BioCISE

Resource Identification for a Biological Collection Information Service in Europe

BGBM

Botanic Garden and Botanical Museum Berlin-Dahlem

Resource Identification

for a

Biological Collection Information Service in Europe (BioCISE)

Results of the Concerted Action “BioCISE Resource Identification”, funded by the European Commission, DG XII, within the EU Fourth Framework's Biotechnology Programme, August 1, 1997 to December 31, 1999

Edited by Walter G. Berendsohn

Botanic Garden and Botanical Museum Berlin-Dahlem
Department of Biodiversity Informatics

The Authors:

Andrea Hahn, Anton Güntsch, Birgit Felinks, Linda Olsvig-Whittaker, Louis Rechaussat, Pedro Fernandes, Pier-Luigi Nimis, Richard White, Walter Berendsohn, and Wouter Los were members of the Concerted Action or belonged to the BioCISE secretariat (see p. iv).

David Lazarus, Institut fuer Paläontologie, Museum für Naturkunde, Zentralinstitut der Humbolt-Universität zu Berlin, Invalidenstraße 43, D-10115 Berlin, Germany

Mark J. Costello and Chris Emblow, Ecological Consultancy Services Ltd (EcoServe), 17 Rathfarnham Road, Terenure, Dublin 6W, Ireland

Jürgen Koenemann and Christoph Thomas, humanIT Human Information Technologies GmbH, Rathausallee 10, D-53757 Sankt Augustin, Germany

Neil Thomson, The Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom

Printed in Germany.

ISBN 3-921800-44-7

Published by the Botanic Garden and Botanical Museum Berlin-Dahlem,
Freie Universität Berlin
Königin-Luise-Straße 6-8, 14191 Berlin, Germany
www.bgbm.fu-berlin.de

Cover: model of *Utricularia* sp. Exhibition, Botanical Museum Berlin-Dahlem (photograph: I. Haas).

Contents

Members of the Concerted Action

Anastasios Anagnostopoulos, The Goulandris Natural History Museum,14562 Athens, Greece

Prof. Dr. Walter Berendsohn (Project co-ordinator), Freie Universität Berlin, ZE Botanischer Garten und Botanisches Museum Berlin-Dahlem, 14191 Berlin, Germany

Pedro Fernandes, Instituto Gulbenkian de Ciencia, 2781 Oeiras Codex, Portugal

Gregor Hagedorn, Biologische Bundesanstalt für Land- und Forstwirtschaft, Institut für Mikrobiologie, 14195 Berlin, Germany,

Dr. Jasmin Jakupovic, (formerly at) Technische Universität Berlin, Germany

Prof. Dr. Jacques Lebbe †, Université Paris VI, Laboratoire Organisation & Evolution des Systèmes, 75252 Paris Cedex 5, France

Dr. Wouter Los, Institute for Systematics and Population Biology / Zoological Museum Amsterdam, Amsterdam, The Netherlands

Prof. Dr. Jos van der Maesen, Agricultural University Wageningen, Dept. of Plant Taxonomy, 6700 ED Wageningen, The Netherlands

Prof. Dr. Pier Luigi Nimis, University of Trieste, Department of Biology, 34127 Trieste, Italy

Dr. Richard J. Pankhurst, Royal Botanic Garden Edinburgh, Taxonomic Computing, Edinburgh EH3 5LR, United Kingdom

Dr. Louis Réchaussat, INSERM, 75654 Paris Cedex 13, France

Dr. Jarmo Saarikko, Finnish Forest Research Institute, 00170 Helsinki, Finland

Dr. Karsten Siems, Analyticon AG, Research Department, 13355 Berlin, Germany

Prof. Dr. Benito Valdés, University of Sevilla, Depto. de Biologia Vegetal y Ecologia, 41012 Sevilla, Spain

Besher Wattar, (formerly at) Novo Nordisk A/S, Scientific Computing, 2880 Bagsvaerd, Denmark

Dr. Richard White, University of Southampton, Biodiversity & Ecology Research Division, Southampton SO16 7PX, United Kingdom

Dr. Linda Olsvig-Whittaker, Nature Reserves Authority, Science and Management Division, Jerusalem 94467, Israel, linda@bgumail.bgu.ac.il

Project Secretariat

Secretariat: Birgit Felinks (1st year) Andrea Hahn (2nd and 3rd year), Anton Güntsch, Sylvia Steinmann. Collaborator: Dr. Cornelia Lehmann, Letras Informatik GmbH, Berlin.

XI. Concepts for a European Portal to Biological Collections

Walter G. Berendsohn, Mark J. Costello, Chris Emblow, Anton Güntsch, Andrea Hahn, Jürgen Koenemann, Christoph Thomas, Neil Thomson and Richard White

The accessibility and thus use of biological collections would be significantly improved by a common "portal" through which information on the holdings of research institutes, museums, survey organizations etc. can be accessed. This portal may take the form of an Internet site where the reader can search for information based on biological names, taxonomic groups, habitat names (e.g. marshes), ecological relationships between species (e.g. parasitism), and geographic sources of specimens and observations of species. Unfortunately, such a search facility is far more complex than immediately apparent. One of the main obstacles is that collections use terminology going back 300 years and it is not feasible for most facilities to update their data comprehensively with changes in taxonomic nomenclature and geographic or political boundaries. However, we posit that this problem, as well as the difficulties created by the fragmentation of the collection community itself (see box on p. 3) can be overcome by a pragmatic and concerted effort of the interested parties.

Why should we try?

Channelling collection information into a common access system makes sense because

- the combination of information from thematically different data areas will enhance knowledge discovery and understanding
- users will be presented with a common interface covering a wide range of known and not yet known inquiries
- it will stimulate efforts to find agreement on good practice, standardization of data items and quality control
- a concerted approach will – to some extent - remove duplication of efforts; scarce technical resources can be put to use in a focused and collaborative way.

