Biodiversity Informatics, 7, 2010, pp.67 – 71

RECOMMENDATIONS OF THE GBIF TASK GROUP ON THE GLOBAL STRATEGY AND ACTION PLAN FOR THE MOBILIZATION OF NATURAL HISTORY COLLECTIONS DATA

WALTER G. BERENDSOHN(1), VISHWAS CHAVAN(2) & JAMES A. MACKLIN(3)

(1) 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

(2) Global Biodiversity Information Facility Secretariat, Universitetsparken 15, 2100, Copenhagen, Denmark.

(3) Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA 02138, USA

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

Abstract. – A Task Group to envision a Global Strategy and Action Plan for the Mobilization of Natural History Collections Data established by the Global Biodiversity Information Facility (GBIF) has formulated three basic recommendations in order to increase the rate of mobilization of natural history collections data and improve the usage of this information resource: (i) GBIF must facilitate access to information about non-digitized collection resources by publicizing the research potential of collections through metadata and assessing the number of non-digitized specimens; (ii) GBIF must work with collections to continue to increase the efficiency of specimen data capture and to enhance data quality by means of technical measures, by means of ensuring attribution and professional credit and influencing institutional priorities, and by engaging with funding agencies; (iii) GBIF must continue to improve and promote the global infrastructure used to mobilize digitized collection data through technical measures, outreach activities and political measures.

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

INTRODUCTION

Mobilizing the biodiversity information intrinsic to the specimen holdings of natural history museums and herbaria of the world was one of the core aims of establishing the Global Biodiversity Information Facility (OECD 1999) and has been an integral part of its work program ever since. The value of specimen data for biodiversity research and natural resources management has been widely recognized (see references in Scoble & Bourgoin, 2010 and Baird, 2010). GBIF has created a functional technical infrastructure to discover and facilitate access to distributed data resources of primary biodiversity data (GBIF, 2008a), including natural history collections data. As of July 2010, GBIF facilitates access to more than 201 million such primary biodiversity data records, about 25.7% of which are specimen data (GBIF, in prep) of which only 52.9% are geo-referenced. These are thought to represent a large proportion of the specimen data existing in digital form. Effectively, the "low hanging fruit" have been picked and although the amount of specimen data increases steadily, the present growth rate is minimal when considering that estimates of the total numbers run in the range of 1.2 to 2.1 billion specimens (Ariño, 2010) or even higher (see reference in Vollmar & al., 2010).

In 2008, the GBIF Secretariat concluded that a strategy and action plan is needed to incite further mobilization of specimen data. GBIF constituted a Task Group of domain experts, with the aim of developing a draft action plan and strategy that is relevant at the global level and informs regional and national action plans for the digitization of natural history collections (GBIF, 2008b). A work plan of the GBIF Task Group for the Global Strategy and Action Plan for Mobilization of Natural History Collections Data was formally proposed to the GBIF Governing Board in November 2008 (GBIF, 2008c).

The Task Group conducted a global survey to identify the barriers and challenges to digitization (Vollmar et al., 2010) and undertook a prototype study to estimate the universe of the natural history collections data (Ariño, 2010). It further held consultations with the natural history collections community and professional societies such as the Society for Preservation of Natural History Collections (Berendsohn et al., 2009), and the Consortium of European Taxonomic Facilities (CETAF). Possible technical solutions were discussed in the biodiversity informatics communities, e.g. in TDWG (Bourgoin et al., 2009). An interim report of the Task Group was presented and circulated at the 16th meeting of the GBIF Governing Board held in October 2009 in Copenhagen. The Task Group submitted its report to the GBIF Science Committee in April 2010. The articles in this volume represent key parts of this report. In the following a summary of the conclusions the Task Group arrived at is given, they are based on the discussions during meetings and workshops and more details are provided in the individual articles cited.

GENERAL CONCLUSIONS

The development of a Global Strategy and Action Plan for the Mobilization of Natural History Collections Data must be guided by a single basic strategic principle: user demand must be the driver of the detailed digitization of individual specimens. Accordingly, priorities for digitization should be set either according to the demand from on-going or projected research, or in accordance with socio-political demands (Conventions etc.). Funding of digitization activities should be linked directly to these priorities, with the costs either incorporated into research proposals, or covered by (international) organizations, foundations, or governments.

