

## USING GEOGRAPHICAL AND TAXONOMIC METADATA TO SET PRIORITIES IN SPECIMEN DIGITIZATION

WALTER G. BERENDSOHN AND PEGGY SELTMANN

*Dept. of Biodiversity Informatics and Laboratories, Botanic Garden und Botanical Museum Berlin-Dahlem, Freie Universität Berlin, Königin-Luise-Straße 6-8, D-14195 Berlin, Germany*

*Correspondence e-mail: w.berendsohn@bgbm.org*

**Abstract.** – Digitizing the information associated with by specimens in natural history collections is a key endeavor providing falsifiable information about past and present biodiversity on a global scale, for application in a variety of research fields far beyond the current application in biosystematics. Existing digitization efforts are driven by individual institutional necessities and are not coordinated on a global scale. This has led to an over-all information resource that is patchy in taxonomic and geographic coverage as well as in quality. Digitizing all specimens is not an achievable aim at present, so priorities need to be set. Most biodiversity studies are both taxonomically and geographically restricted, but access to non-digitized collection information is almost exclusively by taxon name. Creating a “Geotaxonomic Index” providing metadata on the number of specimens from a specific geographic region belonging to a specific higher taxonomic category may provide a means to attract the attention of researchers and governments towards relevant non-digitized holdings of the collections and set priorities for their digitization according to the needs of information users outside the taxonomic community.

**Key words.** – Natural history collections; collections; specimens; specimen data; metadata; digitization; GBIF; biodiversity research.

### INTRODUCTION

Each specimen in natural history collections carries a wealth of information, most notably the past location in space and time of a verifiably identified species. Specimens add the historical component of biodiversity to contemporary observation networks, and deposited voucher specimens are reproducible scientific evidence for species identifications for all areas of biodiversity research.

General agreement exists that having all this information available in electronic form would greatly improve the information base for many research domains, including – but by no means restricted to – systematics.

Substantial efforts have been invested by institutions over the past years in digitizing specimen information, and the Global Biodiversity Information Facility (GBIF) offers the technical infrastructure to make these data records universally available. However, the GBIF Task Group on the Digitization of Natural History Collections realized that

- we are very far from a complete data resource: most specimen data remain accessible only by consulting the actual specimen
- digitizing individual specimens is a very costly process and no funding for globally comprehensive specimen digitization is in sight
- existing digitization efforts are not coordinated, producing an information resource that is patchy in taxonomic and geographic coverage as well as in quality (e.g. Yesson 2007, Balian & al. 2008, Kusber & al. 2009).
- no mechanism exists to request globally information about relevant non-digitized holdings of collection institutions. The potential of specimen information is thus not appreciated by a wider user community.

We posit that we may overcome these obstacles by making user demand the driver of detailed digitization of individual specimens. User demand for detailed specimen data comes from ongoing or projected research or is connected to

socio-political demands (conventions, repatriation, environmental laws, etc.). If institutions lack the resources to fulfill this demand, costs of digitization activities should be covered by the agencies funding the research or requiring the data, i.e., the costs should either be incorporated into research proposals, or be covered by organizations, foundations, or governments with direct interest in the specific area of biodiversity. However, at present neither these costs nor the potentially relevant specimen holdings and their whereabouts can be assessed by users in a comprehensive way.

### THE VISION

To remedy this situation, collection-holding institutions will need to co-ordinate their efforts on national and regional, if not global levels, and agree to implement the necessary mechanisms. Estimating the costs involved in specimen digitization (separately for taxonomic verification, exact georeferencing, and imaging) is one of the tasks. The creation of a high-level metadata catalogue is another. The single most important data item besides the taxonomic identification of specimens is the geographic location of the collection event. This conclusion is based on a preliminary survey of (non-taxonomic) scientific literature reporting research results based on natural history specimen data. Biodiversity research has a strong spatial component; researchers are often investigating biotic changes in specific geographic locations. Examples include species distribution modeling (e.g. Longmore 1986), predicting new species distributions (e.g. Raxworthy et al. 2003), invading organisms (e.g. Suarez et al. 2001; Peterson and Vieglais 2001), biological control of pests (e.g. Soberón et al. 2001), habitat loss (e.g. Pergams and Nyberg 2001), or impact of climate change on species distributions (e.g. Peterson et al. 2002, 2004; Parmesan et al. 1999) and on plant physiology (e.g. Peñuelas and Estiarte 1997). Although geographic data in museum collections have shortcomings, resulting mainly from the ad hoc nature of collecting efforts, collection inventories can be used to identify gaps and prioritize future field work. In combination with observation and ecological data, collection data serve to identify

global priority areas for conservation action based on taxonomic diversity (Vane-Wright et al. 1991; Pressey et al. 1993).