In the process of assessing user requirements and available resources on the provider's side it became clear that such a service had to be very flexible; scalable on the collection owner's (provider's) side, simple in its internal mechanisms, broad in its cover of collections from different sub-disciplines, and providing a user interface adaptable to users' needs. Realising that an extensive Europe-wide specimen ("unit") - based access system is not yet within reach, but that user demand exists for concerted access to collection information right away, it was decided to focus on the creation of a collection-level information system as the kernel of "The BioCISE". However, from the beginning a unit-level approach should be integrated (see Chapter IX). Recent initiatives, particularly the formation of – and EU support for – the European Natural History Specimen Information Network (ENHSIN 2000) are a promising proof of intent on both the information providers' and the funding agencies' side.

Achieving breadth and scalability: Meta-information to the rescue!

As has been shown in Chapter X, the physical as well as the information content of biological collections can be described using meta-information, i.e. information linked to sets of units. The interesting fact about such metadata items is that – with few exceptions – they are applicable to both, unit-level and collection level.

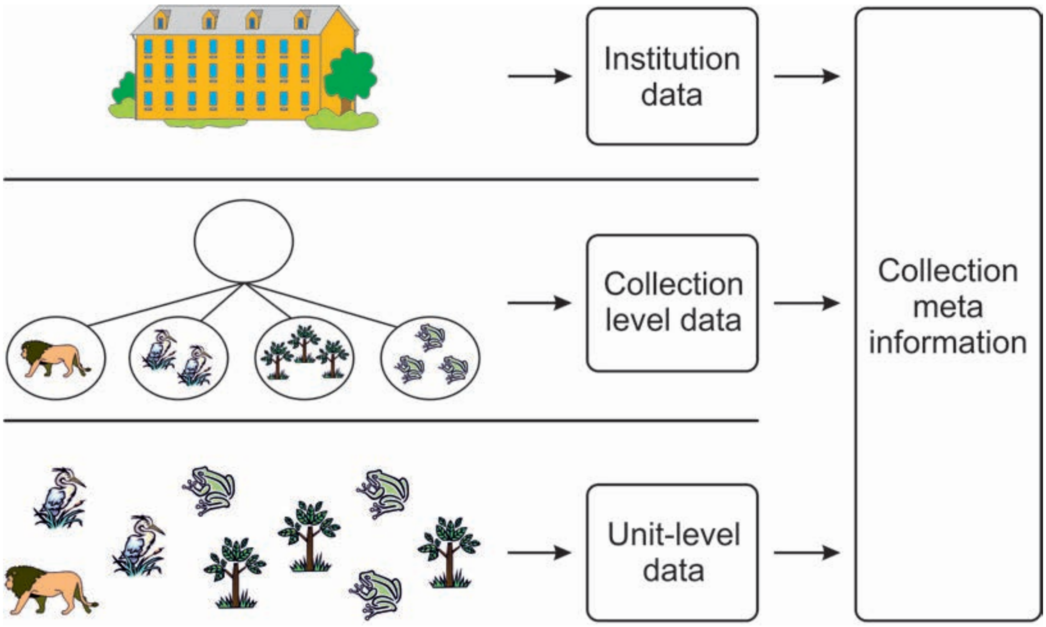

Figure 13: Physical hierarchy of reference points for collection meta-information

This presents a first important chance to achieve scalability of the system. Generally, the information referring to a high level in the physical hierarchy (fig. 13), an entire Natural History Museum for example, will be much less specific than that referring to a single unit. However, exceptions can be found e.g. in specialised collections, which consist of only one species, as well as in reference collections for a specific site.

The meta-information can be at any level of detail, from the very general (plants, Europe, 19th century) down to the very specific (the species, detailed gathering site, date). Between these extremes a fluent transition exists because most of the important data items describing collections and units are belonging to a (more or less) hierarchical classification system (e.g. geographical: country - department). This is the second important factor, which can contribute to achieving a scalable system. A system using meta-information organized into such a schema will allow processing of very detailed collection descriptions but still be able to provide information on collections that cannot or do not want to supply this kind of detail. It will enable researchers to locate needed information or materials and obtain them by conventional means, if necessary.

In addition, meta-information may serve as a valuable enrichment of unit data especially when statements about data quality and procedures are provided. However, at present, the major advantage of metadata lies in the fact that they can provide information about - and facilitate access to - units even where unit-level data are not available.

A knowledge based approach: Metadata 'thesauri'

Identified priorities. The two most important data areas for meta-information about collections are names of organism groups and named areas (the latter using geographical, biogeographical, geological, palaeontological or ecological terminology). Unfortunately, these areas are not stable: terms may have different meanings depending on who applied them and when (e.g. “Germany”, “Liliaceae”). Moreover, geoecological and taxonomic class names represent scientific concepts, thus parallel, partly overlapping hierarchies may exist (e.g. generic and family delimitation in systematics).

Consequently, no single standard hierarchy exists for any of these information domains. The development and application of integrated metadata 'thesauri' and classifications for these data areas is a prerequisite for the functioning of an extendible collection information service. To be clear: these will be pragmatic tools to facilitate access; they will not attempt to redefine terms or derive new classifications. They will allow searches by keyword and by following hierarchical links despite the underlying anarchy. Users must be able to select and/or specify fuzzy concepts such as habitat boundaries (“Rainforest”) or undefined geographic terms (“Central Europe”) that don’t map easily to the political boundaries available in today's gazetteers and geographical information systems. Taxonomic concepts and names representing organism groups present similar problems of parallel and partially overlapping hierarchies. To include them in data access interfaces, information structures and methods which are able to accommodate and process such complex inter-relations between individual metadata elements must be defined.

The thesaurus forms a common source for both indexing of collections and the design of the portal’s user interface (e.g. implementation of a convenient taxonomic browser instead of free text fields). The thesaurus to be constructed has to be powerful enough to treat various semantic relations such as synonyms and hierarchies. It has to put a special focus on taxonomic and geographic terminology to fulfil the special requirements of the biological community. National and thematic networks can derive keywords to describe their collections from this thesaurus to provide a homogeneous data source utilized by the central catalogue. In addition to data provided by the networks, rule based technologies can be used to represent complex weighted relations among thesaurus’ elements and to further enrich the set of keywords by deriving useful categories from collection descriptions. Rule based indexing will reduce the costs for the time consuming process of human indexing.

