






Please cite the Published Version

Asase, A , Sainge, MN , Radji, RA , Ugbogu, OA  and Peterson, AT  (2020) A new model for efficient, need-driven progress in generating primary biodiversity information resources. *Applications in Plant Sciences*, 8 (1). e11318 ISSN 2168-0450

DOI: <https://doi.org/10.1002/aps3.11318>

Publisher: Wiley

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/636101/>

Usage rights:  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

Additional Information: This is an open access article published in *Applications in Plant Sciences*, by Wiley.

Data Access Statement: The data resources generated by this project are available as online resources via the Global Biodiversity Information Facility (<https://doi.org/10.15468/e8rhqm> [Asase, 2018]; <https://doi.org/10.15468/9czcig> [Ugbogu et al., 2019]). Other data sets, yet to be published, are available from the authors upon reasonable request.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

A new model for efficient, need-driven progress in generating primary biodiversity information resources

Alex Asase^{1,6} , Moses N. Sainge² , Raoufou A. Radji³ , Omokafe A. Ugbogu⁴ , and A. Townsend Peterson^{5,6} 

Manuscript received 30 April 2019; revision accepted 28 October 2019.

¹Department of Plant and Environmental Biology, University of Ghana, P.O. Box LG 55, Legon, Ghana

²Tropical Plant Exploration Group (TroPEG), P.O. Box 18, Mundemba, Cameroon

³Department of Botany, University of Lome, 06 BP 6135 Lome, Togo

⁴Forestry Research Institute of Nigeria, PMB 5054, Jericho Hills, Ibadan, Nigeria

⁵Biodiversity Institute, and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

Authors for correspondence: alexasase@ug.edu.gh; town@ku.edu

Citation: Asase, A., M. N. Sainge, R. A. Radji, O. A. Ugbogu, and A. T. Peterson. 2020. A new model for efficient, need-driven progress in generating primary biodiversity information resources. *Applications in Plant Sciences* 8(1): e11318.

doi:10.1002/aps3.11318

PREMISE: The field of biodiversity informatics has developed rapidly in recent years, with broad availability of large-scale information resources. However, online biodiversity information is biased spatially as a result of slow and uneven capture and digitization of existing data resources. The West African Plants Initiative approach to data capture is a prototype of a novel solution to the problems of the traditional model, in which the institutional “owner” of the specimens is responsible for digital capture of associated data.

METHODS: We developed customized workflows for data capture in formats directly and permanently useful to the “owner” herbarium, and digitized significant numbers of new biodiversity records, adding to the information available for the plants of the region.

RESULTS: In all, 190,953 records of species in 1965 genera and 331 families were captured by mid-2018. These data records covered 16 West African countries, with most of the records (10,000–99,999) from Côte d’Ivoire, Ghana, Togo, Nigeria, and Cameroon, and the fewest data records from Mauritania (<100 records). The West African Plants Initiative has increased available digital accessible knowledge records for West African plants by about 54%. Several of the project institutions have put initial project data online as part of their Global Biodiversity Information Facility data contributions. The average cost of data capture ranged from US\$0.50–1.00 per herbarium sheet.

DISCUSSION: Data capture has been cost-effective because it is much less expensive than de novo field collections, allows for development of information resources even for regions in which political situations make contemporary field sampling impossible, and provides a historical baseline against which to compare newer data as they become available. This new paradigm in specimen digitization has considerable promise to accelerate and improve the process of generating high-quality biodiversity information, and can be replicated and applied in many biodiversity-rich, information-poor regions to remedy the present massive gaps in information availability.

KEY WORDS biodiversity informatics; digitization; plants; primary biodiversity data; West Africa.

