



Concept Paper

Open Access to the Digital Biodiversity Database: A Comprehensive Functional Model of the Natural History Collections

Bogdan Jackowiak ^{1,*}, Marcin Lawenda ², Maciej M. Nowak ^{1,3}, Paweł Wolniewicz ², Jerzy Błoszyk ^{1,4}, Michał Urbaniak ², Piotr Szkudlarz ¹, Damian Jędrasiak ², Justyna Wiland-Szymańska ¹, Rafał Bajaczyk ⁵ and Norbert Meyer ²

- ¹ Department of Systematic and Environmental Botany, Faculty of Biology, Adam Mickiewicz University in Poznań, ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; maciej.nowak@amu.edu.pl (M.M.N.); jerzy.bloszyk@amu.edu.pl (J.B.); piotr.szkudlarz@amu.edu.pl (P.S.); justyna.wiland-szymanska@amu.edu.pl (J.W.-S.)
- ² Poznań Supercomputing and Networking Center, ul. Jana Pawła II 10, 61-139 Poznań, Poland; lawenda@man.poznan.pl (M.L.); pawelw@man.poznan.pl (P.W.); michalu@man.poznan.pl (M.U.); djedrasiak@man.poznan.pl (D.J.); meyer@man.poznan.pl (N.M.)
- ³ Biological Spatial Information Laboratory, Faculty of Biology, Adam Mickiewicz University in Poznań, ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
- ⁴ Natural History Collections, Faculty of Biology, Adam Mickiewicz University in Poznań, ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland
- ⁵ Faculty of Biology, Adam Mickiewicz University in Poznań, ul. Uniwersytetu Poznańskiego 6, 61-614 Poznań, Poland; rafal.bajaczyk@amu.edu.pl
- * Correspondence: bogjack@amu.edu.pl



Citation: Jackowiak, B.; Lawenda, M.; Nowak, M.M.; Wolniewicz, P.; Błoszyk, J.; Urbaniak, M.; Szkudlarz, P.; Jędrasiak, D.; Wiland-Szymańska, J.; Bajaczyk, R.; et al. Open Access to the Digital Biodiversity Database: A Comprehensive Functional Model of the Natural History Collections. *Diversity* **2022**, *14*, 596. <https://doi.org/10.3390/d14080596>

Academic Editor: Michel Baguette

Received: 5 July 2022

Accepted: 23 July 2022

Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The Natural History Collections of Adam Mickiewicz University (AMUNATCOLL) in Poznań contain over 2.2 million specimens. Until recently, access to the collections was limited to specialists and was challenging because of the analogue data files. Therefore, this paper presents a new approach to data sharing called the Scientific, Educational, Public, and Practical Use (SEPP) Model. Since the stakeholder group is broad, the SEPP Model assumes the following key points: full open access to the digitized collections, the structure of metadata in accordance with certain standards, and a versatile tool set for data mining or statistical and spatial analysis. The SEPP Model was implemented in the AMUNATCOLL IT system, which consists of a web portal equipped with a wide set of explorative functionalities tailored to different user groups: scientists, students, officials, and nature enthusiasts. An integral part of the system is a mobile application designed for field surveys, enabling users to conduct studies comparing their own field data and AMUNATCOLL data. The AMUNATCOLL IT database contains digital data on specimens, biological samples, bibliographic sources, and multimedia nature documents. The metadata structure was developed in accordance with ABCD 2.06 and Darwin Core standards.

Keywords: biodiversity collections; digitization; open-access database; citizen science; citizen education; environmental management

1. Introduction

Natural history collections (NHCs) have always been fundamental to the study of the diversity and variability of organisms [1]. Specimens collected in the field are essential for descriptions of taxa [2,3] and have become reference points for species identification by successive generations of researchers. However, for several decades, the use of collections has not been limited to taxonomic research. Specimens as information resources are being increasingly widely used in biogeographic, environmental, and ecological research [4]. The value of physical biodiversity specimens has increased significantly through the use of

molecular biology techniques to identify taxa and in phylogeographic studies [5–11]. NHCs are currently experiencing a renaissance as a result of the IT revolution, including the development of Geographic Information System (GIS) [12]. This renaissance is expressed, first, in open access to an increasing amount of digital biodiversity data for everyone interested.

The exponential growth of digital data from NHCs is primarily the result of major initiatives such as the United States National Science Foundation's Advancing Digitization of Biodiversity Collections [13], Australia's Atlas of Living Australia [14], Mexico's National Commission for the Knowledge and Use of Biodiversity [15], Brazil's Centro de Referencia em Informacao [16], Europe's emerging Distributed System of Scientific Collections [17], the African Biodiversity Challenge [18], and China's National Specimen Information Infrastructure [19].

The digitization of NHCs contributes to the development of many areas of biodiversity research [20–26]. Digital botanical collections are increasingly being used in phenological research [27–32], studies on species extinction [33–35] and invasion [36], and modelling species distribution [37,38]. Interest in the information archived in NHCs has increased significantly due to threats to biodiversity [39] and global climate change [40–42]. Digital biocollections have proven to be useful in nature conservation [43], and open access to biodiversity databases is conducive to the development of citizen science and natural education [44–46].

On a global scale, this demand is being met by large international initiatives, such as the Global Biodiversity Information Facility (GBIF) [47–50] and Integrated Digitized Biocollections [51]. They bring together hundreds of regional and local collections in their networks, in which digital databases of biodiversity are based on common international standards. Many regions are underrepresented in global biodiversity databases, and their collections are still seeking a place in the digital information flow [52].

With this in mind, we would like to share our experience in developing a digital biodiversity collection aimed at a wide range of users. The functional Scientific, Educational, Public, and Practical Use (SEPP) Model proposed in this paper assumes that digital data on biodiversity will be used for scientific, educational, public, and practical purposes. In this study, we aim to present an original information system on biodiversity that (i) is based on scientific natural collections, (ii) covers a wide range of organisms (algae, plants, fungi, and animals), (iii) takes into account the needs of various user groups, and (iv) is linked to the international GBIF database. This system has been developed and implemented in the NHCs at Adam Mickiewicz University (AMU) in Poznań, Poland.

2. Materials and Methods

In this paper, we share our experience with the development and implementation of the project “AMU Nature Collections—online (AMUNATCOLL): Digitization and sharing of the natural data resource of the Faculty of Biology at Adam Mickiewicz University in Poznań” [53]. The AMUNATCOLL project was cofinanced by the European Union through the European Regional Development Fund under the Operational Program Digital Poland (OP PC), Priority axis: II E-administration and open government; Action: 2.3 Digital accessibility and usability of public sector information; Submeasure: 2.3.1 Digital access to public sector information from administrative sources and science resources.

The idea of digitizing and sharing data on biodiversity was born at the Faculty of Biology of AMU (FBAMU), which for over 100 years has been a regional centre for biodiversity research and a storage site for NHCs. This idea was implemented in cooperation with the Poznań Supercomputing and Networking Center (PSNC), which is affiliated with the Institute of Bioorganic Chemistry of the Polish Academy of Sciences. The PSNC is an internationally known node of European research in the field of scientific information infrastructure and an important research and development centre in the field of information and communication technologies [54].

2.1. Natural History Collections of Adam Mickiewicz University in Poznań

The NHCs were deposited in a museum unit specifically created for this purpose at the FBAMU in 2004 [55]. These collections document the scientific research of several generations of naturalists conducted in many regions of the world. Most are collections gathered by biologists associated with the University of Poznań, which was established in 1919. The oldest specimens date back to the end of the 18th century. The AMU NHCs are characterized by high biodiversity, which corresponds to the diverse methodology of specimen documentation and storage. Therefore, three main groups of collections can be distinguished: botanical, mycological, and zoological.