Geo-ecological thesaurus. With respect to geo-ecological classifications of collections, two types of questions have to be addressed: The comparatively simple question of where a collection (institute) is located (e.g. “Is there a reference collection of microbial strains in town X?”) and the much more complicated question where specimens collected at a defined site can be found (e.g. “Who has holdings of specimens from the northern Mediterranean coast?” or “Where can I find organisms collected from Late Triassic St. Cassian Formation?”). The data for the former question can be captured from rather standardised address information and related to existing data collections on present administrative boundaries. The compilation of metadata for the second question is difficult for various reasons. In contrast to taxonomic data, no agreed systems of nomenclature for geographic, ecological, or palaeontological “areas” exist. The catalogue has to deal with very variable applications of terms (e.g. “St. Cassian”, “Sankt Cassian”, “Cassinian” in the above example), consider provisions for changes in the delimitation of areas in time (e.g. in the case of “Germany” and “Yugoslavia”), consider the problems of more or less linear references (“Mediterranean coast”, “River Guadalquivir”), vague delimitation (“northern Spain”) and the problems of scientific concepts represented by ecological and palaeontological terms. The identification of existing data collections (e.g. available gazetteers) and the contacting of geographical and ecological institutes will be prominent approaches. The decision on practical cut-off points for the hierarchical representation of the data (as opposed to a synonymised keyword list) is fundamental. Mapping the European languages into such a thesaurus (initially to be implemented largely in English) is a problem, which is tackled extensively in various projects; collaboration has to be sought here as well.

Taxonomic thesaurus. Rules of nomenclature (“Codes”, e.g. ICZN 1999, Greuter et al. 2000) exist for the area of taxonomic data (scientific names of organisms). However, synonyms, conceptual differences between applications of the same class name, as well as the problem of congruence of concepts with differing names also persist in taxonomy. This is the consequence of the naming system being a pragmatic approach to a rather complicated scientific problem: the classification of life on earth according to its natural evolution. Large data collections have already been identified which can be used to compile a catalogue of names down to genus level. This can be used as a backbone classification to which other terms can be associated. “Pseudotaxa”, i.e. higher-level class names for organisms that do not directly correspond to a taxonomic group (e.g. “Medicinal plants”, “Pests”, “Birds of prey”, “Trees”, “Microbes”) must also be treated.

Other data areas. Data areas that do not belong to one of the aforementioned categories must also be tackled. This covers, e.g., temporal aspects (date of collection event, dwelling time of the organisms in palaeontological contexts, etc.), representation of collection purposes (research material, archive of vouchers, exhibition, etc.) and preservation methods (often important for potential use of materials in analysis), among others. The compilation of such thesauri can be based on an analysis of the data provided by the BioCISE survey database and by looking at other similar data collections, to find out about terms applied by users in their queries, and terms used by institutes to describe their collections.

Issues for informatics research

Modelling and implementing catalogues and their utilization in user interfaces are general tasks of informatics research and application development. A partnership with an information technology provider should be sought to avoid duplication of efforts.

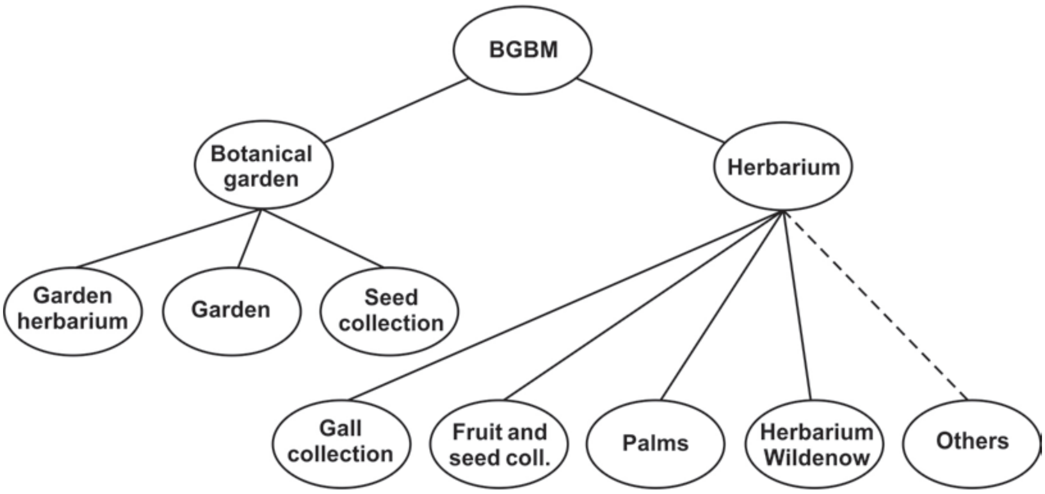

Figure 14: Hierarchical structure of a selection of collections at the Botanic Garden and Botanical Museum Berlin-Dahlem (BGBM)

A proposed representation of entire collections. Since biological collections are generally organized hierarchically, it is straightforward to describe them as trees, the nodes representing sub collections (sets of units), and the connecting lines representing “is part of” relationships (fig. 13). Each node is linked to sets of attributes (e.g. taxonomic identification, ownership, locality) providing the sub-collections’ properties (fig. 14). These attributes are referenced in the metadata thesaurus and thus can be referred to other, more general or more specific terms. Properties can be further quantified by labelled links to express the fuzziness typical for collection descriptions (e.g. “mainly” Coleoptera, “some” Lepidoptera). Available metadata are often incomplete (for example if derived from questionnaires). Adding “dummy” nodes labelled “other” can indicate this (fig. 13).

Güntsch et al. (2000) demonstrated the use of this representation to formulate rules to derive complex concepts describing classes of collections (e.g. “Natural History Museum”, “Botanical Garden”). Future work will have to analyse the inheritance of properties within a tree representing both exact and fuzzy data to achieve a solid theoretical ground for the implementation of indexing modules, data capture tools, and user interfaces.

Information model. Documentation of heterogeneous information resources is a current topic e.g. in library science, museology, and environmental and medical informatics. The definition of meta-information attributes associated with collections of biological material must be based on an evaluation of existing and emerging metadata standards and work effected by current international working groups involved in the standardisation of access to museum resources. General metadata standards such as the Dublin Core definitions (Anon. 1998) must be incorporated, too.