We posit that a demand-driven approach makes the size of the digitization task realistic and fundable, because the task can be focused and outside funding can be mobilized. However, a principal obstacle is the lack of access to potentially useful un-digitized collection holdings. The following three recommendations refer to helping users to address their demand for information, helping collections to answer that demand, and to the provision of the infrastructure necessary to transport the results to the user.

RECOMMENDATIONS

1: GBIF must facilitate access to information about non-digitized collection resources

1.1. Publicize the research potential of collections through metadata

- Metadata (i.e. higher-level information about the content of a collection) are an essential infrastructural component enabling demand-driven specimen-level digitisation (Berents et al., 2010, Scoble & Bourgoin, 2010, Berendsohn & Seltmann, 2010). GBIF Participants and national funding agencies must provide funding for metadata capture, publication and discovery in order to build this essential research infrastructure.
- Collections must investigate the institution-specific costs of metadata capture.
- Collections must expose their metadata holdings to promote the potential data availability to users and to expand the user base.
- GBIF must assist collections to use appropriate means to publish standard metadata to describe their holdings.
    - GBIF and organizations of collection directors and custodians should work together to establish discipline-specific best practice guidelines for metadata capture.
    - GBIF must set up or identify data capture mechanisms allowing collection managers or metadata authors to report numbers of specimens that belong to certain metadata categories.
    - GBIF must foster development of controlled vocabularies useful for authoring enriched metadata documents.
    - The metadata describing a collection represent publishable content. GBIF should promote scholarly credit for metadata authoring, in order to

provide an incentive for quality metadata provision.

- GBIF and the collections should initially focus on metadata records documenting – for the entire collection – the number of un-digitized specimens of a particular (higher) taxon from a specific country or region of origin. Large scale research questions as well as data repatriation efforts are centered upon these data areas. Currently there is no straightforward way for interested parties to assess the amount of data that should by digitized for their purposes (Berendsohn & Seltmann, 2010).
- GBIF should create a mechanism to query the metadata (e.g. Who has how many specimens of Amphibia from Tanzania) and to send requests initiating communications about demand for specimen-level digitization between a user and a collection manager (Berendsohn & Seltmann, 2010).
- The GBIF Secretariat should set up a pilot metadata system (Bourgoin et al., 2009), using the taxonomic groups suggested at the Leiden workshop (Berendsohn et al., 2010) as "do-able targets".

1.2. Assessing the number of un-digitized specimens

- In parallel to the metadata effort, GBIF should use an analytical or statistical approach to determine the size of the holdings of un-digitized specimens world-wide (Ariño, 2010).

2: GBIF must work with collections to continue to increase the efficiency of specimen data capture and to enhance data quality.

2.1. Technical measures

- GBIF must investigate and document ways to industrialize specimen digitization.
- GBIF must foster the development and dissemination of collection type-specific procedures, best practice guidelines and tools aimed at streamlining workflows for data entry, imaging, text and feature recognition, georeferencing, and general quality enhancements.
- GBIF should develop and promote simple, easy-to-use, intuitive and efficient data capture tools that would be accessible to all user skill levels.
- GBIF should foster the creation of 'data dictionaries' in order to disperse the training load and reduce the number of input errors.
- GBIF and relevant societies/organizations must investigate innovative ways to use citizen science approaches for specimen-level digitization.

2.2. Attribution, professional credit, institutional priorities

- GBIF must continue to foster digitization related capacity building and training activities.
- Natural History Collection Management should develop best practice and policy for data mobilization and curation accompanying their management of the physical collection.
- GBIF and other relevant stakeholder organizations must promote the recognition of natural history collection data publishing as a scholarly and scientifically useful exercise.
- Collection administrators and/or management must make efforts to allocate dedicated human resource for digitization activities.
- GBIF must work together with collection managers to help recognize digitization as a component of their formal job description.
- Digitization of type specimens (all type material) should be made a prerequisite for the publication of new names (Berents et al., 2010). Digitization and public data availability for all specimens used in a taxonomic revision should be recognized as best practice and made mandatory by collections.

2.3. Funding

- GBIF must actively engage with funding agencies in convincing them of the significance of support for natural history collections digitization.
- GBIF must continue its seed money award scheme as it has acted as a catalytic agent and

created ripple effects in encouraging collections and many national and international donor agencies to support digitization.