The botanical collections comprise approximately 500,000 specimens, including over 350,000 vascular plants. Specimens of vascular plants are stored in two herbaria affiliated with Index Herbariorum [56]: POZ and POZG. The POZ herbarium (approx. 190,000 sheets) consists of many collections, mainly from Poland but also from various regions of Europe and North America. This herbarium contains over 240 nomenclatural types of various ranks. The main part of the POZG herbarium is the collection of vascular plants by S. Lisowski (approx. 120,000 sheets), an outstanding expert in the flora of Central Africa. This collection includes over 100 nomenclatural types of different ranks. Bryophytes and algae constitute an important part of the botanical collections. The collection of bryophytes includes approximately 90,000 herbarium bags with mosses (POZG) and over 40,000 bags with liverworts (POZW). The herbarium of algae (POZA) comprises over 10,000 microscopic specimens and dried specimens, including collections of stoneworts by I. Dąmbska. A total of over 20,000 macrofungi (POZM) and 30,000 lichen (POZ) specimens have been deposited in mycological collections.

The zoological collections include over 1,700,000 invertebrate specimen catalogued thus far, as well as 50,000 chordates. Over 80% of the collection is arthropods, the most numerous of which are arachnids and insects. The collection of molluscs is also very rich (over 200,000 specimens). The arachnid phylum, numbering over 660,000 identified specimens, is represented primarily by mites (over 500,000 specimens).

Less numerous but equally interesting are spiders, harvestmen, and pseudoscorpions. In the insect phylum, among more than 500,000 specimens, the most abundant are collections of beetles, wasps, sawflies, bees, ants, and butterflies. The collections of fleas, dragonflies, mayflies, and true bugs are also very interesting. The nomenclatural types (1000 specimens) collected therein prove the high scientific value of the zoological collections.

2.2. Definition and Course of Digitization

The applied digitization procedure is consistent with the broad approach introduced by Nelson and Ellis [52]: “We define digitization as the conversion of specimen data from analogue to digital signals. This includes transcribing text data from specimen labels and other specimen-related documents into digital records of those labels and documents regardless of input mode (e.g., voice, keyboard, scanning/optical character recognition (OCR); the translation of physical specimens to digital images of those specimens, including two-dimensional, three-dimensional (3D), computed tomography (CT), and other digital image types that visually represent the physical specimen; the conversion of analogue audio and video recordings to digital recordings; the conversion of textual location descriptions into digital georeferences within an accepted geographical coordinate system and the conversion of other specimen-related data into digital format with technologies that are or might become available”. Considering the specificity of the AMUNATCOLL project, these activities were oriented around selected elements specified in the above definition, with particular emphasis on label text transcription; document (e.g., herbarium sheets) scanning; photography with the light microscope, stereoscopic microscope, scanning microscope, and full-matrix CANON cameras; and conversion of textual location description into digital georeferences.

The life cycle of the AMUNATCOLL project, which aims to provide access to the digital biodiversity data stored at the FBAMU, has four main phases: initiation, preparation, and implementation of the project and the operation of the IT system (Figure 1). The basic methodological assumptions of the system development life cycle (SDLC) [57,58] are used in the implementation of this project. The AMUNATCOLL project was carried out for 39 months. Over 70 employees participated in the project implementation. Approximately 60 people were specifically involved in digitizing the collections, including 5 who specialized in the process of geotagging NHC data [59].

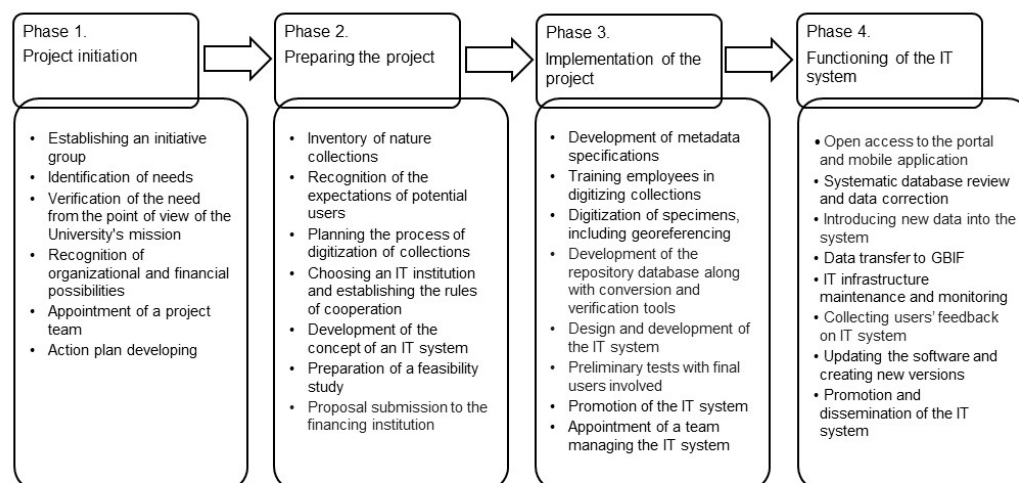


Figure 1. Development of the system for collecting, analysing, and sharing digitized data on biodiversity based on the Natural History Collections of Adam Mickiewicz University in Poznań (AMUNATCOLL IT).

3. Conceptual Model of Digitized Natural History Collections Open to Science, Education, Public and Practical (SEPP) Use

The mission of the AMUNATCOLL IT system is to develop and promote knowledge about biodiversity by collecting and openly sharing digitized data and providing tools for their analysis for research, education, application and social purposes (Figure 2). We assume that the digitized and shared data will be used not only by specialists in the field of life sciences but also by experts in culture, history, and other fields of knowledge. The survey conducted during the project preparation phase showed that not only scientists but also teachers, students, experts in nature conservation and environmental management, officials and specialists in spatial planning, and even legal experts are interested in digitized data on biodiversity. The expectations signalled at that time were taken into account in the conceptual model of the AMUNATCOLL IT system.

The system consists of a digital database and tool set made available through the web portal and a mobile application integrated with the portal. The AMUNATCOLL database is fed by four analogue information sources subject to the digitization process: (i) natural collections, i.e., specimens of algae, plants, fungi, and animals, as well as a collection of soil samples from all over the world, deposited at the FBAMU; (ii) published and unpublished floristic, faunal, and mycobiologic data, mainly from Poland; (iii) nature photography and films documenting scientific research in many regions of the world; and (iv) bibliographic sources cited in the AMUNATCOLL database.

The AMUNATCOLL mobile application allows users to document observations directly in the natural environment and send them to the portal for further analysis. The portal is interactive, enabling users to explore the shared data and provide feedback to the system administrator.

