

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

NICHOLAS KING^{1†}, LEONARD KRISHTALKA^{2†}, AND VISHWAS CHAVAN^{1†}

*1*Global Biodiversity Information Facility Secretariat, Universitetsparken 15, DK-2100, Copenhagen, Denmark. Email: nking@gbif.org and vchavan@gbif.org.

*2*Biodiversity Institute (and) Department of Ecology and Evolutionary Biology, The University of Kansas, Dyche Hall, Lawrence, KS 66045, USA. Email: krishtalka@ku.edu

† Dr. Nicholas King is the Executive Secretary of the Global Biodiversity Information Facility (GBIF). Dr. Leonard Krishtalka is the Chair of the GBIF Science Committee. Dr. Vishwas Chavan is the Senior Programme Officer for DIGIT (Digitization and Mobilization of Primary Biodiversity Data including natural history collections data). This article is a reflection by the authors on behalf of GBIF (Secretariat and the Science Committee) on potential uptake of the recommendations from the GBIF Global Strategy and Action Plan for Mobilisation of Natural History Collections Data (GSAP-NHC) Task Group.

Abstract: The Global Biodiversity Information Facility (GBIF) has a mandate to facilitate free and open access to primary biodiversity data worldwide. This special issue of *Biodiversity Informatics* publishes a collection of papers that summarize the findings of the recent GBIF Task Group on a Global Strategy and Action Plan for Mobilization of Natural History Collections Data (GSAP-NHC). The GSAP-NHC Task Group has made three primary recommendations dealing with discovery, capture, and publishing of natural history collections data. This overview provides insight on various activities initiated by GBIF to assist with an early uptake and implementation of these recommendations. It calls for proactive participation by all relevant players and stakeholder communities. Given recent technological progress and growing recognition and attention to biodiversity science worldwide, we believe that rapid progress in discovery, publishing and access to large volumes of useful collection data can be achieved for the immediate benefit of science and society.

Keywords: Global Biodiversity Information Facility (GBIF), Natural History Collections, Discovery, Digitization, Publishing, Primary Biodiversity Data.

The Global Biodiversity Information Facility (GBIF) has a mandate to facilitate free and open access to primary biodiversity data worldwide (OECD, 1999). Digitization and mobilization of natural history collections data has been an integral part of the Work Programme of the Global Biodiversity Information Facility (GBIF) since its inception in 2001 (OECD, 1999 and GBIF, 2002). To date, GBIF facilitates access to more than 200 million primary biodiversity data records of which 25.7% are specimen-based (GBIF, 2010a). These data have widespread application in various areas such as biodiversity research, natural resources

management, and bio-security (Chapman, 2005; Tann, et. al., 2008; Pyke & Ehrlich, 2010, Baird, 2010, this volume; Berents, et al., 2010, this volume). To expedite the progress of discovery and publishing of natural history collections data through the GBIF network, a Task Group on a Global Strategy and Action Plan for Mobilization of Natural History Collections data (GSAP-NHC) was established in 2008 (GBIF 2008a). This Task Group formulated three primary recommendations and called for local-to-global scale involvement of all players and increased investment by stakeholders.

The Task Group's three primary recommendations for the GBIF community addressed (i) facilitating access to information pertaining to non-digitized collections through enriched metadata; (ii) establishing a conducive data publishing framework; and (iii) improvement and promotion of a global infrastructure to increase efficiency of discovery and publishing of the natural history collections data (Berendsohn, et.al., 2010, this volume). In this article, we outline measures being implemented by the GBIF network that will ensure an early uptake of these recommendations. Further, we reflect on potential approaches to implement these recommendations to ensure expeditious progress in discovery, digitization, and publishing of natural history collections data.

RECOMMENDATION 1: GBIF MUST FACILITATE ACCESS TO INFORMATION ON NON-DIGITIZED NATURAL HISTORY COLLECTION RESOURCES.

A basic GBIF metadata profile¹ has been developed and is being implemented through development of the GBIF Integrated Publishing Toolkit (IPT). The metadata profile incorporates the necessary elements to describe non-digitized collections including the number of non-digital specimens, and approximate cost per specimen digitization. The GBIF Metadata Implementation Task Group (GBIF 2009a) has suggested the use of controlled terminologies for use in metadata authoring. In order to encourage the development of the distributed metadata catalogue infrastructure, in 2009 GBIF awarded nine small grants to GBIF Participants (GBIF 2009b). Furthermore, a prototype of the controlled vocabulary server² that will support public drafting of key extensions and multi-lingual vocabularies is being implemented. In order to encourage scholarly credit for metadata publishing, the concept of 'Data Papers' - publishing of enriched metadata documents in scholarly journals - is being tested through a GBIF pilot study. Three journals namely, ZooKeys, PhytoKeys and BioRisks are participating in this pilot study.