- Science funding agencies and private donors must fund digitization where the research demand exists.
- Collections must investigate the institution-specific costs of specimen data capture.

3: GBIF must continue to improve and promote the global infrastructure used to mobilize digitized collection data

3.1. Technical measures

- GBIF must encourage hosting environments to facilitate discovery and publishing of natural history collections data especially for small and mid-sized museums.
- GBIF must accelerate progress toward allocation and resolution of persistent identifiers at the dataset and data record level, including services and best practice guidelines.

3.2. Outreach activities

- GBIF should make special efforts to include collections from the Southern hemisphere.
- GBIF must pay special attention to small and medium sized collections based in research and academic institutions and their organizations, as they may be able to contribute with minimal investment and encouragement.

3.3. Political measures

- National and/or thematic digitization strategies must be developed to identify and draw from possible synergies with digitization efforts in other domains (e.g. primary data, archives, print publications).

OUTLOOK

The recommendations presented are intended to form the base for further discussion. Some of the questions raised will only be answered when action is taken. It is clear, however, that the natural history collection community has to work closer together to address the question of how to mobilize their information resources in an efficient and usable way. Further implementation of these recommendations calls for socio-political decisions, cultural changes, as well increased infrastructural, technological and financial investment by the stakeholder communities at all levels from local to global scale.

REFERENCES

Ariño, A. H. 2010. Approaches to estimating the universe of natural history collection data. J. Biodiversity Informatics 7: 81-92.

Baird, R. 2010. Leveraging the fullest potential of scientific collections through digitization. Biodiversity Informatics 7: 130-136.

Berendsohn, W. G., Berents, P. and Macklin, J.A. 2009. Results of the Leiden Workshop on digitisation priorities. Report from a workshop during the 2009 annual meeting of the Society for the Preservation of Natural History Collections (SPNHC)[1]. [Accessed September 1, 2010]

Berendsohn, W. G. and Seltmann, P. 2010. Using geographical and taxonomic metadata to set priorities in specimen digitization. J. Biodiversity Informatics 7: 120-129.

Berents, P., Hamer, M. and Chavan, V. 2010. Towards demand-driven publishing: approaches to the prioritization of digitization of natural history collections data. J. Biodiversity Informatics, 7: 113-119.

Bourgoin, T., Berendsohn, W. G. & Macklin, J. A. 2009. The Natural History Collections Geotaxonomic Index. Report from the working session during the 2009 meeting of the organisation for Biodiversity Information Standards (TDWG) in Montpelier[2]. [Accessed September 1, 2010]

GBIF 2008a: GBIF Work Programme 2009-2010[3]. [Accessed September 9, 2010]

GBIF 2008b: Terms of Reference for "Task Group on a Global Strategy and Action Plan for the Mobilisation of Natural History Data"[4]. [Accessed August 29, 2010]

GBIF 2008c: Global Strategy and Action Plan for the Digitization of Natural History Collection Data.

1 http://www2.gbif.org/GSAP-NHC_SPNHC09_Report.pdf
2 http://www2.gbif.org/GSAP-NHC_TDWG09_WorkingSession.pdf
3 http://www2.gbif.org/WP2009-10.pdf
4 http://tinyurl.com/gsaptg

Preliminary report to the GBIF Governing Board 15 in Arusha[5]. [Accessed September 12, 2010]

GBIF in prep.: Discovery and Publishing of Primary Biodiversity Data through GBIF network: The State-of-the-Art and Potentials (version 2.0), 52pp. Copenhagen.

OECD 1999. Final Report OECD Megascience Forum Working Group on Biological Informatics[6]. OECD, Paris, Jan. 1999. [Accessed August 29, 2010]

Scoble, M. J. and Bourgoin, T. 2010. Natural history collections digitization: rationale and value. J. Biodiversity Informatics 7: 77-80.

Vollmar, A., Macklin, J. A., and Ford, L.S. 2010. Natural history specimen digitization: challenges and concerns, J. Biodiversity Informatics 7: 93-112.

5 http://www2.gbif.org/GSAP-NHC-GB15ReportwithAnnex.pdf

6 http://www.oecd.org/dataoecd/24/32/2105199.pdf