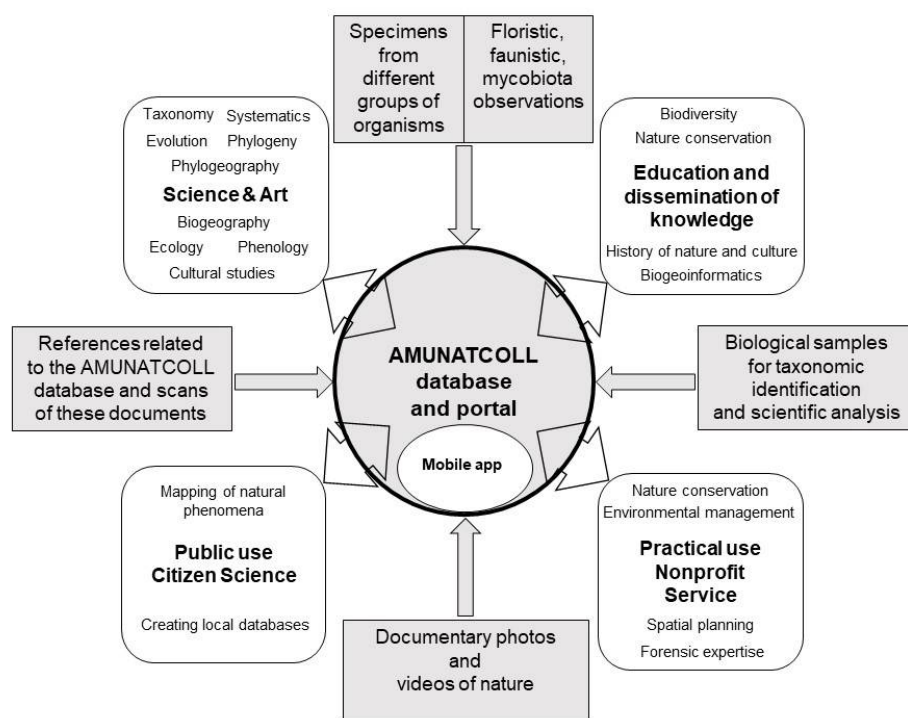


Figure 2. Conceptual model of digitized natural history collections open to science, education, public, and practical (SEPP) use.

Interactivity with biodiversity data is an important feature of the system because the AMUNATCOLL mobile application allows users to document observations directly in the natural environment and send them, including the digitized data, to their portal account for further analysis. Users can also add their own archive location data directly to the map in the portal. Additionally, using the features of creating teams and joining teams, users can share field data and comments.

The specimen data along with their occurrence, fully available at the AMUNATCOLL portal, are successively transferred to the international GBIF database [60,61]. It is worth emphasizing that the AMUNATCOLL IT system is open to the inflow of data from other research centres and individual owners of nature collections.

The presented idea goes towards Digital Extended Specimen (DES) [62] concept in respect of developed functionality as well as following FAIR principles (findability, accessibility, interoperability, and reuse) of specimen metadata. The DES concept was developed through an open community consultation led by GBIF [63] and is a result of aggregation of the open digital specimen concept developed under DiSSCo initiative [64] and the extended specimen concept developed by Lendemer [65] and implemented by Biodiversity Collections Network [66]. Presently, the AMUNATCOLL IT system supports fully or partially the Extended Specimen Network extensions: Primary, Secondary, and Tertiary by enabling access to data such as images, metadata, ecological data, environmental samples, field images, conservation status, and species distribution. Its future development will aim to extend this set at interoperability for other data groups.

4. Collecting and Sharing Biodiversity Data in the AMUNATCOLL IT System

The main path from nature collections and information being deposited in the traditional (analogue) way at the AMU to making them available in a modern and fully open format leads through the following stages: digitization, database design, and portal equipped with search and analysis tools (Figure 3). The AMUNATCOLL database, in a special private area, can also be supplied with digital data sent from a mobile application that has been specially developed for field surveys.

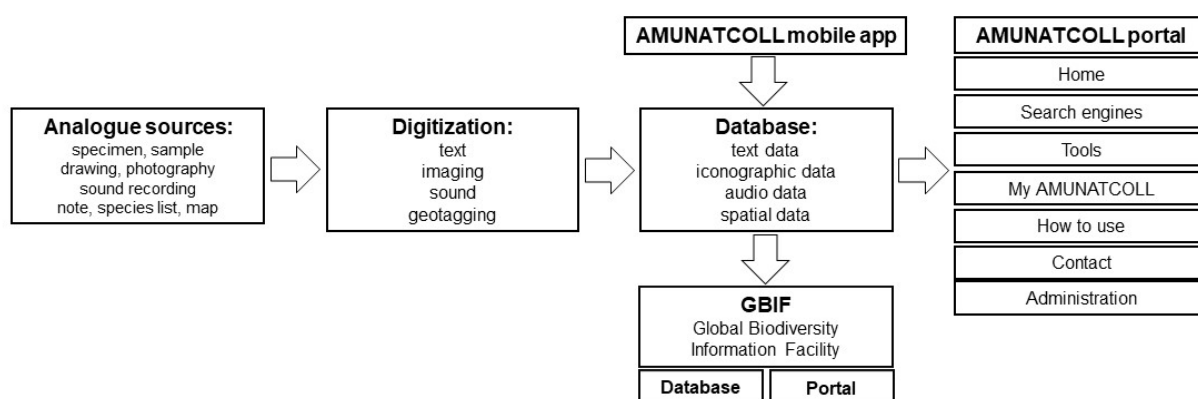


Figure 3. The process of digitizing and sharing biodiversity data in the AMUNATCOLL IT system.

4.1. Metadata Structure and Database

The core of the AMUNATCOLL IT system is a database with content that is based primarily on collections of biological specimens as well as photos and published or previously unpublished field observations (Section 1 in Table 1). These analogue biodiversity documents are described during digitization with metadata that meet ABCD 2.06 standards [67] and correspond to the Simple Darwin Core Categories [68]: Record-level Terms, Taxon, Occurrence, Event, and Location. Depending on the type of document, the number of fields in a single record varies from 68 to 85 (Table 1).

For the purpose of the specimen description, fields for information on the location and method of storage and conservation are also provided. A specific source of information on biodiversity is samples containing taxonomically unidentified biological material (mainly invertebrates). These very valuable samples await scientific analysis, particularly the determination of their species (Section 2 in Table 1). An extensive description of the samples includes, among other things, information about already elaborated systematic groups that allows researchers to focus their analyses on groups that have not yet been identified. The list contains 50 metadata items and is modelled on Section 1. The AMUNATCOLL database also collects multimedia documents of landscapes and habitats, as well as photos of specimens, which, although they do not have all the attributes necessary for inclusion in Section 1, can be used for nonscientific purposes (Section 3). In this case, a single record consists of 51 fields. Section 4 (References) contains 28 metadata describing the references and other source documents cited in the descriptions of specimens, photos, biological samples, and multimedia documents. Two sections (1 and 3) also include six fields for information on the usefulness of digitized objects for specific user groups.

The database implemented in AMUNATCOLL IT provides effective access to all metadata fields as well as organizational, technical, and auxiliary fields. An open-source PostgreSQL server was deployed in the system. It manages relational databases and emphasizes extensibility and SQL compliance. PostgreSQL is designed to handle heavy workloads, such as data warehouses or web services with multiple concurrent users. Furthermore, it is often the primary choice for solutions for biodiversity systems, as is evidenced by examples of institutions such as the Belgian Biodiversity Platform [69] and Island Conservation Society [70] or the TriatoKey application [71]. The database is divided into three logical parts: (i) “amunatcoll”, responsible for storing digitized textual data on specimens; (ii) “dlibra”, which stores metadata on digitized images of specimens and multimedia collected directly in the field; and (iii) “anc_portal”, responsible for the storage of data related to the portal and its functions.