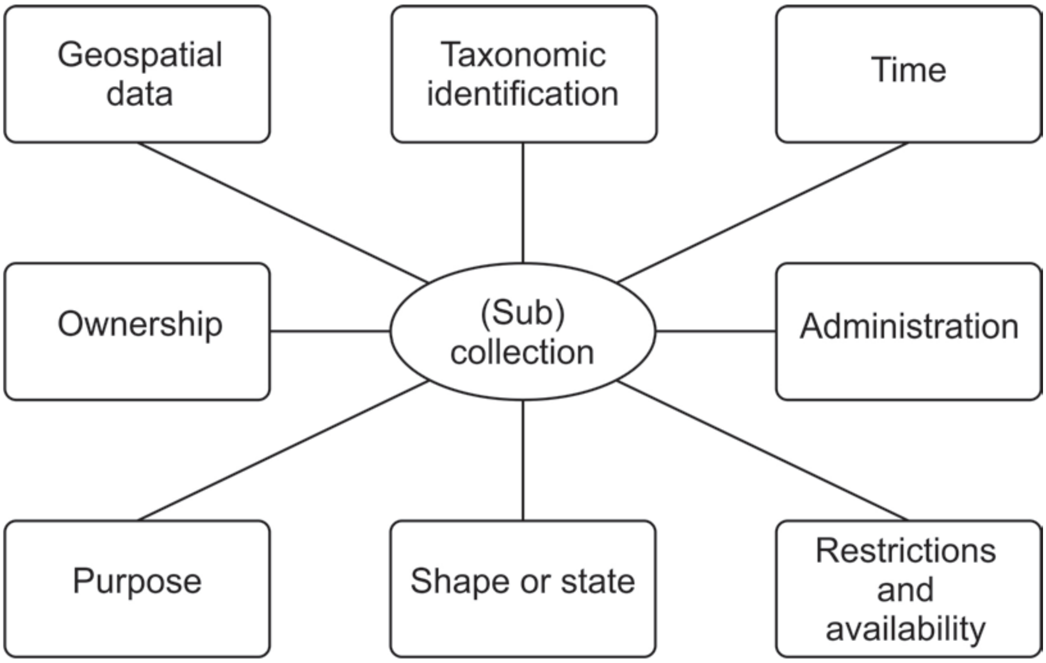

Figure 15: Sets of attributes providing (sub-) collection properties

The representation of hierarchical structures, use of controlled vocabularies, incomplete or fuzzy data, and administrative metadata are problems extending far beyond the scope of biological collections. Similar problems are encountered in the domain of environmental information systems (e.g. in the context of the European Topic Centre on Catalogue of Data Sources, EEA 1999) or have indeed been identified as an unresolved problem for all collections (including museums, documents, archives, subject gateways, etc.) by the collection description working group of UKOLN (UK Office for Library and Information Networking; see Heaney 2000). An important component of the design process for the European Collection Information Service will be the provision of a theoretical model which helps to find a practical solution to the problems addressed.

The results influence the design of the metadatabase system directly, but also indirectly, by way of the content structure definition of the metadata catalogues.

User interface. There are some standard techniques to allow navigation in hierarchies. For example, in Yahoo!-style sites and most online shopping or product selection environment hierarchies, there are lists of links, and following a link will result in a new page being displayed, possibly with further lists of links until the terminal nodes of the hierarchy are reached. These types of interfaces work well with small hierarchies containing well-known entries but are inappropriate for the task at hand. Direct manipulation interfaces with very fast updates of displays in reaction to user input are needed, with displays that allow users to see and select content rather than having to specify their needs formally. These interfaces have to be tailored to the particular information at hand. Furthermore, interfaces will need to be personalised based on user characteristics and current task. For example, graphic browsing of a hierarchy with Latin names of species may be appropriate for experts who are familiar with taxonomic trees and the specialist terms. Conversely, novices may require a different structure and different terminology likely to be supported by visual means such as images and symbols.

Novel interaction techniques have to be developed, because users must be able to browse and search in multiple, linked hierarchies without loosing orientation, the system and its interfaces must represent missing, fuzzy, or incorrect (outdated) information, and users must be able to select and/or specify often fuzzy geoecological concepts.

Novel technical solutions will need to be developed to design and implement user interfaces that address these issues and at the same time support a large set of users with a diverse set of hardware (network bandwidth and processing speed) and software (browsers, Java, etc.) constraints.

Knowledge processing. In addition to evaluating distributed web sources, quite often new information can be extracted from the existing. This may concern the seemingly obvious one not thought to be necessary to put into (key)words: When searching for micro-organisms, looking into microbial collections seems the natural approach. But what is labelled as a “microbial collection”? Does a search for the keyword also answer with a cheese producer’s *Lactobacter* strains, or an algal reference culture collection? Are we talking prokaryotes, unicellular, or just “small”? The task of a knowledge-processing module is to apply man-made rules for such definitions to existing data and thereby, for example, generate new keywords. Through the possibilities of assigning different weightings, probabilities may be calculated: Asking for micro-organisms should deliver all bacterial collections, but also offer others lower down on the list.

One problem will be that, even with good thesauri available, most of the information used in the service will not adhere to a single structure. For example, collections may represent the data related to the gathering of a unit as a single field of text, or they may provide this in a highly structured, atomised form. Detailed knowledge about the hierarchical decomposition of such information will be very useful in the process of extracting information from text sources, especially if combined with the thesaurus.

The information provider's point of view

Synergetic effects: In the national meetings of collection holders organised or co-organised by BioCISE it became clear that the cross-subdiscipline approach was greeted with enthusiasm, but that communication even within sub-disciplines was generally wanting. Participants presented a multitude of isolated information systems developed or in preparation in their institutes (see Chapter VIII). The potential for synergetic effects was obvious. BioCISE was perceived as a possibility to focus resources and to overcome existing institutional rivalries or other political impediments, which may bear on the development of collaboration on the national level. In this context it is also important that almost all information providers would also be users of the service.