However, current access to non-digitized specimens in collections is almost exclusively by taxon name. This is one of the key impediments to effective use of collection information outside the taxonomic community.

To overcome this impediment, we envision a “Geotaxonomic Index“, with metadata consisting of geographical and taxonomic categories on a rough scale combined with the corresponding number of specimens in the respective collection (fig. 1). This Index will enable researchers and other interested parties to identify potentially relevant holdings; request further details (e.g. the temporal component); request the detailed digitization of the information needed; obtain the actual costs (if any) of necessary digitization efforts; and incorporate, where necessary, the costs in their research proposals or budgets. In this manner digitization will be directly driven by research (or political) needs, in addition to the current approach, which is largely driven by individuals or administrative procedures (e.g. loans) of the collection institutions. Once digitized, the data will be made freely available.

It has been argued (Townsend Peterson, pers. comm.) that “a better focus than simple numbers-oriented metadata schemes will be on identification of the most important collections via expert knowledge.” We posit that this may certainly be true when a taxonomic focus is taken (i.e. the taxonomic expert makes the assessment). Even for taxonomic specialists this assessment may be strongly biased, largely depending on the custodial activities and their publicity at the collection institutions themselves. From a geographical point of view, we don’t think that many experts can make a meaningful assessment of the contents of collections and their relevance beyond a few key institutions. We simply have to concede the fact that the content of many of our collections is poorly known, especially of the large collections in the North. So a number-oriented scheme seems to be the only practical global approach.