Table 1. AMUNATCOLL metadata structure.

| AMUNATCOLL Metadata Groups (In Brackets a Reference to Darwin Core Categories) | Number of Metadata in Sections of Digitized Objects | | | | |
|---|--|-----------------------|-----------|------------|--------------|
| | Section 1 | | Section 2 | Section 3 | Section 4 |
| | Specimens, Drawings, Photos | Human Observations | Samples | Multimedia | Bibliography |
| Record (Record-level Terms) institution, collection, specimen/object/photo number, file name, mapping type, source material, date of recording the information, surname and first name of the author | 12 | 4 | 1 | 7 | 1 |
| Taxon (Taxon) scientific names of taxa of various ranks and their authors | 11 | 11 | - | 11 | - |
| Description of the specimen and its origin (Occurrence/Event) nomenclatural type, gender, age, stage, dimensions, evidence material, borrowed, collection development status, author of the collection, collection number of the specified author, author and date of identification, date of collection of the specimen, method of collection, habitat, substratum | 21 | 17 | - | - | - |
| The location of the specimen, sample, or other object (Location) location, location above sea level, latitude and longitude, UTM and ATPOL coordinates and comments, protected areas in Poland, continent, state, voivodship, district and commune in Poland, location within the physiographic division of Poland and biogeographical division of Europe, location within the geographical division of the potential natural vegetation of Poland | 26 | 26 | 28 | 26 | - |
| The place and method of the specimen storage and conservation | 8 | 3 | - | - | - |
| Usefulness of specimens/objects for target groups scientists, teachers, students, pupils, state and local government administration, nature protection services, non-governmental organizations, nature lovers | 6 | 6 | - | 6 | - |
| Citing information about a specimen, sample, or other object (Occurrence) | 1 | 1 | 1 | 1 | - |
| References cited in the AMUNATCOLL database | - | - | - | - | 27 |
| Sample description author and date of collection, nature and general description of the sample, collection method, characteristics of merocenosis, sample humidity, verified groups of animals | - | - | 20 | - | - |
| Total: | 85 | 68 | 50 | 51 | 28 |

4.2. Portal

The architecture of the AMUNATCOLL portal [72] is in line with the assumptions of the conceptual model of the IT system (Figure 2). It takes into account both the abundance and complexity of information on biodiversity resources deposited in the database as well as the requirements of potential users. The modular and multilayered structure of the portal offers users quick and easy access to data while enabling the administration panel to manage the system.

The homepage contains basic information about the AMUNATCOLL IT system, including database resources, Geographic Information System for Biodiversity (BioGIS) tools,

the mobile application, potential users, and the contractors and owners of the system (Figure 4).



Figure 4. The homepage of the AMUNATCOLL portal <https://amunatcoll.pl/> (accessed on 20 December 2021).

A much more extensive description, including the mission of the system, can be found in the “AMUNATCOLL” module. However, the most important functions for the user are in the Search Engines, Tools, and My AMUNATCOLL (Figure 5) modules. The system allows users to search for information about specimens, biological samples, multimedia documents, bibliographic sources cited in the database, and educational materials. Specimens and taxa can be viewed using several types of search engine. The general search engine is available to everyone, while the others are available to logged-in users. The three specialized search engines differ in the number of selection criteria and conditions that must be met for the specimens sought. For the simple search engine, a user can choose from one to seven criteria that the specimens must meet (Figure 6a). These are deliberately frequently selected specimen attributes: genus or species name in Latin, superior taxonomic unit, author of the collection, year of collection, country, type of collection, and iconographic documentation. In addition, it is feasible to search for specimens by marking their area of origin on the map. In the case of the extended and advanced search engines, the user can choose from 1 to 37 criteria classified into four groups (Figure 6b,c). Furthermore, it should be mentioned that when many criteria are selected, the possibility of defining logical operators was introduced. In the simple and extended search engines, the “and” condition is used by default. On the other hand, in the advanced search engine, the operators “or” and “no” are also implemented and when used are explicitly mentioned and modified.

Specimens can also be accessed by selecting collections or systematic groups, the lists of which are made available in the panels of relevant search engines. Systematic groups are represented by higher rank taxa, for which information is not included in the metadata specification itself. Data on taxonomically unidentified samples can be obtained using 49 selection criteria classified into four groups (Figure 7a). Eight criteria were entered for the multimedia search, and six criteria were entered for the bibliography (Figure 7b,c). Authorized legitimate users can create educational and training materials

from AMUNATCOLL database documents that are available to everyone and can be downloaded from the successively completed list.

| Home | AMUNATCOLL | Search engines | Tools | My AMUNATCOLL | How to use | Contact | Administration |
|------|--------------------|---|------------------------|-----------------|-----------------|--------------------|-------------------|
| | Mission | Specimen and taxa | Statistical analysis | My observations | Search engines | Contact form | Reports and stats |
| | Portal | General | Reports | My albums | Tools | | Protected taxons |
| | Mobile application | Specialized | Spatial analysis | My maps | My observations | Protected area | |
| | BioGIS | <div>Simple</div> <div>Extended</div> <div>Advanced</div> | Dot distribution map | My teams | | Reports and stats | |
| | Our users | | Area class map | | | Manage users' role | |
| | About us | | Choropleth map | | | Files report | |
| | | Collections | Diagram map | | | Files sources | |
| | | Systematic groups | Attribute grouping map | | | Team leaders | |
| | | Samples | Cluster map | | | Editor tools | |
| | | Multimedia | Timelaps map | | | Excel files tool | |
| | | Bibliography | | | | Task history | |
| | | Education | | | | Database history | |

Figure 5. The architecture of the AMUNATCOLL portal.

Figure 6. View of specialized search engines of the AMUNATCOLL portal: (a) simple search engine, (b) extended search engine—criteria groups, and (c) extended search engine—a selected criteria group.

Samples search engine

Add ▾

Sample identification ▾ value of the given form field to mark this field 'not empty' in the results

Sample localization ▾

Sample description ▾

Sample bibliography ▾

SEARCH CLEAR

Multimedia search engine

Title Keywords Country

Date of information preservation Author Protected area name

Genus / Species Object number

Enter character * as a value of the given form field to mark this field 'not empty' in the results

SEARCH CLEAR

Bibliography search engine

Quoting Authors/Editors ⓘ Publication year ⓘ

Publication title Keywords Publication type

Any

Enter character * as a value of the given form field to mark this field 'not empty' in the results

SEARCH CLEAR

Figure 7. View of the search engines for taxonomically unidentified samples (a), multimedia (b), and bibliography (c).

The “Tools” module, which is closely related to the search engines, incorporates two functions: creating statistical reports and visualizing the results of analyses on charts and maps. A detailed description of these tools is beyond the scope of this paper, so we use an example of an analysis aimed at answering the following questions: (i) How many and which scans of *Papaver rhoeas* specimens are represented in the AMUNATCOLL botanical-mycological database collection? (ii) In which countries were they collected? (iii) How are these sites distributed?

After three search criteria (genus/species, iconography, and collection type) are selected, the number of specimens that meet them is displayed (Figure 8a). Adding the grouping factor (country) allows the user to obtain a table with the number and percentage distribution of the specimens and to generate appropriate charts (Figure 8b). Using BioGIS tools, the obtained results can be presented on various types of maps (e.g., dot distribution map, cluster map) (Figure 8c,d). This opens up wide possibilities of analysis built upon the BioGIS tools selected by the user according to his or her needs.

The user can choose the following map types: dot distribution map, area class map, choropleth map, diagram map, cluster map, attribute grouping map, and time-lapse map. Depending on the type of map and the nature of the presented phenomenon, the user can define map properties such as colour, shape and size of signatures, colour and width of borders, colour and transparency of separated areas, and number and width of intervals within the variability range of the analysed feature or phenomenon. The user can define the description properties and design a map legend (including title, author, description length, font sizes, legend location, and width).