1 <http://rs.gbif.org/schema/eml/eml-gbif-profile.xsd>

2 <http://vocabularies.gbif.org>

RECOMMENDATION 2: GBIF MUST WORK WITH THE NATURAL HISTORY COLLECTIONS COMMUNITY TO CONTINUE TO INCREASE THE EFFICIENCY OF SPECIMEN DATA CAPTURE AND TO ENHANCE DATA QUALITY.

As part of wider promotion of relevant biodiversity informatics best practices, in 2008 the GBIF Training Manual on Digitization of Natural History collections data was published (GBIF, 2008b). The methodologies therein are being used widely by the community. While GBIF promotes several community developed digitization tools, in 2011 we plan to develop a catalogue of tools that could be used by potential data publishers at varied skill-set levels. GBIF also plans to commission studies on ways to industrialize discovery, digitization, and publishing of natural history collections data, and engage citizens and naturalists into digitization activities. Whilst GBIF has in the past awarded small grant 'seed funding' to encourage individual digitization activities (GBIF, 2003; GBIF, 2004; GBIF, 2006; and GBIF, 2008c), in future we aim to encourage demand-driven discovery and publishing initiatives through strategic partnerships with key professional societies at global and regional levels. The GBIF-commissioned Data Publishing Framework Task Group has recommended several activities such as 'Data Papers', 'a Data Usage Index', and 'Data Citation Mechanisms' to recognize data publishing as a scholarly activity (Chavan and Ingwersen, 2009). The majority of these mechanisms will be implemented in a phased manner over the next few years. In addition, the GBIF Science Committee has recommended the formation of a high-level Task Group to engage with Directors of major natural history collections institutions and key national, regional, and international funding agencies to seek increasing investment and support towards data discovery, digitization, and publishing activities. The recent announcement by the U.S. National Science Foundation of a program to fund the Advanced Digitization of Biological Collections (ADBC) is a major step in this direction³. The high-level GBIF Task Group will also engage with professional societies to advocate digitization of collections as an essential

3 http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503559

component of curatorial best practices and advancing research in their biodiversity science disciplines.

RECOMMENDATION 3: GBIF MUST CONTINUE TO IMPROVE AND PROMOTE THE GLOBAL INFRASTRUCTURE USED TO MOBILIZE DIGITIZED COLLECTIONS DATA.

Whilst GBIF, through the above-mentioned high-level GBIF Task Group, intends to engage with large institutions in the northern hemisphere, it is equally important to engage with collection facilities in the ‘South’, as well as a small- and medium-size museums worldwide. To facilitate the development of demand-driven strategies and action plans for discovery, digitization and publishing of primary biodiversity data, including those associated with natural history collections, a best practice guide will soon be released for community uptake (GBIF, 2010b). These guides are also intended to assist collections institutions to secure increased funding for data publishing activities due to their adherence to the GBIF global standards and identified global priorities. Furthermore, GBIF will roll out plans for establishing a distributed data hosting infrastructure (GBIF, 2008d) within the GBIF network to facilitate publishing of data by small and medium-size data holders. A Beginners Guide to Persistent Identifiers is under development to encourage and assist the practice of allocation and resolution of persistent identifiers at the dataset and data record level (GBIF, 2010c). A Joint GBIF-TDWG Task Group on Multimedia Resources has developed a Multimedia Resources Metadata Schema (MRTG), which will be implemented as a DarwinCore extension in the next version of the IPT, facilitating discovery of multimedia objects (images, video, and audio) related to specimens (GBIF, 2009c; Morris, et.al., 2008, 2009). This schema is currently undergoing a TDWG ratification process. A GBIF Position Paper on enhancing fitness-for-use of primary biodiversity data (GBIF, 2010d) makes specific suggestions towards improving the geospatial fitness-for-use of data. These recommendations will be implemented in a phased manner in the near future.

THE WAY FORWARD – COMMUNITY UPTAKE

These measures indicate that GBIF is either implementing or has plans to initiate a suite of activities that will help expedite progress in discovering and publishing scientifically and socially relevant natural history collections data. However, there are several factors beyond GBIF’s control that determine the degree of success in ensuring free and open access of the world’s natural history collections data. The most important of these is the commitment of the collections community itself to accept and implement these recommendations. We believe that implementation of these recommendations at the national scale will lead to construction of ‘national information infrastructure’ (Chavan and Krishnan, 2003) facilitating expedited discovery, digitization and publishing of natural history collections data.