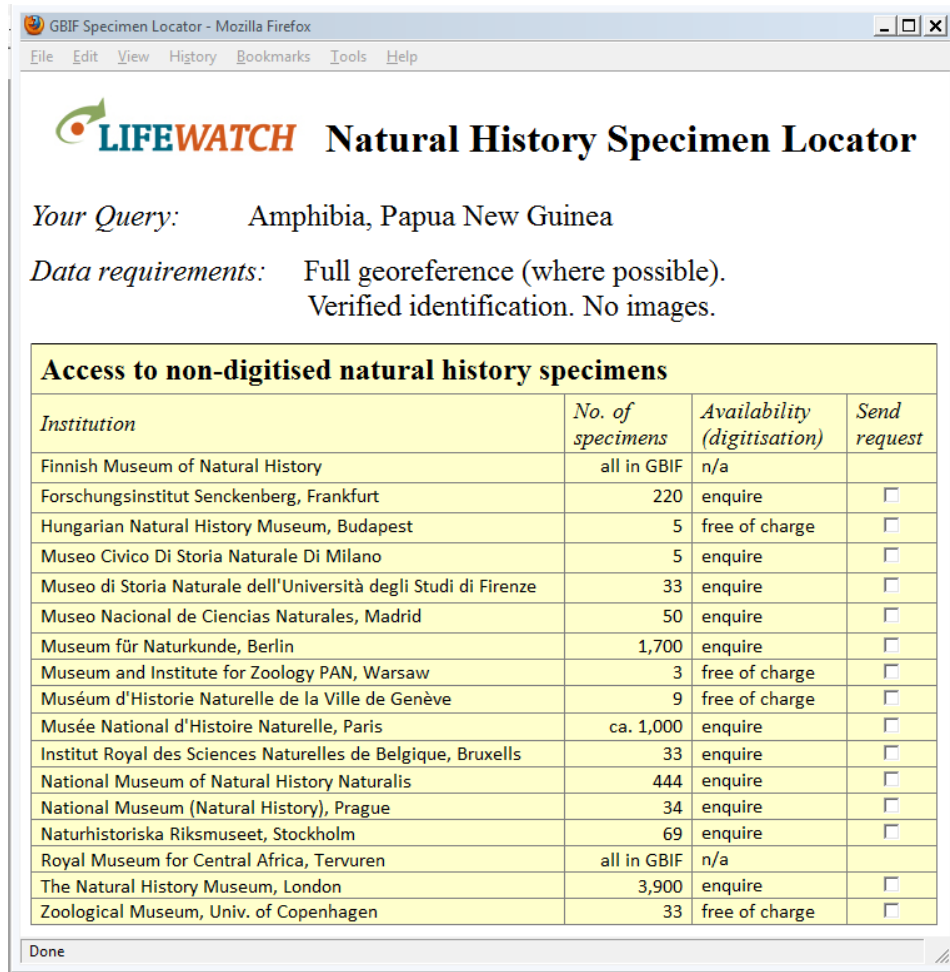


Figure 1. Mock-up of web query service for collection data discovery. The European LifeWatch infrastructure initiative has included this approach to collection metadata in its work plan. Please note that although a European collection directors meeting supported the approach, the depicted setup is entirely imaginative and the cited collections (names taken from the CETAF membership list) have not been consulted with regard to their holdings!

Apart from individual researchers, countries of origin desire to gain access to digitized data on specimens collected on their territory. A long-standing example for such efforts is the data repatriation activity of the Mexican National Commission for the Knowledge and Use of Biodiversity (CONABIO). From the start in 1992, CONABIO realized the relevance of integrating repatriated data from a number a foreign scientific collections. CONABIO provided funds to some researchers to visit foreign scientific collections to collect data, or established agreements to obtain images of specimens to digitize specimen data in Mexico. CONABIO has achieved the digitization

of about 1,200,000 records from specimens collected in Mexico and held by foreign natural history museums, mainly in the US, plus 26 countries (Koleff 2010, pers. comm.). These data have contributed significantly to CONABIO's achievements in areas such as predicting the threat of invasive species, the spread of disease vectors, risk-assessment of GMO crop introductions to wild relatives, regional prioritization of environmental decision making, and conservation efforts involving rural communities, to name but a few (Soberón and Koleff, 1999; Koleff et al. 2004; CONABIO 2005, 2007, 2010).

Other countries that wish to take action similar to CONABIO would greatly profit from an

information source like the Geotaxonomic Index. They would be able to rather precisely calculate the costs of such an endeavor. Funding could be achieved using bilateral agreements between countries, as well as by tapping into international resources such as the Global Environmental Facility. In both cases, individual researchers and national organizations concerned with biodiversity of their countries, the existence of the Geotaxonomic Index would constitute a major step forward, providing the base for targeted information requests. In contrast to the digitization of all specimens establishing the Index may also be feasible financially, because the cost of metadata capture should be several scales lower than that of digitizing data for individual specimens.

#### **FEASIBILITY OF THE APPROACH**

A basic consideration with respect to collections is that, because of the long-term perspective of natural history institutions, they also provide a solid and long-term sustainable base for the establishment of biodiversity information networks. We therefore posit that the information resource itself should be as decentralized as possible, sitting as close as possible to the actual physical collection. This is exemplified by the current GBIF network and needs only to be extended to cover metadata provision.

It is also clear that the creation of the Geotaxonomic Index has to be driven by the natural history institutions themselves as a straightforward extension of their historical role as biodiversity information centers. Unfortunately, the compilation of the necessary metadata will still require a substantial effort, and institutional resources are on average very scarce. Up to now, all attempts to interest funding organizations in this task as an enabling, one-time activity have failed. However, to underpin applications and to continue this quest we need to obtain more data on the size, cost and feasibility of the approach.

Preliminary studies as to the feasibility and the effort needed to produce the metadata catalogue have been undertaken in Germany. P. Seltsmann interviewed 15 curators of collections during visits to 10 institutions in Germany and one in the UK: Natural History Museum Berlin, State Museum of Natural History Görlitz – Botany collections,

Zoological Collections of the Martin Luther University Halle-Wittenberg, Alexander König Research Museum in Bonn, Julius Riemer Museum for Natural History and Ethnology in Wittenberg, Bavarian State Collections for Zoology and Bavarian State Collections for Botany in Munich, German Entomological Institute in Müncheberg, Botanic Garden and Botanical Museum Berlin-Dahlem and Natural History Museum in London. Although by no means representative, the institutions and collections were selected to cover a wide range of sizes, organizational models, and taxonomic groups.