Biological diversity is distributed unevenly worldwide, with well-known biogeographic patterns such as the latitudinal diversity gradient (Mannion et al., 2014), in which the tropics hold many more species than higher-latitude regions. Other examples of such uneven distribution of biological diversity include the existence of centers of endemism and so-called “hotspots” of biodiversity (Myers et al., 2000). These imbalances are, for whatever reasons, often correlated (if not causally) with differences in wealth, in which the wealthy, so-called Global North is relatively biodiversity-poor, whereas the so-called Global South is often biodiversity-rich and economically poor (Balmford et al., 2005; Rands et al., 2010). A third dimension of

this equation centers on the geographic distribution of biodiversity data worldwide. As with global wealth, this commodity tends to be concentrated in the Global North and to be relatively sparse and rare in tropical countries. The reasons behind this unequal distribution of information have to do with histories of occupation and colonization (Peterson et al., 2016a), as well as with histories of science and its development largely in the Global North, at least until relatively recently (Collen et al., 2008).

The field of biodiversity informatics has developed rapidly in recent years, with the appearance of important and exciting new tools for analyzing and interpreting biodiversity data. These tools, when

linked properly to information availability, have in turn yielded exciting advances in understanding key aspects and dimensions of biodiversity: its distribution in space and time and with respect to environments, and optimal strategies for its conservation (Peterson et al., 2010). Another area of massive recent advance in biodiversity informatics has been in the broad availability of large-scale information resources such as the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org>), VertNet (<http://vertnet.org>), speciesLink (<http://splink.cria.org.br>), REMIB (<http://remib.org>), iDigBio (<http://www.idigbio.org>), and others.

However, online biodiversity information is notoriously and dramatically biased spatially, as has been demonstrated in numerous recent studies (Yesson et al., 2007; Boakes et al., 2010; Stropp et al., 2016; Sosef et al., 2017), which aligns with the information inequities enumerated above. Underdevelopment of biodiversity information resources emerges as a result of slow capture and digitization of existing data resources on the part of the individuals and institutions holding those data and/or specimens; this is a common problem, as the systematics enterprise has become budget stressed and resource starved (Bini et al., 2006). Unfortunately, these processes are long term in nature, and there is no evidence that they are in the process of being resolved.

This paper offers a first view of a new model designed specifically to remedy these imbalances. We noted striking contrasts between the enthusiasm for an entry into the world of biodiversity informatics among scientists from across West Africa, and the information resources that were so painfully slow to appear to support and nourish that enthusiasm. Rather than languish or work with too-small data resources while awaiting digitization efforts on the part of the institutions that hold the key specimens, we designed an alternative approach, building on previous explorations by Mexican institutions (Soberón and Llorente-Bousquets, 1993; Sarukhán et al., 2014). Specifically, the model explored herein is one in which scientists in the region can take charge of their own information destinies, and initiate and implement steps to create new information resources.

THE WEST AFRICAN PLANTS INITIATIVE

The West African Plants (WAP) Initiative is a consortium of West African researchers in botany; its goal is to digitize and mobilize available, high-quality, primary biodiversity occurrence data resources for West African plant diversity. The project covers a region extending from Senegal and Gambia east to Cameroon and north to Mali, and aims to create a rich data resource that can be made openly available and brought into currency for scientific research, policy decisions, and ultimately for sustainable development. A further goal of the initiative is to develop relevant human resources and skills regarding the plants of West Africa. The consortium includes partners from four West African institutions (University of Ghana, University of Lomé [Togo], Forestry Research Institute of Nigeria, and Tropical Plant Exploration Group [Cameroon]), in coordination with six institutions across Europe and North America (Naturalis Biodiversity Center [The Netherlands]; Royal Botanic Gardens, Kew and World Museum [United Kingdom]; Muséum National d'Histoire Naturelle [France]; and the Missouri Botanical Garden and New York Botanical Garden [United States]). All 10 institutions participating in the project have significant West African botanical

holdings in the form of herbarium sheets in numbers ranging from 10,000 to 160,000 or more specimens.

Within both West Africa and Africa more generally, major efforts to digitize and mobilize available primary biodiversity information resources have taken place. For example, the Southern African Botanical Diversity Network (SABONET) project involved computerization of data on labels of herbarium specimens of major herbaria in southern Africa (Siebert and Smith, 2004). The African Plants Initiative (API) project also digitized images of type specimens and associated information on African flora in herbaria in both the Northern and Southern Hemispheres, and made these accessible via electronic and other means for scholarly use (Smith, 2004). These previous efforts have been primarily taxonomic initiatives; WAP, on the other hand, is a new initiative with a different motivation and approach that is primarily focused on biodiversity informatics.

