on the general assumptions made during the project preparation stage, solutions that are currently available to all interested parties were developed. It should be noted that future users, who represent different target groups, had a significant influence on the final design of the portal and the mobile application. These individuals participated in specially organized presentations and workshops to test new versions of the tools that were to be implemented in the AMUNATCOLL IT system. These meetings also provided an opportunity to promote the project.

In the final phase of the project, crucial tasks for project success were completed: (1) the database repository with the required converters was developed; (2) the data and the analytical tools were made openly available on the project's portal; (3) users were provided with access to the mobile application; and (4) the database was made available via the GBIF website.

6. Functioning of the AMUNATCOLL IT system

The result of the project is a functioning AMUNATCOLL IT system that consists of a database containing more than 2.2 million records and interfaces that make those records available online. The digital data on the NHCs, which are deposited at the FBAMU, are openly available through the project's portal (Amunatcoll. Portal 2022), the mobile app (Amunatcoll. Mobile app 2022) and the Global Biodiversity Information Facility website (Amunatcoll at Gbif 2022) (Fig. 5). The portal provides a rich set of tools for data analysis and results presentation in both graphical and cartographic form. The mobile application is designed to facilitate field studies of flora, fauna and natural habitats (Nowak *et al.* 2022).

The AMUNATCOLL IT system operates on the basis of the SEPP (Scientific, Educational, Public, and Practical use model) model that was developed as part of this project (Jackowiak *et al.* 2022b). Based on the experience gained while completing the project, new solutions for geotagging were proposed (Nowak *et al.* 2021). The scientific benefits of these solutions are illustrated by the visitation and download data that are available through the GBIF. During the first few months after project completion, the AMUNATCOLL database was visited more than 9000 times, and the data thus obtained were cited approximately 20 times.

The AMUNATCOLL IT system was also tested as part of a citizen science project aimed at documenting an invasive species (*Ambrosia artemisiifolia*) in the Poznań area (Grewling *et al.* 2022). The portal and the mobile app have already been used several times in the context of academic classes in Poland, France, the Czech Republic and Madagascar (Nowak M. M., oral information).

On the basis of such experience and self-tests, adjustments are currently being made to the database, and some IT solutions are being modified in terms of both software and hardware.