Discussions also took place in the context of the incorporation of a metadata catalogue in the master plan for the European LifeWatch research infrastructure (Tack 2009) during several meetings of the preparatory phase project. Further input was obtained from a workshop held in Leiden (Berendsohn & al. 2009) in the context of the meeting of the Society for the Preservation of Natural History Collections in July 2009 and a meeting of the Directors of Collections committee of the European Distributed Institute of Taxonomy in June 2008. Both meetings were supporting the creation of a Geotaxonomic Index. Finally, a working session during the 2009 meeting of the organization for Biodiversity Information Standards (TDWG) discussed the technical implementation of the Geotaxonomic Index. The participants agreed that the Index should be linked to the Biodiversity Collections Index (Hyam 2008), which in turn will be integrated with the Global Biodiversity Resource Discovery Service (GBRDS) being set up by GBIF. The draft of the Natural Collections Descriptors standard (NCD v. 09, Thomson & al. 2008) adequately covers the metadata items needed, with the exception of a quantitative descriptor (number of specimens, samples or batches), which needs to be included in the standard (see Bourgoin & al. 2009 for details).

#### **PRELIMINARY RESULTS**

##### ***CATEGORIZATION***

Aiming at a strategy to capture metadata in the most time- and cost-effective manner, we first attempted to categorize the enormous diversity of collection holdings according to 4 criteria that

directly influence the methodology of metadata digitization: physical collection type, location of the information, data content of the geographical information, and the type of written information.

*1. Physical collection type.*– Natural history specimens are quite diverse in their physical form, depending on the nature of the sample and on the type of preservation and storage. This strongly influences the accessibility of the specimen’s data. We found it useful to distinguish the following types of collections:

- Dried herbarium collections (“cryptogams”, “phanerogams”)
- Alcohol collections (e.g. pisces, arachnid, centipede, echinoderm, crustacean, hymenoptera collections; partly reptile and amphibian collections)
- Pinned invertebrate collections (e.g. lepidoptera, coleopteran collections)
- Skull, bone, pelt collections, collections including stuffed specimens (e.g. aves, mammal collections, partly reptile collections)
- Paleontological collections
- Microscopic slide collections
- Environmental samples (water, ice, sediment, air or soil samples, results from fogging and trapping containing many different organisms not [yet] sorted or separated).

*2. Data location.*– The location of the relevant geographic metadata (collection localities) is strongly related to collection type, but not unambiguously so. Generally, notes about collection localities are included:

- On specimen labels: This is the case in many collection types, but accessibility of the label varies strongly (e.g. between those on herbarium sheets and those on a needle underneath a pinned insect).
- On index cards: Usage of index cards and the availability of collection localities thereon vary widely among collections and are highly dependent on the holding institution.

- In accessions ledgers (written or computerized): Here, two types have to be distinguished: ledgers that hold detailed specimen data (one entry per specimen) and less detailed ledgers holding metadata (one entry per accession event, e.g. “300 Coleoptera from Brazil and 100 from Ecuador”; “5.000 specimens from Darwin”)<sup>1</sup>.
- In additional folders complementing accession registers (e.g. in the form of label copies): This is practiced in the cryptogam collections of the Bavarian State Collection for Botany and may be practiced in other collections, too.
- In collector’s catalogues and collector’s field books: These are extreme variable in form and content and may not have been deposited with the institution holding the collection.
- By inference. In some cases, other data may indicate geographic information. The most obvious cases are taxa known to be endemic in a certain area. Another lead may be provided by collectors specialized in a defined geographic region who have deposited their material in the collection.
- From curatorial knowledge.
- In Literature. Monographic treatments and some checklists contain specimen data with reference to higher level geographic categories (e.g. countries) and the collection location.

*3. Data content.*– Specimen information is highly diverse as to language and detail. Moreover, geographical categories are partly dependent on the historical context - especially for historical

---

<sup>1</sup> Detailed accessions ledgers are typical for small to middle-sized collections such as pisces, reptile, amphibian or mammal collections, whereas herbarium collections and most parts of insect collections deal with less detailed ledgers – if they exist at all. For example, the entomological collection of the NHM London with approximately 28 Million specimens is keeping such “rough” ledgers whereas the German Entomological Institute in Müncheberg with approximately 5 Million specimens is not; the herbarium collections at the Bavarian State Collection for Botany keep less detailed ledgers but the herbarium collections at the Botanical Garden and Botanical Museum Berlin-Dahlem do not keep accession ledgers that could be used in this context.

specimens, for which it may be impossible to assign geographical metadata even to the level of a present country of origin.