These recommendations call for a cultural shift in the collections community—in curatorial practices, in developing demand-driven strategy and action plans, in recognition of digitization as a mainstream, if not priority, activity for collections institutions, as well as scholarly recognition for discovery and publishing efforts by both individuals and institutions (Krishtalka and Humphrey, 2000). We suggest that collections-related professional societies such as the Society for Preservation of Natural History Collections (SPNHC), the Natural Science Collections Alliance (NSCA), and the Consortium of European Taxonomic Facilities (CETAF) can play a vital advocacy role in promoting the discovery and publishing of collections data as a core institutional activity. We also strongly encourage academic and scientific publishers to mandate that systematic and taxonomic publications be accompanied by the simultaneous discovery, capture, and publication of associated specimen data—as is now the case for publications based on genomic data (Ventura, 2005, Lindberg & Humphreys, 2008). Recent technological advances make it feasible to have such simultaneous discovery and publishing of collections data together with scholarly publications (Penev, et.al., 2009 and Penev, et.al., 2010). National, regional, and international funding agencies can play a catalytic role by supporting digitization activities on the basis of demand-driven strategies, so that

such data when accessible are scientifically, socially, ecologically, and economically relevant and useful. We further recommend that funding agencies stipulate that the discovery and publishing of data from legacy digitization projects or future collection activities are a prerequisite for financial support to such activities.

Given ongoing advances in information technologies, and growing recognition and attention to biodiversity science worldwide, we believe that a rapid progress in discovery, publishing, and access to large volumes of useful collections data can be achieved for the immediate benefit of science and society. The degree of success is largely dependent on how rapidly the collections community can collaborate and adapt to implementing these recommendations. We call on the collections community to join GBIF in expediting uptake of the techniques and technologies already available, and to participate in their further development and enhancement.

REFERENCES:

- Baird, R. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics*, 7: 130-136.
- Berendsohn, W.G.; Chavan, V.S.; and J. A. Macklin 2010. Recommendations of the GBIF Task Group on Global Strategy and Action Plan for the mobilization of natural history collections data. *Biodiversity Informatics*, 7: 67-71.
- Berents, P., Hamer, M., and V. Chavan. 2010. Towards demand-driven publishing: approaches to the prioritization of digitization of natural history collections data. *Biodiversity Informatics*, 7: 113-119.
- Chapman, A. D. 2005. Uses of primary species-occurrence data, version 1.0. Copenhagen: Global Biodiversity Information Facility. 106 pp. ISBN: 87-92020-01-1.
- Chavan, V. and S. Krishnan (2003) Natural history collections: a call for national information infrastructure. *Current Science*, 84(1): 34-42.
- Chavan, V.S., and P. Ingwersen, 2009. Towards a data publishing framework for primary biodiversity data: challenges and potentials for the biodiversity informatics community. *BMC Bioinformatics*, 10 (Suppl 14): S1, doi: 10.1186/1471-2105-10-S14-S2.
- GBIF 2002. GBIF Annual Report 2001-2002. Accessible online at http://www2.gbif.org/annual_report_2001_2002.pdf (Accessed on September 17, 2010).
- GBIF 2003. DIGIT seed money award, 2003. Accessible online at <http://www.gbif.org/informatics/primary-data/seed-money-awards/2003/#c1007>. (Accessed on September 17, 2010).
- GBIF 2004. DIGIT seed money award, 2004. Accessible online at <http://www.gbif.org/informatics/primary-data/seed-money-awards/2004/#c1008>. (Accessed on September 17, 2010).
- GBIF 2006. DIGIT seed money award, 2005-2006. Accessible online at <http://www.gbif.org/informatics/primary-data/seed-money-awards/2005-2006/#c1011>. (Accessed on 17 September 2010).
- GBIF 2008a. Terms of Reference for “Task Group on a Global Strategy and Action Plan for the Mobilisation of Natural History Data”. Accessible at <http://tinyurl.com/gsapptg> (Accessed on September 17, 2010).
- GBIF 2008b. GBIF Training Manual 1: Digitisation of Natural History Collections Data, version 1.0. Copenhagen: Global Biodiversity Information Facility. Accessible at <http://www.gbif.org> (Accessed on September 17, 2010).
- GBIF 2008c. GBIF seed money award, 2007-2008. Accessible online at <http://www.gbif.org/informatics/primary-data/seed-money-awards/2007-2008/#c1009>. (Accessed on September 17, 2010).
- GBIF 2008d. GBIF Work Programme 2009-2010. 2008. Copenhagen. Global Biodiversity Information Facility. pp. 57. Accessed online at <http://www2.gbif.org/WP2009-10.pdf> (Accessed on September 17, 2010).
- GBIF 2009a. Report of the GBIF Metadata Implementation Framework Task Group. 2009. Copenhagen. Global Biodiversity Information Facility. pp. 38. Accessible online at <http://www2.gbif.org/GBIF-MIFTG-Report.pdf> (Accessed on September 17, 2010).
- GBIF 2009b. GBIF small grant awards for metadata catalogues. 2009. Copenhagen. Global Biodiversity Information Facility. Accessible online at <http://tinyurl.com/metadatasmallgrant> (Accessed on September 17, 2010).