IPR and other legal concerns: Still existing impediments to the networking of unit-level data resources are the unresolved questions of IPR in databases and database networks (see Chapter II). Another important area of concern is the unresolved question of obligations imposed by the Convention on Biological Diversity (particularly in the case of living collections, which undoubtedly represent “genetic resources”). The use of metadata was greeted by some of the institutes with more advanced data holdings as a possibility to await a solution of these problems before going public with their data holdings, but at the same time being able to advertise their collection's scientific information content.

Promotion of collections. Collection holders also consider it essential to improve public understanding of the importance of natural science collections, and of the relationship between collection conservation efforts with the ability to manage, preserve, and interpret our natural heritage as well as the world we live in. One of the most important future tasks will be to show that the varieties of collections are actually daily used, and that they are of public interest. Hence the credit of the users will serve as motive to establish and to maintain a Collection Information System.

Basic organization of the Service

Data capture: As mentioned before, national meetings and spin-off activities led to the realization that the European service must rely on national networks or nodes to actually collect the information. For projects implementing such nodes national funding may be found – as has already happened in the case of Belgium, Austria, and Germany, and as it will hopefully be achieved in other European countries. The political argument in favour of such funding lies in the obligations incurred by government in the context of international conventions, the contribution to an over-all European research infrastructure, and synergetic effects to be achieved on all levels. However, governments and funding bodies do seem to believe that unit-level data capture is within the scope of individual institute's activities, so that successful attempts to find extra financing for such endeavours will probably remain the exception. For collection-level information, the BioCISE survey database can serve as an initial dataset to build upon.

National nodes. Regarding the access to technical innovation, Europe has grown closer over the past years. In the implementation, and above all, population, of biological collection information systems, however, it may still take a long time to reach a common level. This does not demean the value of the collections themselves but their chances to be recognised, preserved, and properly valued. The aim of the National Node set-up is to provide equal chances for all of Europe's biological collections to be represented in a common information system. The metadata access system will help to override the inherent inequality resulting from still widely differing access to information systems, databasing expertise and staff supply. In some countries (e.g. Belgium), national information systems are already well developed, needing little initial input to adapt to contribute to a European system. In other cases providing the basic means (software and training) for the initial set-up as a National Node will be needed.

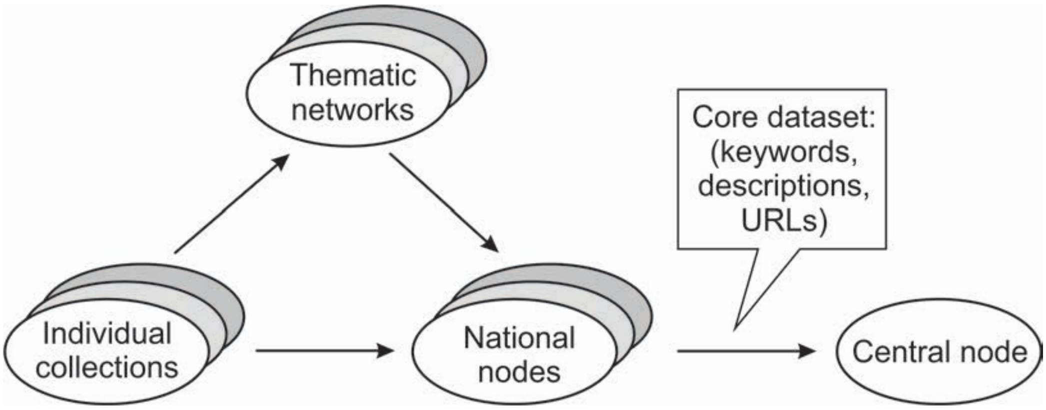

Figure 16: Information flow from individual collections to Central Node

The National Nodes are to co-ordinate networking at the national level, but to provide a common access point and a conceptional framework for the European system, a central system ("Central Node") is needed (which, however, has to be kept at a minimum to ensure sustainability). To make the information gathered by the National Nodes useful for the European system, standard protocols for metadata content and usage must be applied. This is most efficiently possible by providing the National Nodes with software which communicates with the central system's thesauri etc. and which can be accessed by the central system for information retrieval. The software has to be designed in an easy-to-extend fashion, using widely available software as its base, so that National Nodes can easily extend their activities.

The National Nodes should be hosted by organisations committed to research or information provision on a national level, e.g. institutes being part of national academies, or the academies themselves, organizations representing the national clearing house mechanism, or other agencies with an obligation of fostering the collection and maintenance of biological data. It is not mandatory, but of advantage, if they are information providers and/or users themselves, thus being beneficiaries of the improved information access the European Service will provide. They will take on the responsibility of hosting the national meta-database and setting up a website.

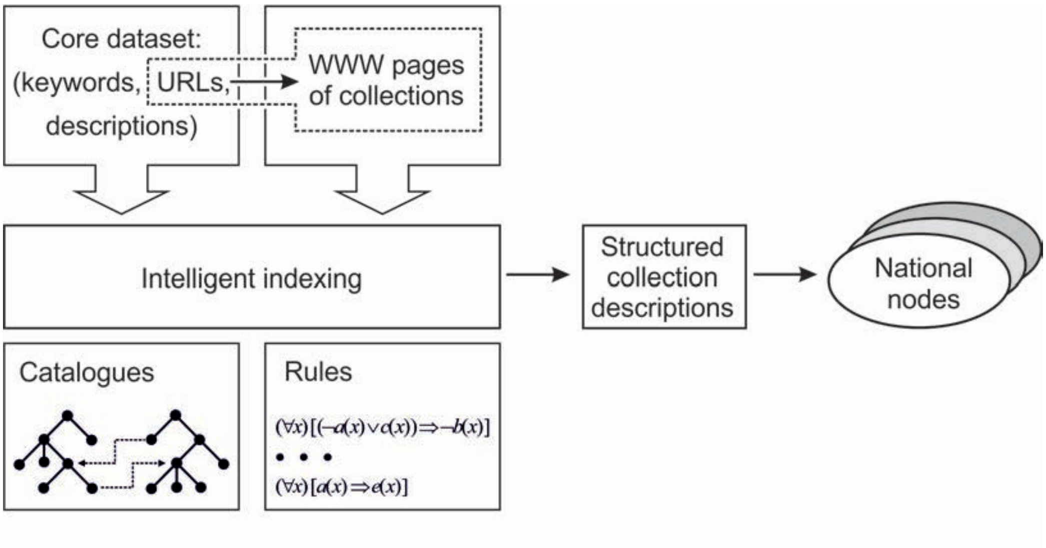

Figure 17: Feedback of information from Central Node towards National Nodes

Communication with the Central Node includes keeping up an interface to supply the necessary core data to perform searches over all connected databases (fig. 15), and to feed back enhanced quality data to the national system (fig. 16). The enhanced data result from the Central Node's application of knowledge processing and advanced indexing tools and the addition of information from different sources. Links to already operating thematic networks will be established, as demonstrated by the BioCISE project.