4. *Typescript.*– Handwritten specimen labels or ledger entries may be impossible to discern by unqualified personnel.

#### **FIRST CONCLUSIONS**

In most cases, obtaining really exact metadata will require digitization of individual specimens (thus contradicting the purpose of metadata capture). A certain amount of fuzziness in the metadata holdings has thus to be accepted. From the survey conducted, it becomes clear that the methods to be applied depend strongly on the institutional traditions of record keeping and do vary strongly according to collection type. However, to make the metadata useful enough to become the driver of prioritization, clear aims as to the minimum level of taxonomic and geographic categories have to be defined. Guidelines are lacking at present. On the taxonomic side, this varies from group to group depending on the systematics, for example, for flowering plants family level is useful but for the very large families. So this needs to be fixed group-by-group. For geography, the aspiration should be to consistently reach the level of modern country (with some additional categories to allow capture of larger inclusive terms of historical importance, such as Yugoslavia, Soviet Union, or German East Africa). Consistent data, derived by whatever means including statistical extrapolation, will be useful in the Geotaxonomic Index.

The most timesaving and effective method to obtain geographic information would probably be to filter appropriate data from accession ledgers, where present. The procedure could consist of (i) scanning ledgers using large format book scanners (drawing on the experience from the Biodiversity Heritage Library or similar mass-digitization projects), (ii) for typewritten records: rapid data capture through optical character recognition (OCR), (iii) semiautomatic mark-up of key data (taxon, collection locality, number of specimens); and (iv) processing of data to add standardized taxon and geography levels. Subsequent quality checks should be run against samples of the physical collection and existing databases of digitized individual specimens. An analogue

procedure can be used for card indices containing appropriate information.

Scanning of ledgers has or is being executed in many institutions, for example at the Natural History Museum's Zoology Department in London and the Zoological Museum Alexander Koenig (e.g. for pisces, reptile and amphibian collections). The primary motive has often been the preservation of historical documentation rather than databasing. However, where the first step of scanning has been executed, the digital information is already mobilized and available, which greatly enhances the possibilities for cost effective data capture (e.g. by outsourcing).

Where no written records about the collection exist, other approaches have to be taken, which mainly depend on the type of collection.

(i) In sheet mounted herbarium collections, the information on the label attached to the sheet has to be read to extract the metadata. Leafing through the sheets may in some collections be possible without overly stressing the material. In herbaria using an open storage systems (where sheets do not have to be removed from the containers), leafing through the sheets would make the registration a relatively rapid process (estimated at less than a single person year per million herbarium specimens). Otherwise, every sheet has to be extracted from its storage location and registered. In a first step, an inventory of species names is taken, compiling a list conforming to the sequence of specimens in the herbarium. To that list specimen data can be added. The term for the country, even if expressed in different languages or in handwriting, is normally very easily deciphered, capturing it takes a second, where clearly indicated, where not, the label can be photographed for later processing. However, even with open storage systems it seems appropriate to add a few seconds to take the specimen out and directly obtain an image of the entire specimen including the label. With few exceptions (folded labels etc.), photographing the specimen effectively captures the entire information on the sheet. Higher-end commercial one-shot digital camera can be used, these are now starting to reach image qualities that are sufficient for OCR and feature detection techniques (Steinke, 2008, 2009). First tests at the Berlin herbarium led to an

estimate of less than 30 seconds per specimen for the imaging process (including attachment of a barcode label), with a single person carrying out the work. Adding the effort for the species list and for discerning (most of) the unclear countries of origin, we arrive at an estimate of 6 person-years of largely unqualified work to achieve exact and verifiable metadata for a million individual herbarium specimens. Where herbarium specimens are not mounted on sheets but kept in capsules (carton envelopes, mainly used for mosses, lichens and fungi), the specimen label should be extracted and photographed (as practiced e.g. at the Botanical State Collection in Munich – D. Triebel, pers. comm.). This solution may also be applicable for unmounted herbarium material. Additional data (collectors, collection numbers, descriptions, exact location, annotations, accession numbers, etc.) can be extracted without handling the specimen and thus anywhere in the world, using automated processes or citizen science approaches (using social networks to capture the label content).