METHODS

The project linked caretakers of specimens at institutions on three continents with scientists from across West Africa who had an intense interest in accessing data on plants of their region. The West African proponents have the manpower and motivation, whereas European and North American participants increasingly share the desire for data associated with their collections to be digitized and shared globally. In the first phases of this project, we adopted previous workflows for digitization of biological collections (e.g., Nelson et al., 2012; Thiers et al., 2016) with specific modifications to capture data in formats that would be directly and permanently useful to the “owner” herbaria. Data capture started in 2015 and took place only at West African institutions. All data capture was strictly in accordance with DarwinCore standards (Wieczorek et al., 2012), achieved either by (a) capturing data records from existing images (e.g., images supplied by the Naturalis Biodiversity Center at this pilot phase) or (b) capturing data from images taken quickly and efficiently by project interns in West African herbaria. These images were captured using a Bencher BECM2PF Copymate II Fluorescent Tabletop Producer (110-240V; Bencher Inc., Antioch, Illinois, USA) via a system that controlled for image quality and reduced time spent in post-imaging batch processing and light adjustment. Data capture was achieved using BRAHMS database software (<http://herbaria.plants.ox.ac.uk/bol/>), customized for this project. Data capture was carried out by graduate students and technicians trained by this project who were involved at various stages in the project as interns. Selection of plant groups for digitization was guided by the specific interests of the West African partners.

This general model—that data be captured, improved, and mobilized by those who wish to have access to the data, rather than by the institutions holding the collections—is probably particularly needed in some regions of the world. As an initial step toward identifying such regions, we explored the 57,586,379 records of the kingdom Plantae that are based on preserved specimens (i.e., herbarium sheets) and that are served via the GBIF data portal (Global Biodiversity Information Facility, 2017a); this work is preliminary and is part of a broader exploration of the networks of providers and users that characterize the biodiversity informatics world at present (Peterson et al., in prep.). We processed these data from raw numbers in each country to densities of records per km² and then calculated the proportion of the total number of records in each country that are provided by institutions in

high bioinformatics potential, we instead look to guide countries as to which paths to explore in developing their own biodiversity information resources efficiently.

RESULTS

In all, 190,953 records of species in 1965 genera and 331 families were captured by mid-2018. Hierarchical taxonomic representation (i.e., order, family, genus) of the most common plant genera (i.e., with ≥ 500 records) among data records generated is summarized in Fig. 1. These primary biodiversity data records covered 16 West African countries, with most of the records (10,000–99,999) from Côte d'Ivoire, Ghana, Togo, Nigeria, and Cameroon, and the fewest data records from Mauritania (<100 records) (Fig. 2). These data records can be tied to subnational regions like states, departments, and provinces, reflecting broad geographic coverage by WAP data resources, but again with gaps and foci that limit the utility of the data (Fig. 3). The temporal distribution of the data records showed the earliest specimens in the first years of the 19th century and a peak in collections activity in the latter half of the 20th century (Fig. 4). Different countries saw increased or reduced collections activity over the centuries, probably as a result of a combination of political situations and collections priorities and interests (Fig. 4).

In broad terms, then, the WAP Initiative generated >190,000 primary biodiversity occurrence data records in a span of two years. Presently, 351,700 such plant herbarium sheet data records are available from the region in the form of digital accessible knowledge records via GBIF (query available for replication at <https://doi.org/10.15468/dl.yjr4tf> [Global Biodiversity Information Facility, 2017b]); therefore, the WAP Initiative has increased available digital accessible knowledge records for West African plants by about 54%. Perhaps most significantly, the average cost of data capture in this project ranged from US\$0.50–1.00/herbarium sheet for West African herbaria, but only US\$0.50/herbarium sheet for images supplied to the project by European institutions. The average cost of data capture per specimen includes costs associated with tasks across the entire digitization workflow (e.g., preparation tasks, staff allowance, post-digitization tasks). Finally, our efforts to identify regions with data to be captured, improved, and mobilized by those who wish to have access to plant biodiversity data emphasize much of Africa, the Middle East, and Central Asia, with South Asia and Southeast Asia not far behind (Fig. 5, bottom panel).