Automatic extraction of keywords. The efficiency of information access will to a large extent depend on the efficiency of linking metadata items in the thesauri to the collections or sub-collections. The central node will receive information from the national nodes in the form of core attributes and free text descriptions. One of the core attributes is the URL of the collection's website (if any). This can be used to implement web-robot techniques to analyse the websites and automatically extract keywords. For this, an advanced multilingual free text indexing tool has to be developed. Rule based techniques will then be used to work on both, the information retrieved from national nodes and that from Websites, to generate value-added indices. These can be used directly in the central user interface, but should also be communicated back as added value to the national nodes.

User access. Users will be able to choose between access via the European Service, National nodes, thematic networks as well as access to individual collections (fig. 17). Where available, the Service and the national nodes will offer these choices to the user.

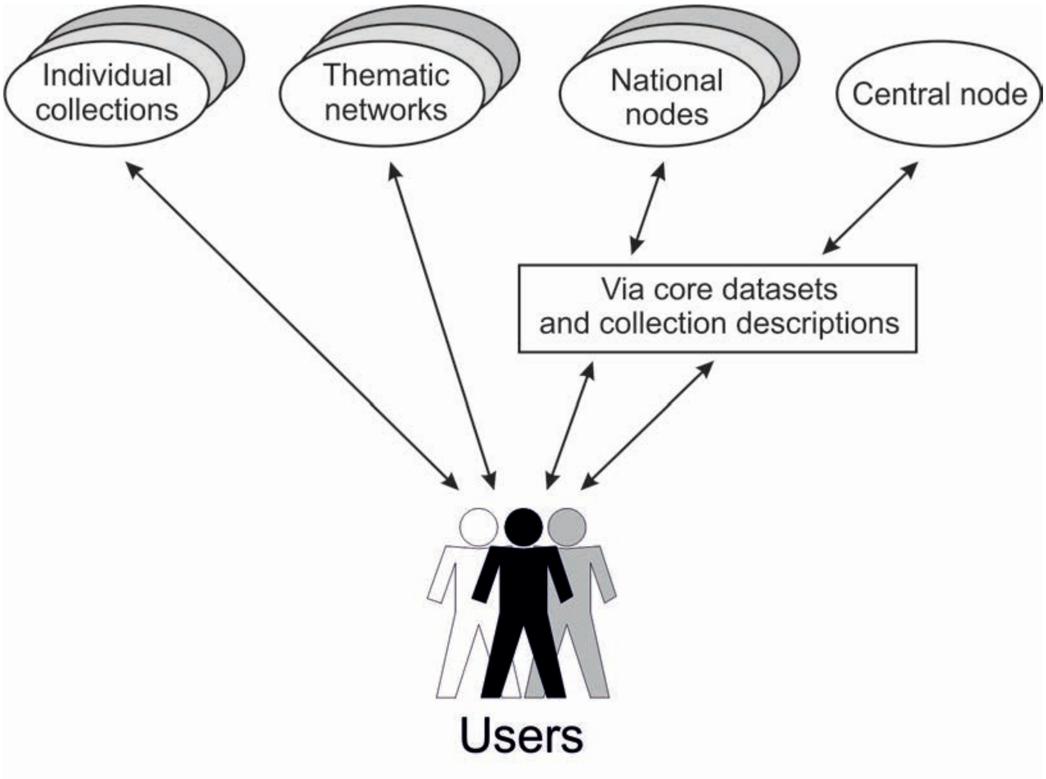

Figure 18: Access to collection information

Sustainability. Prospects for long term funding for the operation of a Biological Collection Information Service on the international scale are rather bleak. On the other hand, among participants in the various workshops and other providers and users interviewed, consensus prevailed that the service itself should be free of charge for the user. Such a facility certainly represents an infrastructure providing access to specimen data, which are the results of more than two centuries of mostly government-financed scientific endeavour. Furthermore, many of the users would be information providers as well. Taking over of marginal costs was still seen as a possibility, as was charging commercial users of the service. However, the general feeling was that the administrative and IPR-related complications inherent to that approach by far outweighed the income generated. For example, several of the thematic networks which are accessible through BioCISE provided their data with the explicit condition that they should be available free of charge.

Again, flexibility of the system is decisive: it has to be able to survive changes in its administrative set-up and location as well as changes in on the information providers side. The system software should require minimum maintenance; once installed, the maintenance and provision of base data will lie in the responsibility of national and thematic networks driven by user and provider communities. The BioCISE project has already established links to 5 national and thematic networks, and the set-up of other National Nodes has been organised in view of pending project proposals. Established contacts to European agencies and the CHM, and liaison with the Consortium of Large-scale European Taxonomic Facilities (CETAF) further broaden the base for attempts to achieve long-term sustainability.

Summary

Collection information serves as a basis for biodiversity research, which is part of the obligations incurred by government in the context of international conventions. Interconnecting such databases of a variety of nations and scientific research topics is a justified cause to involve international research funding agencies. A biological collection information service integrating the full spectrum of resources would play an important role in the Global Biodiversity Information Facility (GBIF 2000) envisioned by the OECD Megascience Forum Working Group on Biological Informatics (Edwards 1999).