(ii) Alcohol and other liquid storage collections usually don't lend themselves to photographic data capture, because the labels are contained within the container or glued to the outside of the differently rounded containers. However, the number of individual specimens is often comparable to those in herbarium collections, so handling specimens individually (again starting with a species list established in sequence of the collection) would not be an insurmountable obstacle. Here, data capture has to rely on the qualification of the personnel to be able to discern geographic locations.

(iii) The situation is quite different in pinned collections, because these are clearly those with the highest specimen numbers, and also because getting at the labels may pose a serious obstacle. In most cases, it involves directly handling the often highly fragile specimens. Especially in old collections, data capture often has to be accompanied by curatorial measures to replace fragile pins, etc. Obviously, capture of metadata from individual specimens is not normally the way forward here. Indirect evidence from taxonomy (endemism) is also not often helpful, because the knowledge of taxon distributions is patchy at best (except, perhaps, in larger butterfly species).

Curatorial knowledge about the collections and indirect evidence from known collector's itineraries are probably the only immediate way forward where no ledgers or other accession information can be gathered. However, many of these collections hold more or less complete species lists, which, together with an assessment of specimen numbers, should be published using the network of the Global Biodiversity Information Facility.

(iv) Skulls, bones, pelts and stuffed animals: Because of the comparatively low numbers of specimens, a direct approach as in liquid storage collections is possible, and often complete databases of such specimens already exist.

(v) For paleontological collections, the idea of a geotaxonomic index as proposed here is probably most difficult to sell, due to the much more complex nature of "locality" in such specimens (as well as additional taxonomic complexity). Metadata restricted to the modern geographic location where the specimens were found are not exceedingly useful, neither for the paleontologist nor for the general user. Further discussion with and within the paleontological community is necessary.

(vi) Microscopic slide collections. For many groups of organisms, these form the most important evidence in natural history collections apart from environmental samples. In many cases, these are well documented and the existing inventories could be used to extract geographic information. However, access to the information of historical collections may be limited or face serious problems (e.g. Lazarus & Jahn 1998).

(vii) Environmental samples. Here we group a number of different collection holdings that are distinguished by their collective content: many specimens and species, but usually a single location. This may contain water, ice, soil, sediment, and air samples or the unsorted results from light-traps, fogging or other trapping methods. Geographic access to the very existence of such materials may be of high importance for researchers.

Summarizing, it is important to stress that huge differences exist in data storage and management and thus, in data availability between



collection types, and also with respect to the same collection types in different institutions. The conclusions presented are clearly very preliminary and need further details and discussion in order to serve as the base for a plan for global action. Apart from the technical discussion, three further points need to be considered:

First, the Geotaxonomic Index can only be successful if it is supported by a substantial number of institutions. Although there is considerable interest on the side of the institutions to better document their holdings, most institutions do not have the resources to carry out the task. Clearly, research funding should not and will not support such an endeavor. Additional funding has to come from research infrastructure programmes, where biodiversity has up to now had only limited success compared to other research communities. An encouraging sign has been the funding of a preparatory phase of the large scale LifeWatch initiative and of several smaller but substantial infrastructure projects by the European Union. The competition for such funding is fierce; a prerequisite for success is to close ranks across disciplinary boundaries within the biodiversity research community.

Second, it has to be verified if an empirical approach may offer an alternative pathway. One of the reviewers suggested using the existing digitized data from GBIF to locate gaps and identify discrepancies between current sampling from ad-hoc digitization and known species richness. We doubt that this is going to produce a base for prioritization at this stage, given that only about 60 Million specimen records (of an estimated 2-3 billion total) are available. Moreover, even if the gaps can be mapped as suggested it is rather naïve to expect governments and funding agencies to go out and “ask collections to provide assessments of their holdings that show how those collections can fill the gaps”. In contrast, we have (though admittedly anecdotal) evidence that detailed digitization can be funded as part of the information gathering effort within a research project, or in response to an information request from a government (with the demand from developing countries now becoming more vociferous, see for example Figueiredo & Smith 2010),