- GBIF, 2009c. Multimedia Resources Task Group. Copenhagen: Global Biodiversity Information Facility. Accessible at <http://www.gbif.org/informatics/primary-data/task-groups/mrtg/>. (Accessed on September 17, 2010).
- GBIF 2010a. Discovery and Publishing of Primary Biodiversity Data through GBIF Network: The State-of-the-Art and Potentials. 2010: Copenhagen. Global Biodiversity Information Facility. (in press).
- GBIF 2010b. Best Practice Guide on Demand-driven Data Discovery and Publishing Strategy and Action Plans (in preparation). 2010: Copenhagen. Global Biodiversity Information Facility. (in press).
- GBIF 2010c. GBIF Beginner's guide to persistent identifiers: call for comments on draft document. Copenhagen. Global Biodiversity Information Facility. Accessible online at <http://www.gbif.org/communications/news-and-events/showsingle/article/gbif-beginners-guide-to-persistent-identifiers-call-for-comments-on-draft-document/>. (Accessed September 20, 2010).
- GBIF 2010d. GBIF Position Paper on future directions and recommendations for enhancing fitness-for-use across the GBIF Network, version 1.0. Authored by Hill A.W., Otegui, J., Ariño, A. H., and R. P. Guralnick. 2010. Copenhagen: Global Biodiversity Information Facility, 25 pp. ISBN: 87-92020-11-9. Accessible online at <http://www.gbif.org> (Accessed on September 17, 2010).
- Krishtalka, L. and P. S. Humphrey. 2000. Can natural history museums capture the future? *BioScience* 50 No. 7: 611-617.
- Lindberg, D.A.B. and B. L. Humphreys. 2008. Rising expectations: Access to biomedical information. *Yerb. Med. Inform.* 391): 165-172.
- Morris, R., Olson, A., O' Tauma, E., Riccardi, G., Whitbread, G., Hagedorn, G., Teage, I., Heikkinen, M., Leary, P., Barve, V., and V. Chavan. 2008. Recommendations of the GBIF Multimedia Resources Task Group. Copenhagen: Global Biodiversity Information Facility., September 2008. Pp. 18. Accessible at <http://www2.gbif.org/MRTG-Recommendations-29-09-2008.pdf>. (Accessed on September 17, 2010).
- Morris, R., Olson, A., Freeland, C., Hagedorn, G., Carausu, M., O'Tauma, and V. Chavan. 2009. Mobilising multimedia resources in biodiversity: 2nd report of the GBIF Multimedia Resources Task Group. Copenhagen: Global Biodiversity Information Facility., March 2009. Pp. 22. Accessible at <http://www2.gbif.org/MRTG2ndReportFinal6Apr09.pdf>. (Accessed on September 17, 2010).
- OECD 1999. Final report of the OECD megascience forum: Working group on biological informatics, January 2009. Organisation for Economic Co-operation and Development . pp. 74. Accessible at http://www.gbif.org/fileadmin/Temp_for_New_Web_Site/BIrepfin.pdf (Accessed on September 17, 2010).
- Penev, L., Erwin, T., Miller, J., Chavan, V., Moritz, T., and C. Griswold. 2009. Publication and dissemination of datasets in taxonomy: ZooKeys working example. *ZooKeys*, 11: 1-8, doi: 10:3897/zookeys.11.210.
- Penev, L., Agosti, D., Georgiev, T., Catapano, T., Miller, J., Blagoderov, V., Roberts, D., Smith, V.S., Simon Rycroft, B., Scott, B., Johnson, N. F., Morris, R. A., Sautter, G., Chavan, V., Robertson, T., Remsen, D., Stoev, P., Parr, C., Knapp, S., Kress, W. J., Thompson, C. F., and E. Terry. 2010. Semantic tagging of and semantic enhancements to systematics papers: ZooKeys working examples. *ZooKeys*, 50: 1-16. doi: 10:3897/zookeys.50.538.
- Pyke, G. H. and P. R. Ehrlich. 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biol. Rev.*, 85: 247-266.
- Tann, J., Kelly, L., and Flemons, P. 2008. Atlas of Living Australia – User needs analysis. Accessible online at: <http://www.ala.org.au/documents/user-needs-analysis-report.html>. (Accessed on September 20, 2010).
- Ventura, B. 2005. Mandatory submission of microarray data to public repositories: how is it working? *Physiol. Genomics*. 20: 153-156, doi:10.1152/physiolgenomics.00264.2004.