The creation of a Biological Collection Information Service in Europe is believed to be a feasible goal if an approach relying on scalable metadata provision through a network of national nodes is used. The major initial contributions an international implementation project would have to make are the following:

- Design of a model for the accommodation of metadata items and their relationships in extendible structured thesauri to be used in information access and indexing.
- Data acquisition for thesauri for taxonomic, geo-ecological and other collection-related data areas from existing sources.
- Design and implementation of a rule-processing software for the automated generation of keywords.
- Setup and development of national nodes; provision of software, where necessary.
- Standardization of core data for the harmonization of data flows between the nodes (content and protocol standardization, e.g. by means of XML data definitions).
- Technical implementation of the information service integrating the metadata model, thesauri and provided data into a user-friendly access system.
- Prototypical integration of individual collection and observation records.

A project period of 3 years with adequate resources is deemed necessary to achieve a functioning service. Questions of sustainability of a collection information service, the adequate consideration of intellectual property rights, and approaches towards data quality standards will have to undergo continued discussion.

References

ABRS 1998: The Global Taxonomy Initiative: Shortening the distance between discovery and delivery. Australian Biological Resources Study, Environment Australia. – Canberra

Alkin, R. 1998: Effective management and delivery of biodiversity information. – Pp. 87-102 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Allmon, W. D. 1997: Collections in Paleontology. – In: Paleontology in the 21st century. Reports and Recommendations. – [http://www.nhm.ac.uk/hosted_sites/paleonet/paleo21/ccep.html]

Anderson, J. E. 1998: Handling the information explosion: The challenge of data management. – Pp. 27-35 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Anon. 1998 [10 Oct]: The Dublin Core: A simple content description model for electronic resources. – In: Dublin Core Metadata Initiative. [http://purl.oclc.org/dc/index.htm].

Anon. 2000 [28 Mar]: FAQ. – In: Dublin Core Metadata Initiative. [http://purl.oclc.org/dc/index.htm].

ASC 1993: An information model for biological collections (draft). Report of the Biological Collections Data Standards Workshop, August 18-24, 1992. Association of Systematic Collections, Committee on Computerization and Networking. [gopher://kaw.keil.ukans.edu:70/11/standards/asc].

Batterbee, R. W. 1981: Changes in the diatom microflora of a eutrophic lake since 1900 from a comparison of old algal samples and the sedimentary record. – Holarct. Ecol. 4: 73-81.

Beach, J. 1998: Responses from Taxacom and ZBIG-L e-mail lists to a query asking scientists (primarily) what their most useful query would be of a networked set of biological collection catalogues. (May, 1998). ZBIG Meeting II, 4 June 1998. – Unpublished.

Berendsohn, W. G. 1997a: CDEFD publications. [http://www.bgbm.fu-berlin.de/CDEFD/]

Berendsohn, W. G., Anagnostopoulos, A., Hagedorn, G., Jakupovic, J., Nimis, P.L., Valdés, B., Güntsch, A., Pankhurst, R. J. & White, R. J. 1999a: A comprehensive reference model for biological collections and surveys. – Taxon 48: 511-562.

Berendsohn, W. G. 1999b: Digital specimens / Digitale Belege. – Berlin. [http://www.bgbm.fu-berlin.de/biodivinf/Projects/DigitalSpecimens.htm]

Berendsohn, W. G. (ed.) 1999c: International gazetteers. – In: TDWG & BioCISE: Standards, information models, and data dictionaries for biological collections. – Berlin. [http://www.bgbm.fu-berlin.de/TDWG/acc/Referenc.htm]

BIOSIS 1999: Biological informatics. – York. [http://www.york.biosis.org/zrdocs/zoolinfo/biol_inf.htm]

Bisby, F.A. 1998. Putting names to things and keeping track: the Species 2000 programme for a coordinated catalogue of life. – Pp. 59-68 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Blackmore, S. 1998: The life sciences and the information revolution. – Pp. 441-450 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Chapman, A. D. 1992: Quality control and validation of environmental resource data. - In: Data quality and standards, proceedings of a seminar organised by the Commonwealth Land Information Forum, Canberra December 1991. – Canberra. [Also published electronically at: http://www.erin.gov.au/life/general_info/validation_1.html]

Cook, L. (ed.) 2000: The national plant collections directory 2000. National Council for the Conservation of Plants and Gardens. – Woking.

Cotterill, F. P. D. 1999: Towards exorcism of the ghost of W. T. Thistleton-Dyer: a comment on “over-duplication” and the scientific properties, uses and values of natural science specimens. – Taxon 48: 35-39.

Diederich, J., Fortuner, R. and Milton, J. 1998: A general structure for biological databases. – Pp. 47-58 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Duckworth, W. D., Genoways, H. H. & Rose, C. L. 1993: Preserving natural science collections: Chronicle of our environmental heritage. - Washington.

EEA 1999: European Environment Agency - European Topic Centre on Catalogue of Data Sources – Hannover. [http://www.mu.niedersachsen.de/cds/]

ENHSIN 2000: ENHSIN – European Natural History Specimen Information Network. [http://www.nhm.ac.uk/science/rco/enhsin/]

Edwards, J. L. 1999: The Global Biodiversity Information Facility: an international Network of interoperable biodiversity databases. – ASC Newletter 27(3-4): 6-7.

Farr, E. & Zijlstra, G. (ed.) 1999: Index nominum genericorum (plantarum). – Washington. [http://www.nmnh.si.edu/ing/]

Freiberg, H. (ed.) 1999: Projekt Naturdetektive. – Deutscher Clearinghouse Mechanism des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit, Schulen ans Netz e.V. und DigiVision GbR. [http://www.naturdetektive.de/]

GBIF 2000: GBIF – The global biodiversity information facility. – Stuttgart. [http://www.gbif.org/]

Gollin, M. A. 1999: New rules for natural products research. – Nature Biotechnology 17: 921-922.

Greuter, W., McNeill, J., Barrie, R., Burdet, H.-M., Demoulin, V., Filguerias, T. S., Nicolson, D. H., Silva, P.C., Skog, J. E., Trehane, P., Turland, N. J. & Hawksworth, D. L. (ed.) 2000: international code of botanical nomenclature (Saint Louis Code). – Regnum Veg. 138.