Third, the “payoff” of metadata capture relative to partial full data capture needs to be considered (Townsend Peterson, pers. comm.). With respect to herbarium sheet collections (see above), we have concluded that full data capture by means of photographic images is the most promising way ahead, with the Geotaxonomic Index produced as a first side effect. In general, the Task Group concludes that metadata capture should not inhibit detailed data capture (see Berents & al., this issue). So where institutional resources need to be used, the effort to obtain metadata rich enough to guide digitization efforts has to be weighed against the benefits of actual detailed capture of biodiversity data. However, we think that – in contrast to current practice – the proposed approach to metadata capture may make additional external funding available for data mobilization.

#### ACKNOWLEDGEMENTS

We gratefully acknowledge the corrections, suggestions and criticisms provided by 3 reviewers as well as the “Comments and Reflections on the GSAP Report” given by Townsend Peterson on the Report for the GBIF Science Committee of the GBIF Task Group on the Global Strategy and Action Plan for Mobilization of Natural History Collections Data. Data gathering was made possible by the European LifeWatch Preparatory Phase project (FP7 INFRA-2007-2.2-01, contract 211372) and the support of the Freie Universität Berlin for the German national node of the Global Biodiversity Information Facility.

#### REFERENCES CITED

- Balian, E. V., Segers, H., Lévêque, C. and Martens, K. 2008. Freshwater animal diversity assessment: an overview of the results, *Hydrobiologia*, 595: 627-637. [doi:10.1007/s10750-007-9246-3].
- Berendsohn, W. G., Berents, P. and Macklin, J. A. 2009. Report of the Leiden Workshop on digitisation priorities<sup>1</sup>. Annual Meeting of the Society for the Preservation of Natural History Collections. Leiden. [Accessed September 1, 2010]
- Bourgoin, T., Berendsohn, W. G. and Macklin, J. A. 2009. Report of the Working Session on "The Natural History Geotaxonomic Index"<sup>2</sup>. TDWG

<sup>1</sup> [http://www2.gbif.org/GSAP-NHC\\_SPNHC09\\_Report.pdf](http://www2.gbif.org/GSAP-NHC_SPNHC09_Report.pdf)

<sup>2</sup> [http://www2.gbif.org/GSAP-NHC\\_TDWG09\\_WorkingSession.pdf](http://www2.gbif.org/GSAP-NHC_TDWG09_WorkingSession.pdf)