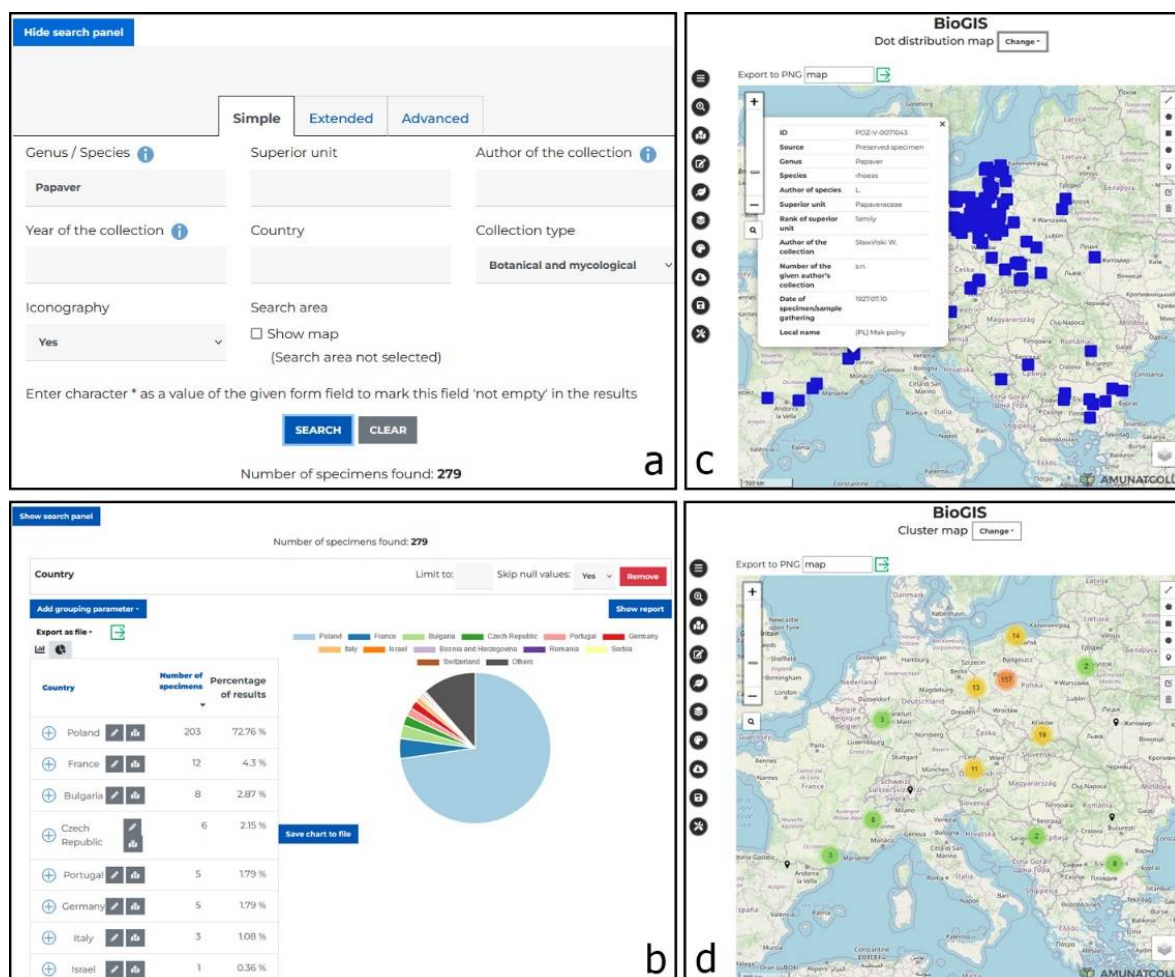


Figure 8. The use of statistical and spatial analysis tools available in the AMUNATCOLL panel: (a) searching for specimens according to selected criteria, (b) quantitative summaries of the search results in the form of tables and graphs, and (c) presentation of the results on maps.

The system includes thematic layers such as the grid of geographical coordinates, the grid of the Universal Transverse Mercator (UTM) [73], the grid of the Atlas of distribution of vascular plants in Poland (ATPOL) [74], countries and voivodships, counties, communes, national parks, and landscape parks in Poland.

In addition, the user has the option to define his or her own vector layers and develop his or her own maps.

4.3. Mobile Application

The AMUNATCOLL mobile application is aimed at users observing flora, fauna, mycobiota, and natural habitats in the field for scientific, educational, and professional purposes related to nature conservation [75]. This mobile application enables the documentation of natural observations in the form of text descriptions, photos, and voice recordings (Figure 9). The observation form contains a list of predefined fields and open fields that the user can define him- or herself. The list of predefined fields includes ordinal data, data identifying the observation (number, date, author), the coordinates of the place of observation, area size, and vegetation cover. A researcher observing species can specify their scientific names from the list provided and assign them characteristics such as quantity and coverage, sex, age, stage of development, and dimensions. An ecologist documenting natural habitats has the option of selecting habitat types from the list provided or entering names according to individual classification. A special form is available for descriptions of

zoological samples, such as soil samples, litter, bird nests, and deadwood. The geographic locations of the field observations saved in the application and the collected samples can be displayed at all times on the map so that the user has a constant view of his or her location while working in the field. To deliver even more control over fieldwork, the application can remember the user's course over the route. The mobile application can be linked to an individual user account established on the AMUNATCOLL portal. Data transferred to a personal account can be processed using all analytical tools implemented on the portal. If these data meet the conditions specified by the AMUNATCOLL administrator, they can be included in the general database and made publicly available to all system users.

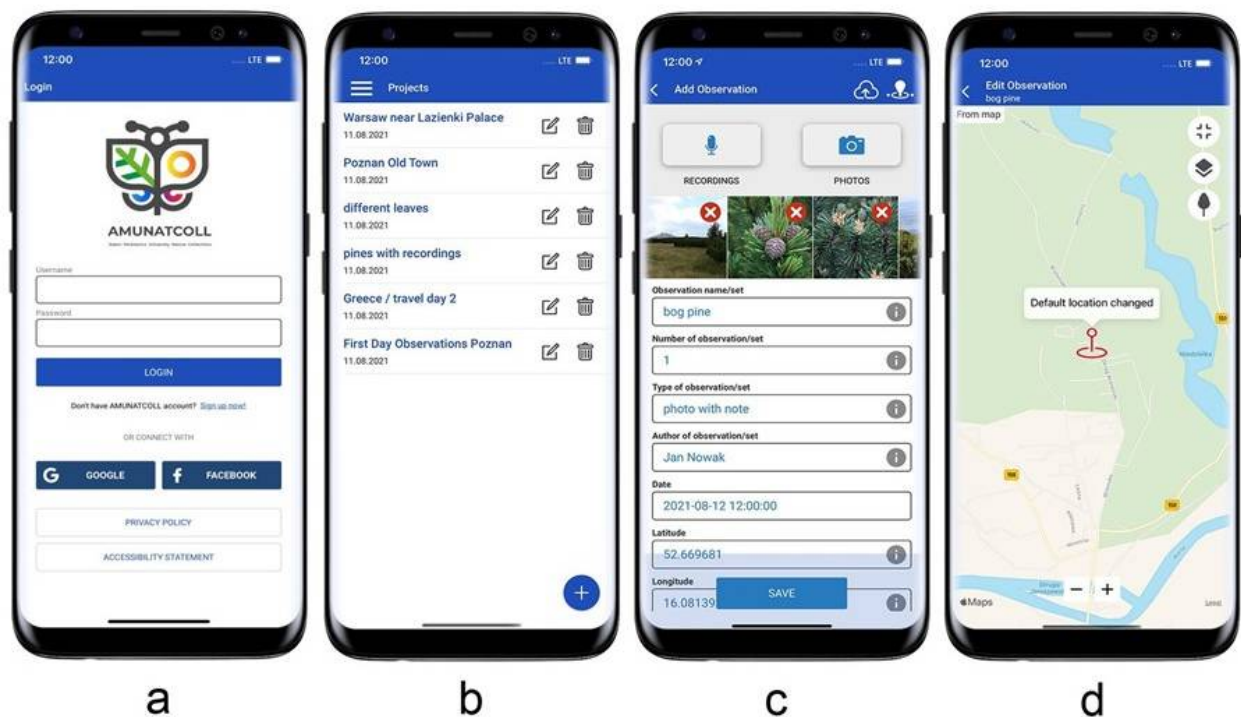


Figure 9. Selected screenshots of the AMUNATCOLL mobile application: (a) login screen, (b) screen with the list of projects, (c) observation screen, and (d) location of the observations on the map. <https://apps.apple.com/pl/app/amunatcoll/id1523442673> and AMUnatcoll <https://play.google.com/store/apps/details?id=pl.pcass.amunatcoll.mobile> (accessed on 20 December 2021).