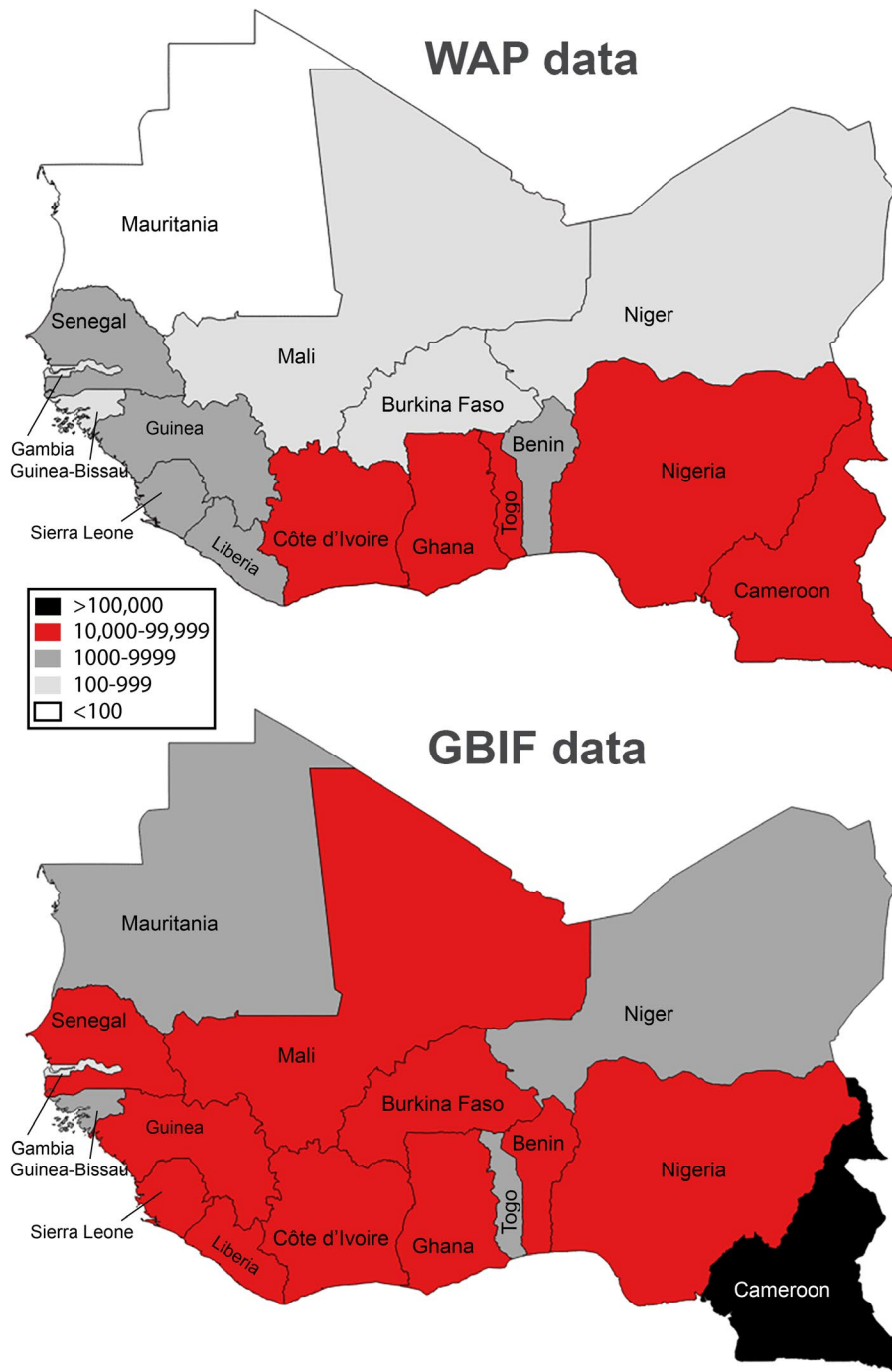


FIGURE 2. Summary of numbers of plant primary biodiversity data records captured as part of the West African Plants (WAP) Initiative, as compared to data otherwise available already via the Global Biodiversity Information Facility (GBIF).