Grube, M. & Nimis, P.L. 1997: Mediterranean lichens on-line. – Taxon 46: 487-493.

Güntsch, A. & Vander Velde, A. 1998 [Nov]: Collaboration of BIODIV and BioCISE – [http://www.bgbm.fu-berlin.de/biocise/TheProject/IntroCollab.htm].

Güntsch, A., Hahn, A. & Berendsohn, W. G. 2000: Repräsentation der Struktur biologischer Sammlungen als Grundlage für die Schaffung eines europäischen Metainformationssystems. – Pp. 37-51 in: Kramer, R. & Hosenfeld, F. (ed.): Umweltdatenbanken im Web, Workshop-Beiträge und Ergebnisse. – Karlsruhe.

Heaney, M. 2000 [14 Jan]: An analytical model of collections and their catalogues. – Oxford. [http://www.ukoln.ac.uk/metadata/rslp/model/amcc-v31.pdf]

Heywood, C. A., Heywood, V. H. & Wyse Jackson, P. 1990: International directory of botanical gardens V. – Königstein.

Holmgren, P. K. & Holmgren N. H. 1990: Index Herbariorum. Part 1: Herbaria of the World. – New York. [Also: http://www.nybg.org/bsci/ih/]

Hoppe, J. R., Boos, E. & Gottsberger, G. 1996: The database system SysTax - an aid for systematics and taxonomy and the management of botanical gardens and herbaria. –Albertoa 4 : 107-108

ICZN 1999: International Commission on Zoological Nomenclature (ed.): International code of zoological nomenclature. Fourth Edition. – London.

IOPI 1999: International Organization for Plant Information, provisional global plant checklist. – Berlin. [http://www.bgbm.fu-berlin.de/IOPI/GPC/]

ISIS 1999: International Species Information System. – [http://www.isis.org/]

Juggins, S., Flower, R. J. & Battarbee, R. W. 1996: Palaeolimnological evidence for recent chemical and biological changes in the U.K. acid waters monitoring network sites. – Freshw. Biol. 36: 203-219.

Lane, M. A. 1996: Roles of natural history collections. – Ann. Missouri Bot. Gard. 83: 536-545

Lane, M. A. 1998: Biological informatics – weaving a web of wealth. – Canberra.

LTER 1999: U.S. Long Term Ecological Research Network. – Albuquerque. [http://lternet.edu/]

MDA 1997: Museums and information technology - uses of information technology in museums. – Cambridge. [http://www.open.gov.uk/mdocassn/mti_rep1.htm]

Naumann, C. M. & Greuter, W. 1997: Naturwissenschaftliche Forschungssammlungen in Deutschland: Die biologischen Sammlungen. Funktion, Situation, Perspektiven. Denkschrift im Auftrag der Direktorenkonferenz naturwissenschaftlicher Forschungssammlungen Deutschlands (DNFS). – Unpublished.

OECD 1999: Report of the OECD Megascience Working Group on Biological Informatics. – Paris. [http://www.oecd.org/dsti/sti/s_t/ms/index.htm]

Powell, A. (ed.) 1998 [Oct]: Metadata - collection level description – Collection Description Working Group, work in progress. – [http://www.ukoln.ac.uk/metadata/cld/wg-report/].

Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. & Lorensen, W. 1991: Object-oriented modeling and design. – London.

Schalk, P. H. & Los, W. H. 1998: The need to rebuild our university education systems on an information technology basis. – Pp. 395-398 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Schumann, G. L. 1998: Electronic teaching aids for students and practitioners. – Pp. 347-357 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Scott, P. R. 1998: The incredible pace of change: Information technology in support of plant pathology. – Pp. 1-14 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Sieders, G. 1998: Plantenbladeren als sleutel tot duizendjarig CO_2-archief. – Bionieuws 19: 5.

Smith, I. M. 1998: Keeping pathogens in their place: International plant quarantine databases. – Pp. 69-77 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. – Oxon, New York.

Soberón, J., Llorente, J. & Benítez, H. 1996: An international view of national biological surveys. – Ann. Missouri Bot. Gard. 83: 562-573.

Swengal, F. [undated]: Global Zoo Directory. – [http://www.cbsg.org/gzd.htm]

Ter Braak, C. J. F. & van Dam, H. 1989: Inferring pH from diatoms: a comparison of old and new calibration methods. – Hydrobiologia 178: 209-223.

The Plant Names Project 1999: International Plant Names Index. [http://www.ipni.org].

Van Dam, H. 1996: Partial recovery of moorland pools from acidification: indications by chemistry and diatoms. – Netherlands J. Aqua. Ecol. 30: 203-218.

Van Dam, H. & Beljaars, K. 1984: Nachweis von Versauerung in west-europäischen kalkarmen stehenden Gewässern durch Vergleich von alten und rezenten Kieselalgenproben. – Pp. 184-188 in: Wieting, J., Lenhart, B., Steinberg, C., Hamm, A. & Schoen, R. (ed.), Gewässerversauerung in der Bundesrepublik Deutschland. – Berlin.

Van Dam, H. & Mertens, A. 1993: Diatoms on herbarium macrophytes as indicators for water quality. – Hydrobiologia 267/270: 437-445.

Vieglas, D. 1999: Integrating disparate biodiversity resources using the information retrieval standard Z39.50. TDWG 1999 Abstracts. – Cambridge, USA, [http://www.tdwg.org/rep1999.html#dave]

Wagner, F. 1998: The influence of environment on the stomatal frequency in Betula. – LPP contribution series 9.

Williams, N. 1997: Drug firms back move to link databases. – Science 277: 902.

The acquisition, cultivation, preservation, and storage of objects in biological collections is an integral part of biological research in many sub-disciplines. Field and research notes and specimen labels contain valuable and detailed data and the object itself can be a physical resource for research and industry.
This publication reports on the concluded concerted action project BioCISE, which set out to identify and analyse biological collection information and its environs with the aim to prepare a sound collaborative technical and structural base for a Biological Collection Information Service in Europe and a strategy for its implementation.

ISBN 3-921800-44-7