4.4. API Interface

Access to the AMUNATCOLL IT system takes place through an interface implemented in REST (REpresentational State Transfer) technology (Figure 10). The portal and the mobile application use the same programming interface. Most of the offered functionalities are available only to the logged-in user; therefore, access to individual interface methods is secured with JSON Web Token (JWT). Selected information from the taxonomic database is made available to external entities, primarily the GBIF. For this purpose, the “BioCASE Provider Software” (BPS) service, which is compatible with the “Biological Collection Access Service” [76], was established. This international network of biodiversity repositories combines specimen data from wildlife collections, botanical gardens, zoos, and other research institutions around the world with information from huge observational databases, including the AMUNATCOLL database.

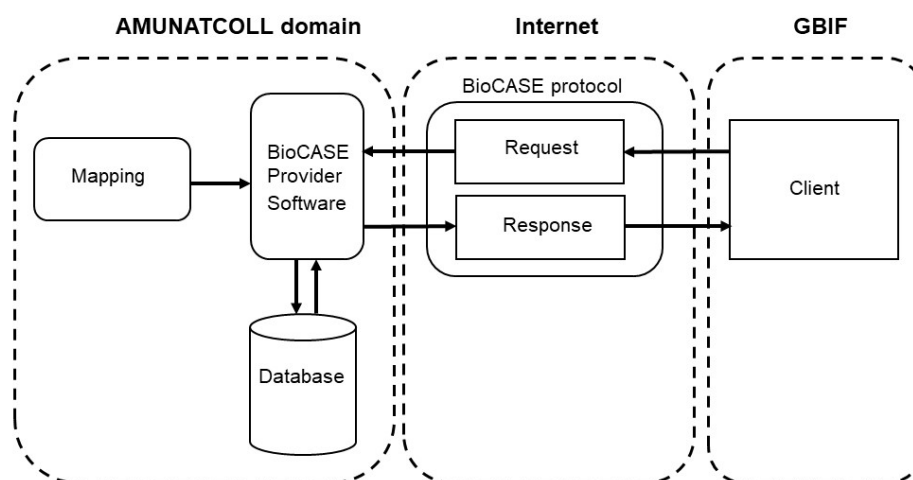


Figure 10. Communication between the AMUNATCOLL IT system and external biodiversity databases.

5. Conclusions

Digital databases based on NHCs are already well established in the science of biodiversity. Large international data aggregators, such as the GBIF or Integrated Digitized Biocollections, are of global importance in facilitating access to biological data. A significant role is also played by small and medium-sized biodiversity collections, which on the one hand enable biodiversity research on a regional and local scale and on the other hand supply information to global databases. Such databases, apart from their scientific role, are also of great importance in the effective protection of biodiversity by developing areas of social activity such as education, citizen science, and environmental management. The social aspect and the related benefits for the wide group of stakeholders should be borne in mind whenever natural resources are being digitized and information systems are being designed to make them available. Digitization of specimens provides a great impetus to undertake new directions of scientific research. Digitization also creates an important opportunity to open access to nature collections to teachers, students, officials responsible for environmental protection, citizens fascinated by nature, and finally everyone who wants to learn to know, investigate, and protect the natural world.

All of the above assumptions drove the design and development of the AMUNATCOLL IT system, which is based on the NHCs of AMU. The ultimate project goals were achieved, inter alia, by

- basing the metadata structure on international standards (ABCD 2.06 and Darwin Core) and enabling the mechanism to transfer data to the GBIF database;
- expanding the metadata structure with fields useful in education, popularization of biodiversity knowledge, and practical nature conservation;
- considering physical specimens and multimedia documents collected directly under empirical environmental study in the digitization process;
- implementing advanced search and analysis tools, especially a rich set of tools for spatial analysis, in the portal;
- integrating the portal with the dedicated mobile application, intended for documenting natural observations directly in the field;
- facilitating a space on the portal for user observations documented by means of a mobile application to work on NHCs;
- creating an open system for collectors wanting to deposit their digital specimens, documents, or observations in the AMUNATCOLL database.

We believe that the SEPP Model of the IT system presented briefly in this work can be an inspiration for many institutions that are developing and implementing collection digitization processes for the benefit of knowledge and the protection of biodiversity.

Author Contributions: All authors contributed to the study design, material preparation, data collection, and analysis. The concept of the publication and the first draft of the manuscript were presented by B.J., M.L. and M.M.N. participated in the development of subsequent versions. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Union through the European Regional Development Fund under the Operational Program Digital Poland (OP PC), Priority axis: II E-administration and open government; Action: 2.3 Digital accessibility and usability of public sector information; Submeasure: 2.3.1 Digital access to public sector information from administrative sources and science resources (Grant number POPC.02.03.01-00-0043/18).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available in the AMUNATCOLL portal <https://amunatcoll.pl/>. (accessed on 20 December 2021).

Acknowledgments: We would like to thank the Ministry of Funds and Regional Policy in Poland for the financial support for the AMUNATCOLL project (Grant number POPC.02.03.01-00-0043/18). Special thanks go to the associates who supported us during the project implementation: Magdalena Dylewska, Maria Jaraszkiewicz, Małgorzata Klimas and Dorota Obiegała.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lane, M.A. Roles of Natural History Collections. *Ann. Mo. Bot. Garden.* **1996**, *83*, 536–545. [CrossRef]
2. Anonymous; [International Commission on Zoological Nomenclature]. *International Code of Zoological Nomenclature*, 4th ed.; London International Trust for Zoological Nomenclature: London, UK, 1999; pp. 1–306.
3. Turland, N.J.; Wiersema, J.H.; Barrie, F.R.; Greuter, W.; Hawksworth, D.L.; Herendeen, P.S.; Knapp, S.; Kusber, W.-H.; Li, D.-Z.; Marhold, K.; et al. (Eds.) *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*; Regnum Vegetabile 159; Koeltz Botanical Books: Glashütten, Germany, 2018.
4. Pyke, G.H.; Ehrlich, P.R. Biological collections and ecological/environmental research: A review, some observations and a look to the future. *Biol. Rev.* **2010**, *85*, 247–266. [CrossRef] [PubMed]
5. Krishtalka, L.; Humphrey, P.S. Can natural history museums capture the future? *BioScience* **2000**, *50*, 611–617. [CrossRef]
6. Hebert, P.D.N.; Cywinska, A.; Ball, S.L.; deWaard, J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B* **2003**, *270*, 313–321. [CrossRef] [PubMed]
7. Suarez, A.V.; Tsutsui, N.D. The value of museum collections for research collections. *BioScience* **2004**, *54*, 66–74. [CrossRef]
8. Hebert, P.D.N.; Gregory, T.R. The promise of DNA barcoding for taxonomy. *Syst. Biol.* **2005**, *54*, 852–859. [CrossRef]
9. Savolainen, V.; Cowan, R.S.; Vogler, A.P.; Roderick, G.K.; Lane, R. Towards writing the encyclopaedia of life: An introduction to DNA barcoding. *Philos. Trans. R. Soc. B* **2005**, *360*, 1805–1811. [CrossRef]
10. Wandeler, P.; Hoeckl, P.E.A.; Keller, L.F. Back to the future: Museum specimens in population genetics. *Trends Ecol. Evol.* **2007**, *22*, 634–642. [CrossRef]
11. Borisenko, A.V.; Sones, J.E.; Hebert, P.D.N. The front-end logistics of DNA barcoding: Challenges and prospects. *Mol. Ecol. Resour.* **2009**, *9*, 27–34. [CrossRef]
12. Lohonya, K.; Livermore, L.; Penn, M. Georeferencing the Natural History Museum's Chinese type collection: Of plateaus, pagodas and plants. *Biodivers. Data J.* **2020**, *8*, e50503. [CrossRef]
13. Advancing Digitization of Biodiversity Collections (ADBC). Available online: https://www.nsf.gov/publications/pubsumm.jsp?WT.z_pims_id=503559&ods_key=nsf15576 (accessed on 26 January 2022).
14. What is the ALA? Available online: <https://www.ala.org.au/about-ala/> (accessed on 26 January 2022).
15. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Available online: <https://www.gob.mx/conabio> (accessed on 26 January 2022).
16. Brazil's Centro de Referência em Informação Ambiental Fundação de Amparo à Pesquisa do Estado de São Paulo. Available online: <http://splink.cria.org.br/indicators/?criaLANG=en> (accessed on 26 January 2022).
17. The Distributed System of Scientific Collection. Available online: <http://disco.eu/> (accessed on 26 January 2022).
18. South African National Biodiversity Institute. Available online: <http://biodiversityadvisor.sanbi.org> (accessed on 26 January 2022).
19. The China's National Specimen Information Infrastructure. Available online: <http://www.nsii.org.cn> (accessed on 26 January 2022).
20. Nelson, G.; Paul, D.; Riccardi, G.; Mast, A. Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys* **2012**, *209*, 19–45. [CrossRef]
21. Smith, V.S.; Blagoderov, V. Bringing collections out of the dark. *ZooKeys* **2012**, *209*, 1–6.