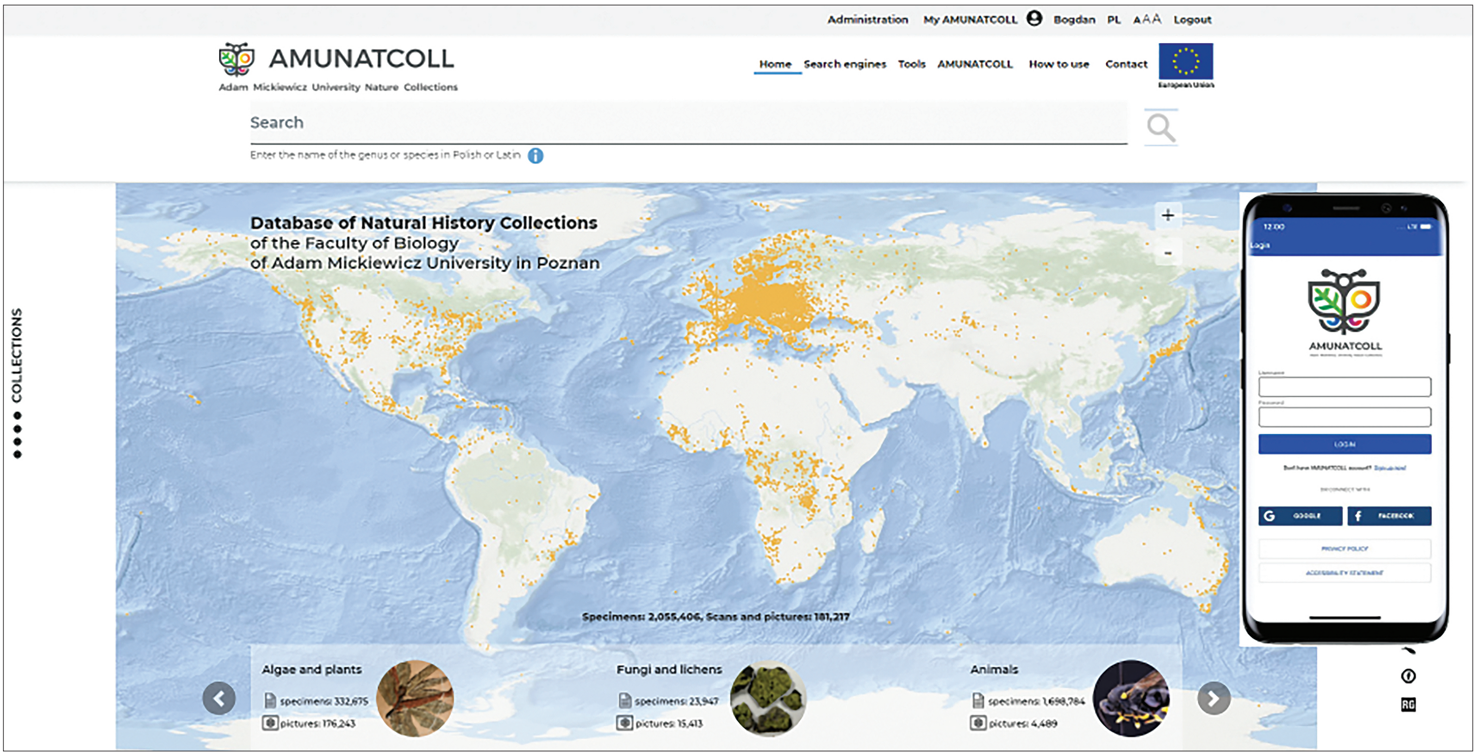

Fig. 5. AMUNATCOLL portal and mobile application are available to all interested users

7. Conclusions

AMUNATCOLL is an interdisciplinary project that requires close cooperation between teams working in the fields of biology and computer science. As in every project in the context of IT, work on AMUNATCOLL began with careful analysis, design and implementation of all the most important modules and functions of the ICT system, which were anticipated in the funding application and the feasibility study.

First, work on the AMUNATCOLL project included the creation of a set of metadata pertaining to various aspects of the correct taxonomic description of the specimens as well as the requirements of the target groups. On this basis, the structure of the database was designed to facilitate the storage of taxonomic, iconographic, organizational and technical information. The process of data preparation and importation was supported by numerous application and logistics solutions, which allowed us to streamline the entire process to improve our ability to process a large number of records over a relatively short period of time.

Simultaneously, the project developed end applications for users: a portal and a mobile application. In addition to the exploration of specimens, the portal system thus implemented facilitates data analysis and spatial visualization. The mobile application that was developed can increase the involvement of target groups by offering them the opportunity to conduct their own observations (by collecting data) and create private collections.

In the context of planning for project tasks related to IT systems, the security of accessing, storing and sharing data cannot be forgotten. This task, which was a constant background featuring of this project, involved preparing, making available and maintaining the IT infrastructure at a level that would allow all of the tasks mentioned above to be carried out in an undisturbed manner.

After the initial version of the system (i.e., in this context, the portal and the mobile application) has been prepared, target groups should be consulted, with the aim of collecting observations and recommendations that can be introduced successively to new versions of the portal. It is also extremely important to use the final stage of the project to improve the quality of the software in terms of functionality and security. This task can be accomplished through numerous tests and audits of the software conducted in laboratory conditions by experts in the fields of user interface analysis and application security.

The data repository provided by the project is a good example of the EOSC (European Open Science Cloud) approach to provide an open access not only to publications and project results but especially to raw data. The repository of AMUNATCOLL is available in the National Data Storage infrastructure and gives an opportunity for scientists to make further analysis on top of the data base.

Acknowledgements. This work has been supported by the AMUNATCOLL project and has been partly funded by the European Union and Ministry of Digital Affairs from the European Regional Development Fund as part of the Digital Poland Operational Program under grant agreement number: POPC.02.03.01-00-0043/18. This paper expresses the opinions of the authors and not necessarily those of the European Commission and Ministry of Digital Affairs. The European Commission and Ministry of Digital Affairs are not liable for any use that may be made of the information contained in this paper.

Author Contributions:
Research concept and design: B. Jackowiak, J. Błoszyk, M. Dylewska, M. M. Nowak, P. Szkudlarz, M. Lawenda, N. Meyer
Collection and/or assembly of data: B. Jackowiak, J. Błoszyk, M. M. Nowak, P. Szkudlarz, M. Lawenda
Data analysis and interpretation: B. Jackowiak, M. M. Nowak, P. Szkudlarz, M. Lawenda
Writing the article: B. Jackowiak
Critical revision of the article: B. Jackowiak, M. M. Nowak, P. Szkudlarz, M. Lawenda
Final approval of article: B. Jackowiak, M. M. Nowak, P. Szkudlarz, M. Lawenda, N. Meyer

References

Amunatcoll. Portal 2022. Available online: https://amunatcoll.pl/

Amunatcoll. Mobile app 2022. Available online: https://apps.apple.com/pl/app/amunatcoll/id1523442673 and AMUnatcoll https://play.google.com/store/apps/details?id=pl.pcss.amunatcoll.mobile

Amunatcoll at Gbif 2022. The occurrence of Natural History Collections of the Faculty of Biology AMU at GBIF. Available online: https://www.gbif.org/occurrence/search?hosting_organization_key=5480dea7-2a71-409b-a832-cbc5f1b5a2e6

Arthur H. J. 1916. The Variable Desert. The Scientific Monthly 3(1): 41-50.

Błoszyk J. & Konwerski S. 2017. Zbiory Przyrodnicze Wydziału Biologii UAM kontynuatorem 160-letnich tradycji muzealnictwa przyrodniczego w Wielkopolsce. Ośrodek Kultury Leśnej w Gołuchowie: 128-139.