DISCUSSION

Antecedents

This project is a direct intellectual descendant of the monumental efforts by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) in Mexico. CONABIO, initiated in 1992, faced a difficult challenge: to be a reliable source of authoritative

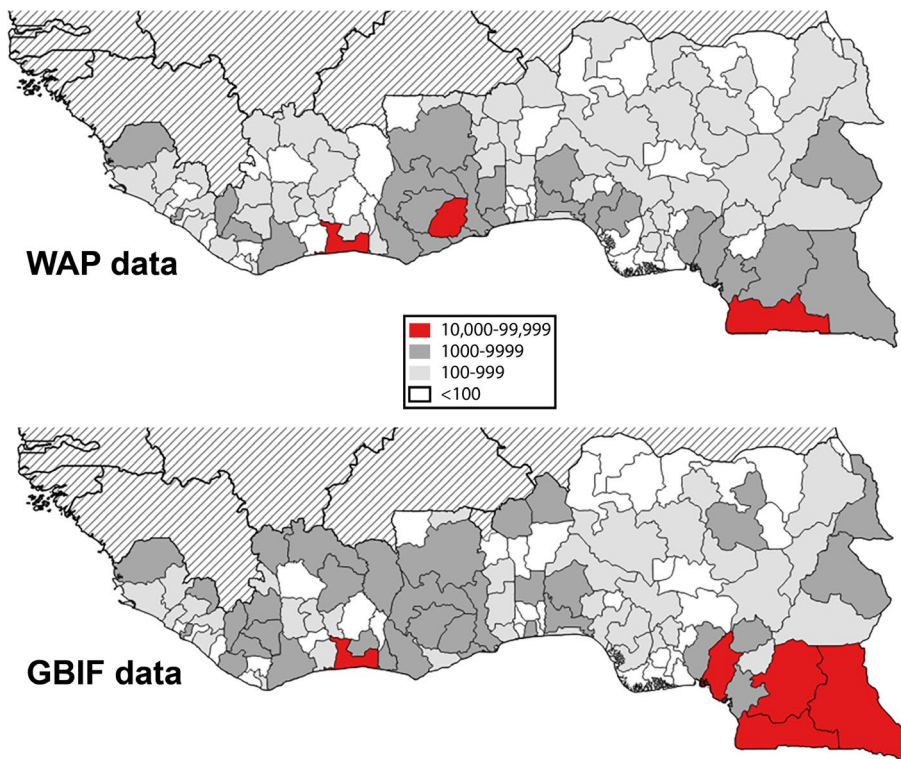


FIGURE 3. Distribution of primary biodiversity data records generated by the West African Plants (WAP) Initiative, as compared to data otherwise available already via the Global Biodiversity Information Facility (GBIF), at the level of states and provinces across eight West African countries. Surrounding countries all had state-level record richness <100, and so are not included in this summary.

biodiversity data for the Mexican government (as well as for its scientific community) when the data were not organized or digitized at the time of CONABIO's founding. This was particularly challenging because collections digitization was still in its infancy and no open biodiversity information networks (e.g., VertNet, GBIF) yet existed (Canhos et al., 2004). CONABIO met this challenge by two main strategies: by funding data-related biodiversity research and specimen curation and by sending out “data capture teams” to institutions around the world, particularly those that had no plans to initiate their own data capture. Data were captured by hand in most cases, sometimes with specimen-by-specimen capture, and in other cases based on images or photocopies of catalogs or specimen labels (Navarro-Sigüenza et al., 2002, 2003). Although many of the current tools did not yet exist (e.g., DarwinCore [Wieczorek et al., 2012], capture interfaces involving images [Nelson et al., 2012], and automated georeferencing [Guralnick et al., 2006]), this work proceeded specimen by specimen and resulted in numerous comprehensive, large-scale data resources for CONABIO and for the Mexican science and policy community (CONABIO, 2012).