22. Page, L.M.; MacFadden, B.J.; Fortes, J.A.; Soltis, P.S.; Riccardi, G. Digitization of biodiversity collections reveals biggest data on biodiversity. *BioScience* **2015**, *65*, 841–842. [CrossRef]
23. Cantrill, D. The Australasian Virtual Herbaria: Tracking data usage and benefits for biological collections. *Appl. Plant Sci.* **2018**, *6*, e1026. [CrossRef]
24. Howarth, F. The Future of Research in Natural History Museums. In *The Future of Natural History Museums*; Dorfman, E., Taylor and Francis Group, Eds.; Routledge: London, UK, 2018; pp. 65–81.
25. Norris, C. The Future of Natural History Collections. In *The Future of Natural History Museums*; Dorfman, E., Taylor and Francis Group, Eds.; Routledge: London, UK, 2018; pp. 13–28.
26. Soltis, P.S.; Nelson, G.; James, S.A. Green digitization: Online botanical collections data answering real-world questions. *Appl. Plant Sci.* **2018**, *6*, e1028. [CrossRef]
27. Willis, C.G.; Ellwood, E.R.; Primack, R.B.; Davis, C.C.; Pearson, K.D.; Gallinat, A.S.; Yost, J.M.; Nelson, G.; Mazer, S.J.; Rossington, N.L.; et al. Old Plants, New Tricks: Phenological Research Using Herbarium Specimens. *Trends Ecol. Evol.* **2017**, *32*, 531–546. [CrossRef]
28. Yost, J.M.; Sweeney, P.W.; Gilbert, E.; Nelson, G.; Guralnick, R.; Gallinat, A.S.; Ellwood, E.R.; Rossington, N.; Willis, C.G.; Blum, S.D.; et al. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. *Appl. Plant Sci.* **2018**, *6*, e1022. [CrossRef]
29. Park, D.S.; Breckheimer, I.; Williams, A.C.; Law, E.; Ellison, A.M.; Davis, C.C. Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States. *Philos. Trans. R. Soc. B* **2019**, *374*, 20170394. [CrossRef]
30. Pearson, K.D. A new method and insights for estimating phenological events from herbarium specimens. *Appl. Plant Sci.* **2019**, *7*, e1224. [CrossRef] [PubMed]
31. Taylor, S.D. Estimating flowering transition dates from status-based phenological observations: A test of methods. *PeerJ* **2019**, *7*, e7720. [CrossRef] [PubMed]
32. Miller, T.K.; Gallinat, A.S.; Smith, L.C.; Primack, R.B. Comparing fruiting phenology across two historical datasets: Thoreau's observations and herbarium specimens. *Ann. Bot.* **2021**, *128*, 159–170. [CrossRef] [PubMed]
33. Rivers, M.C.; Taylor, L.; Brummitt, N.A.; Meagher, T.R.; Roberts, D.L.; Lughadha, E.N. How many herbarium specimens are needed to detect threatened species? *Biol. Conserv.* **2011**, *144*, 2541–2547. [CrossRef]
34. Lughadha, E.N.; Miller, C. Accelerating global access to plant diversity information. *Trends Plant Sci.* **2019**, *14*, 622–628. [CrossRef] [PubMed]
35. Rocchetti, G.A.; Armstrong, C.G.; Abeli, T.; Orsenigo, S.; Jasper, C.; Joly, S.; Bruneau, A.; Zytaruk, M.; Vamosi, J.C. Reversing extinction trends: New uses of (old) herbarium specimens to accelerate conservation action on threatened species. *New Phytol.* **2020**, *230*, 433–450. [CrossRef] [PubMed]
36. Gallinat, A.S.; Russo, L.; Melaas, E.K.; Willis, C.G.; Primack, R.B. Herbarium specimens show patterns of fruiting phenology in native and invasive plant species across New England. *Am. J. Bot.* **2018**, *105*, 31–41. [CrossRef]
37. Glon, H.E.; Heumann, B.W.; Carter, J.R.; Bartek, J.M.; Monfils, A.K. The contribution of small collections to species distribution modelling: A case study from Fuireneae (Cyperaceae). *Ecol. Inform.* **2017**, *42*, 67–78. [CrossRef]
38. Erickson, K.D.; Smith, A.B. Accounting for imperfect detection in data from museums and herbaria when modelling species distributions: Combining and contrasting data-level versus model-level bias correction. *Ecography* **2021**, *44*, 1342–1352. [CrossRef]
39. Hedrick, B.P.; Heberling, J.M.; Meineke, E.K.; Turner, K.G.; Grassa, C.J.; Park, D.S.; Kennedy, J.; Clarke, J.A.; Cook, J.A.; Blackburn, D.C.; et al. Digitization and the Future of Natural History Collections. *BioScience* **2020**, *70*, 243–251. [CrossRef]
40. Andrew, C.; Diez, J.; James, T.Y.; Käuserud, H. Fungarium specimens: A largely untapped source in global change biology and beyond. *Philos. Trans. R. Soc. B* **2019**, *374*, 20170392. [CrossRef] [PubMed]
41. Heberling, J.M.; Prather, L.A.; Tonsor, S.J. The Changing Uses of Herbarium Data in an Era of Global Change: An Overview Using Automated Content Analysis. *BioScience* **2019**, *69*, 812–822. [CrossRef]
42. Lang, P.L.M.; Willems, F.M.; Scheepens, J.F.; Burbano, H.A.; Bossdorf, O. Using herbaria to study global environmental change. *New Phytol.* **2019**, *221*, 110–122. [CrossRef] [PubMed]
43. Canteiro, C.; Barcelos, L.; Filardi, F.; Forzza, R.; Green, L.; Lanna, J.; Leitman, P.; Milliken, W.; Morim, M.P.; Patmore, K.; et al. Enhancement of conservation knowledge through increased access to botanical information. *Conserv. Biol.* **2019**, *33*, 523–533. [CrossRef] [PubMed]
44. Ballard, H.L.; Robinson, L.D.; Young, A.N.; Pauly, G.B.; Higgins, L.M.; Johnson, R.F.; Tweddle, J.C. Contributions to conservation outcomes by natural history museum-led citizen science: Examining evidence and next steps. *Biol. Conserv.* **2017**, *208*, 87–97. [CrossRef]
45. von Konrat, M.; Campbell, T.; Carter, B.; Greif, M.; Bryson, M.; Larrain, J.; Trouille, L.; Cohen, S.; Gaus, E.; Qazi, A.; et al. Using citizen science to bridge taxonomic discovery with education and outreach. *Appl. Plant Sci.* **2018**, *6*, e1023. [CrossRef] [PubMed]
46. Butcher, K.R.; Power, M.J.; Larson, M.; Orr, M.P.; Velasquez-Franco, S.; Hudson, M.A.; Bailey, V.J. Museum Leadership for Engaging, Equitable Education: The Transformative Potential of Digitized Collections for Authentic Learning Experiences. *Curator: Mus. J.* **2019**, *64*, 383–402. [CrossRef]
47. What is GBIF? Available online: <https://www.gbif.org/what-is-gbif> (accessed on 26 January 2022).
48. Wheeler, Q.D. What is GBIF? *BioScience* **2004**, *54*, 717. [CrossRef]