Borisenko A. V., Sones J. E. & Hebert P. D. N. 2009.The front-end logistics of DNA barcoding: Challenges and prospects. Mol Ecol Resour 9: 27-34.

Castillo-Figueroa D. 2018. Beyond specimens: linking biological collections, functional ecology and biodiversity conservation. Revista peruana de biología 25(3): 343-348.

CBD 1992. https://www.cbd.int/

DAPRP 2022. https://www.temidium.pl/artykul/nowa_ustawa_o_ponownym_wykorzystywaniu_informacji_sektora_publicznego-6895.html#sdfootnote1anc.

Grewling Ł., Szymańska A., Kostecki Ł., Nowak M., Gwiazdowska J., Morawiec-Kowalska E., Nowak M. & Jackowiak B. 2022. What's new about ragweed in Poznań, Poland? IRS 2022 International Ragweed Society Conference, September 8-10, 2022, Budapest, Hungary. http://internationalragweedsociety.org/irs2022budapest/.

Gropp R. E. 2018. Specimens, Collections, and Tools for Future Biodiversity-Related Research. *BioScience* 68(1): 3-4.

Hebert P. D. N., Cywinska A., Ball S. L., deWaard J. R. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B, 270: 313-321.

Hebert P. D. N. & Gregory T. R. 2005. The promise of DNA barcoding for taxonomy. Syst Biol 54: 852-859.

Hilton E. J., Watkins-Colwell G. J. & Huber S. K. 2021. The Expanding Role of Natural History Collections. Ichthyology & Herpetology 109(2): 379-391.

Jackowiak B., Błoszyk J., Celka Z., Konwerski Sz., Szkudlarz P. & Wiland-Szymańska J. 2022a. The natural history collections of Adam Mickiewicz University in Poznań (Poland): an outline of their history and content. Biodiv. Res. Conserv. 65: 1-22.

Jackowiak B., Lawenda M., Nowak M. M., Wolniewicz P., Błoszyk, J., Urbaniak M., Szkudlarz P., Jędrasiak D., Wiland-Szymańska J., Bajaczyk R. & Meyer N. 2022b. Open Access to the Digital Biodiversity Database: A Comprehensive Functional Model of the Natural History Collections. Diversity 14: 596.

Krishtalka L. & Humphrey P. S. 2000. Can natural history museums capture the future? BioScience 50: 611-617.

Lane M. A. 1996. Roles of Natural History Collections. Annals of the Missouri Botanical Garden. 83: 536-545.

Lawenda M., Wiland-Szymańska J., Nowak M. M., Jędrasiak D. & Jackowiak B. 2022. The Adam Mickiewicz University Nature Collections IT system (AMUNATCOLL): metadata structure, database and operational procedures. Biodiv. Res. Conserv. 65: 35-48.

Meineke E. K., Davies T.J., Daru B. H. & Davis C. C. 2018. Biological collections for understanding biodiversity in the Anthropocene. Phil. Trans. R. Soc. B 374: 20170386.

MEA 2005. http://www.millenniumassessment.org/en/index.aspx.

Nelson G. & Ellis S. 2018. The history and impact of digitization and digital data mobilization on biodiversity research. Philos. Trans. R. Soc. B, 374: 20170391.

Nowak M. M, Lawenda M., Wolniewicz P., Urbaniak M. & Jackowiak B. 2022. The Adam Mickiewicz University Nature Collections IT system (AMUNATCOLL): portal, mobile application and graphical interface. Biodiv. Res. Conserv. 65: 49-67.

Nowak M. M., Słupecka K. & Jackowiak B. 2021. Geotagging of natural history collections for reuse in environmental research. Ecol Indic 131: 108131.

OPDP 2022. https://www.polskacyfrowa.gov.pl/strony/o-programie/zasady/dla-kogo-jest-program/.

Pearson K. D. & Mast A. R. 2019. Mobilizing the community of biodiversity specimen collectors to effectively detect and document outliers in the Anthropocene. American Journal of Botany 106(8): 1052-1058.

SavolainenV., Cowan R. S., Vogler A. P., Roderick G. K. & Lane R. 2005. Towards writing the encyclopaedia of life: An introduction to DNA barcoding. Philos. Trans. R. Soc. B 360: 1805-1811.

Soberón J. & Peterson T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Phil. Trans. R. Soc. Lond. B. 359: 689-698.

Suarez A. V. & Tsutsui N. D. 2004. The value of museum collections for research collections. BioScience 54: 66-74.

Wandeler P., Hoeckl P. E. A. & Keller L. F. 2007. Back to the future: Museum specimens in population genetics. Trends Ecol. Evol. 22: 634-642.

Yap C. K., Go R., Bujang J. S., Syazwan W. M., Azrizal-Wahid N., Azmai M. N. A., Annavi G., Nulit R., Ibrahim M. H., Omar H., Mustafa M., Aziz N. A. A. 2022. The role of museum of biological collections in environmental research: a short note. MOJ Eco Environ Sci. 7(2): 53-55.