Today, thanks to those multifaceted efforts by numerous taxon-based teams, CONABIO has placed Mexico among the global leaders in biodiversity informatics and biodiversity information availability, such that the country's knowledge of its biodiversity-related natural resources is equal to that of countries with greater financial resources. These information resources have permitted a meteoric increase in the quantity and quality of the biodiversity science developed in Mexico (Peterson et al., 2016a), as well as a key role for

CONABIO in guiding and informing natural resource policy across Mexico and the Mexican government (CONABIO, 2012). To a more limited extent, other regions and taxonomic specialist communities have experimented with need-based biodiversity information capture. For example, for Brazil's Reflora Project, the Paris and Kew herbaria provided large numbers of images of Brazilian herbarium specimens to the Brazilian team, which captured the associated data and provided high-quality data in return, thereby creating exciting new, large-scale information resources (Pignal et al., 2012; Sousa-Baena et al., 2013). These antecedent efforts proved the concept to which we have adhered in the WAP Initiative: biodiversity data can be captured and improved efficiently and effectively when those who lead the initiative are precisely those who dream of having such data resources available.

Was the West African Plants Initiative successful?

Biodiversity information for the West African region has heretofore been characterized by the existence of major gaps across geography (Koffi et al., 2015; Asase and Peterson, 2016; Gango and Kakpo, 2016) and, more broadly, across Africa (Dauby et al., 2016; Sosef et al., 2017). In this sense, the WAP Initiative has made major and significant progress, adding >190,000 new biodiversity records to the in-

formation available for the plants of the region. We note that capture of existing data, as prioritized in this project, makes considerable sense: (1) it is cost-effective because it is much less expensive than *de novo* field collections, (2) it allows development of information resources even for regions in which political situations make contemporary field sampling impossible, and (3) it provides a historical baseline against which to compare newer data as the latter become available (Peterson et al., 2016b).

A particularly important dimension in which the WAP Initiative has been successful is that of being cost-effective. A recent publication (Peterson and Soberón, 2017) presented data on costs of data resources developed by CONABIO over a large sample of projects and demonstrated a clear negative relationship between cost per record and overall magnitude of the project. It also showed that costs involved in creating CONABIO's data resources could be quite high, only rarely going below about US\$5.00 per data record. In this initiative, we had per-record costs of US\$0.50–1.00.

Further measures of success will center on whether the institutions “owning” the specimens follow through and put the new data records online when the newly captured data are sent to the institutions where the specimens are located. Already, several project institutions have put initial project data online as part of their GBIF data contributions, but success should be defined as *all* project-generated data being available online. Only when such data are fully online can they be applied in scientific analyses, in tandem with data provided by the same and other institutions under the aegis of other initiatives and/or institutional efforts.