49. Flemons, P.; Guralnick, R.; Krieger, J.; Ranipeta, A.; Neufeld, D. A web-based GIS tool for exploring the world's biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA). *Ecol. Inform.* **2007**, *2*, 49–60. [CrossRef]
50. Lane, M.A.; Edwards, J.L. The Global Biodiversity Information Facility (GBIF). In *Biodiversity Databases Techniques, Politics, and Applications*; Curry, G.B., Humphries, C.J., Eds.; Systematics Association Special Volume; CRC Press Taylor and Francis Group: Boca Raton, FL, USA; London, UK; New York, NY, USA, 2007; pp. 1–4. Available online: <https://www.taylorfrancis.com/chapters/oa-edit/10.1201/9781439832547-1/global-biodiversity-information-facility-gbif-meredith-lane-james-edwards> (accessed on 26 January 2022).
51. Welcome to the iDigBio Portal. Available online: <https://www.idigbio.org/portal> (accessed on 26 January 2022).
52. Nelson, G.; Ellis, S. The history and impact of digitization and digital data mobilization on biodiversity research. *Philos. Trans. R. Soc. B* **2018**, *374*, 20170391. [CrossRef]
53. AMUNATCOLL. AMU Nature Collections on the Internet: Digitization and Making Available the Resources of Biodiversity Data of the FACULTY of Biology at Adam Mickiewicz University in Poznań. Available online: <http://anc.amu.edu.pl/eng/index.php> (accessed on 26 January 2022).
54. What's PSNC. Available online: <https://www.psync.pl/about/whats-psnc/> (accessed on 26 January 2022).
55. Konwerski, S.; Błoszyk, J. Zbiory Przyrodnicze Wydziału Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu—struktura i funkcjonowanie [Natural History Collections of the Faculty of Biology at the Adam Mickiewicz University in Poznań—structure and functioning]. *Opusc. Mus.* **2013**, *21*, 137–144.
56. Index Herbariorum. Available online: <http://sweetgum.nybg.org/science/ih/> (accessed on 26 January 2022).
57. Grance, T.; Hash, J.; Stevens, M. Security Considerations in the Information System Development Life Cycle: Recommendations of the National Institute of Standards and Technology. In *NIST Spec. Public. 800-64 Rev. 1*; Computer Security Division Information Technology Laboratory National Institute of Standards and Technology: Gaithersburg, MD, USA; Government Printing office Washington: Washington, DC, USA, 2004; pp. 1–25.
58. Grance, T.; Kent, K.; Kim, B. Computer Security Incident Handling Guide. National Institute of Standards and Technology Special Publication. In *NIST Spec. Publ. 800-61*; Computer Security Division Information Technology Laboratory National Institute of Standards and Technology: Gaithersburg, MD, USA; Government Printing office Washington: Washington, DC, USA, 2004; pp. 1–148.
59. Nowak, M.M.; Słupecka, K.; Jackowiak, B. Geotagging of natural history collections for reuse in environmental research. *Ecol. Indic.* **2021**, *131*, 108131. [CrossRef]
60. Natural History Collections of the Faculty of Biology AMU at GBIF. Available online: <https://www.gbif.org/dataset/84b18cce-083a-4464-bee8-25b2083a17cd> (accessed on 26 January 2022).
61. 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 (accessed on 26 January 2022).
62. Webster, M.S.; Buschbom, J.; Hardisty, A.; Bentley, A. The Digital Extended Specimen Will Enable New Science and Applications. *Biodivers. Inf. Sci. Stand.* **2021**, *5*, e75736. [CrossRef]
63. Consultation of Digital Extended Specimens—Phase 2. Available online: <https://discourse.gbif.org/t/digital-extended-specimens-phase-2/2651> (accessed on 20 July 2022).
64. The Specification for Open Digital Specimens (openDS) at GitHub. Available online: <https://github.com/DiSSCo/openDS> (accessed on 20 July 2022).
65. Lendemer, J.; Thiers, B.; Monfils, A.K.; Zaspel, J.; Ellwood, E.R.; Bentley, A.; LeVan, K.; Bates, J.; Jennings, D.; Contreras, D.; et al. The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* **2020**, *70*, 23–30. [CrossRef] [PubMed]
66. Biodiversity Collections Network. Available online: <https://bcon.aibs.org/> (accessed on 20 July 2022).
67. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG). Available online: <http://www.tdwg.org/standards/115> (accessed on 26 January 2022).
68. Wieczorek, J.; Bloom, D.; Guralnick, R.; Blum, S.; Döring, M.; Giovanni, R.; Robertson, T.; Vieglias, D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* **2012**, *7*, e29715. [CrossRef] [PubMed]
69. Belgian Biodiversity Platform. Available online: <https://www.biodiversity.be/1767> (accessed on 26 January 2022).
70. Senterre, B.; Wagner, M. Standardization of Data Collection and Creation of a Biodiversity Database: A PostgreSQL-PostGIS Database for Island Conservation Society (Seychelles). In *Consultancy Report*; Government of Seychelles, United Nations Development Programme: Victoria, Seychelles, 2016; pp. 1–40.
71. de Oliveira, L.M.; de Brito, R.N.; Guimarães, P.A.S.; Mastrângelo, V.; Dos Santos, R.A.; Diotaiuti, L.G.; de Cássia Moreira de Souza, R.; Ruiz, J.C. TriatoKey: A web and mobile tool for biodiversity identification of Brazilian triatomine species. *Database* **2017**, *2017*, bax033. [CrossRef] [PubMed]
72. AMUNATCOLL. Portal. Available online: <https://amunatcoll.pl/> (accessed on 26 January 2022).
73. The Universal Transverse Mercator. Available online: https://en.wikipedia.org/wiki/Universal_Transverse_Mercator_coordinate_system (accessed on 26 January 2022).
74. Zając, A. Atlas of distribution of vascular plants in Poland (ATPOL). *Taxon* **1978**, *27*, 481–484. [CrossRef]

-
75. AMUNATCOLL. Mobile app. Available online: <https://apps.apple.com/pl/app/amunatcoll/id1523442673> and AMUnatcoll <https://play.google.com/store/apps/details?id=pl.pcass.amunatcoll.mobile> (accessed on 26 January 2022).
 76. Biological Collection Access Service. Available online: <https://www.biocase.org/> (accessed on 26 January 2022).