- 2009 Annual Conference. Montpellier. [Accessed September 1, 2010]
- CONABIO 2005. Informe de actividades 1992-2004.<sup>1</sup> Mexico. [Accessed December 23, 2010]
- CONABIO 2007. Informe de actividades 2005-2006.<sup>2</sup> Mexico. [Accessed December 23, 2010]
- CONABIO 2010. Informe de actividades 2007-2009.<sup>3</sup> Mexico. [Accessed December 23, 2010]
- Figueiredo, E. and Smith, G. F. 2010. The colonial legacy in African plant taxonomy.<sup>4</sup> *South African Journal of Science*, Vol 106(3/4), Art. #161, 4 pages. [Accessed August 28, 2010]
- Hyam, R. D. 2008. The Biodiversity Collections Index – Linking 140 Million Records? P. 18 in: Weitzman, A.L., and Belbin, L. (eds.): *Proceedings of TDWG (2008)*, Fremantle, Australia.
- Koleff, P., Fernández C., Martínez J.M. and Moreno, E. 2004. Información sobre la biodiversidad de México en el extranjero.<sup>5</sup> *Biodiversitas* 54 :1-7. [Accessed December 23-]
- Kusber, W.-H., Zippel, E., Kelbert, P., Holetschek, J., Güntsch, A. and Berendsohn, W. G. 2009: From cleaning the valves to cleaning the data: Case studies using diatom biodiversity data on the Internet (GBIF, BioCASE). *Studi Trent. Sci. Nat.* 84: 111-122.
- Lazarus, D. B. and Jahn, R. 1998. Using the Ehrenberg Collection. *Diatom Research*, 13(2):273-291.
- Longmore, R. 1986: *Atlas of Elapid Snakes of Australia*. Australian Flora and Fauna Series 7.
- Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J.K., Thomas, C.D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., T Tennent, J., Thomas, J.A., Warren, M. 1999. Poleward shift of butterfly species' ranges associated with regional warming. *Nature* 399: 579-583.
- Peñuelas, J. and Estiarte, M. 1997. Trends in plant carbon concentration and plant demand for N throughout this century. *Oecologia* 109: 69–73.
- Pergams, O.R.W. and Nyberg, D. 2001. Museum collections of mammals corroborate the exceptional decline of prairie habitat in the Chicago region. *Journal of Mammalogy* 82: 984-992.
- Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H. and Stockwell D.R.B. 2002. Future projections for Mexican faunas under global climate change scenarios. *Nature* 416: 626-629
- Peterson, A.T. and R. Scachetti-Pereira 2004. Potential Geographic Distribution of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in North America. 151, 170-178.
- Peterson, A.T. and Vieglais, D.A. 2001. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem. *Bioscience* 51: 363-371.
- Pressey, R.L., Humphries, C.J., Margules, C.R., Vane-Wright, R.I., Williams, P.H. 1993. Beyond opportunism: key principles for systematic reserve selection. *Trends in Ecology & Evolution*. 8: 124-128.
- Raxworthy, C.J., Martinez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. and Peterson, A.T. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837-841.
- Soberón, J., Golubov, J. and Sarukhán, J. 2001. The Importance of *Opuntia* in Mexico and routes of invasion and impact of *Cactoblastis cactorum* (Lepidoptera: Pyralidae). *Florida Entomologist* 84: 486-492.
- Soberón, J. and Koleff, P. 1999. The national biodiversity information system of Mexico. In: Raven, P. H. (ed.) *Nature and human society. The quest for a sustainable world*. National Academy Press. Washington, pp. 586-595.
- Suarez, A.V., Holway, D.A. and Case, T.J. 2001. Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from Argentine ants. *Proceedings of the National Academy of Sciences of the United States of America* 98: 1095-1100.
- Steinke, K.-H., Dzido, R., Gehrke, M. and Prätel, K. 2008. Feature recognition for herbarium specimens (Herbar-Digital). *Proceedings of TDWG*. Perth.
- Steinke, K.-H. 2009. Lokalisierung von Schrift in komplexer Umgebung. *Tagungsb. Jahrestagung deutschen Ges. Photogrammetrie*. Jena.

1 [http://www.conabio.gob.mx/institucion/informes/pdf/informe\\_doce\\_anios.pdf](http://www.conabio.gob.mx/institucion/informes/pdf/informe_doce_anios.pdf)

2 [http://www.conabio.gob.mx/institucion/informes/pdf/InfConabio2005\\_2006.pdf](http://www.conabio.gob.mx/institucion/informes/pdf/InfConabio2005_2006.pdf)

3 <http://www.conabio.gob.mx/institucion/informes/pdf/III%20informe%20CONABIO.pdf>

4 <http://sajs.co.za/index.php/SAJS/article/viewArticle/161/231>

5 <http://www.conabio.gob.mx/otros/biodiversitas/doctos/pdf/biodiv54.pdf>

- Tack, J. (ed.) 2009. LifeWatch – e-Science and Technology Infrastructure for Biodiversity Data and Inventories. INBO, Brussels.
- Thomson, N., Hyam, R., Rinaldo, C., Butler, C., Holland, D., Mathé, B., Waibel, G., Addink, W., Altenburg, R. and Döring, M. 2008. Natural Collections Description (NCD) – A data standard for exchanging data describing natural history collections<sup>1</sup>. Biodiversity Information Standards (TDWG). [Accessed July 27, 2010]
- Vane-Wright, R.I., Humphries, C.J. and Williams, P.H. 1991. What to protect? Systematics and the agony of choice. *Biological Conservation* 55: 235-254.
- Yesson C., Brewer P. W., Sutton, T., Caithness, N., Pahwa J. S., Burgess, M., Gray, W.A., White, R. J., Jones, A. C., Bisby, F.A. and Culham A. 2007. How global is the Global Biodiversity Information Facility? *PLoS One* 2/11: 1-10.

---

<sup>1</sup> [http://www.tdwg.org/fileadmin/subgroups/ncd/NCD\\_090.doc](http://www.tdwg.org/fileadmin/subgroups/ncd/NCD_090.doc)