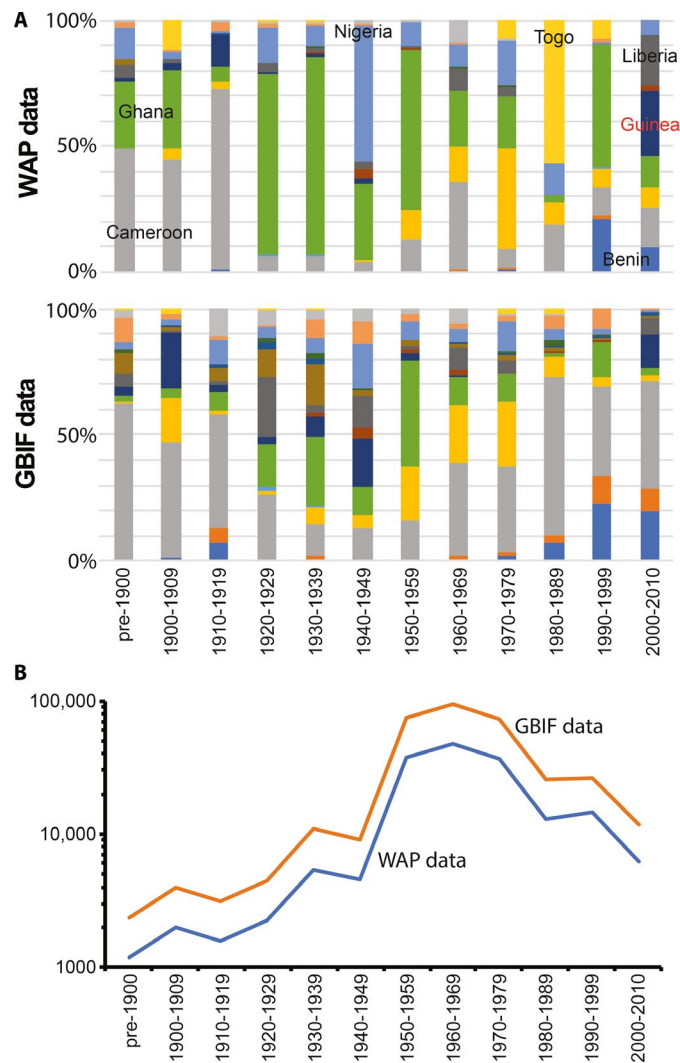


FIGURE 4. Temporal trends in West African Plants (WAP) Initiative data and Global Biodiversity Information Facility (GBIF) data. (A) Proportional representation of eight West African countries over the period 1900–2009. Country labels indicate which color pertains to which country. (B) Number of records through time, showing year-to-year variation. Note that the number of records is shown on a log₁₀ scale.

Finally, we consider the West African Plants Initiative to have been a significant success because it has demonstrated the collective, emergent properties of broad, international collaborations. In this project, botanists from four West African countries cooperated and coordinated closely to generate a product that is more significant to science than any could have generated alone. The West African team has in turn coordinated efforts with institutions in Europe and North America, receiving significant and invaluable support from European herbaria in terms of large stores of images of herbarium sheets from across the West African region.

Other venues and steps forward

Building and enhancing capacity in biodiversity informatics among West African students and scientists is critically important to effective and efficient data mobilization, data sharing, and data use.

Abundant new biodiversity data may prove of little importance to the science infrastructure of the region if not accompanied by new scientific talent that allows the data to be put to good use. Several of our project participants have already initiated further studies to fill this personnel gap, but support for their studies and support for institutions to receive them upon completion of their studies are still uncertain.

Once captured, primary biodiversity data (as in this project) must be cleaned and georeferenced to yield high-quality data resources. Such data will have many immediate applications such as documenting basic biodiversity patterns (Arita et al., 2008), identifying priority areas for conservation efforts (Loyola et al., 2007), detecting biotic change (Peterson et al., 2015), anticipating biotic responses to local and global change (Kearney et al., 2010), and guiding sustainable development and decision-making about natural resources and the environment (Chapman, 2005; Sousa-Baena et al., 2013). As such, skills development in data management and data improvement are key to improving captured data and developing high-quality botanical information resources. A further need is skills in data analysis, such as in multivariate statistics, place-prioritization efforts, and ecological niche modeling; such skills are necessary if mobilized data are to be used to inform national and regional decision-making (see tools presented in Peterson and Ingenloff, 2016).

The scope of this particular project can and should be expanded, in terms of both West African and European/North American institutions, to improve geographic coverage and record density. Such an expansion in participation will also enhance the georeferencing process, because national scientists presumably have better knowledge of the geography of their own countries. West African countries such as Senegal, Guinea, and Burkina Faso have significant in-country collections, and all West African countries have significant representation in collections in European and North American institutions for which data have yet to be captured. Significant holdings known to us include the herbaria of the Natural History Museum (London), the Smithsonian Institution, and Harvard University, and we hope that these institutions will participate in future chapters of this initiative.

A new model for biodiversity informatics progress

This paper outlines, illustrates, and expands on what can be a new model for progress in biodiversity informatics, particularly for developing countries that are often rich in biological diversity but poor in biodiversity information. The quandary is that scientists and policy-makers in those developing countries are often highly motivated and active in efforts to develop information resources, yet data are too frequently not available or not abundant; starting from zero with new sampling is not possible owing to cost, logistical considerations, land use change, and/or politics. However, institutions in Europe or North America, and sometimes elsewhere, often hold specimens with rich information coming from those same countries, but often have not captured those data completely, leaving information dormant and unavailable for applications in science, conservation, and natural resources policy.

In the solution that we explore herein, the scientists and policy-makers in the countries and regions of origin of the biodiversity data are those who motivate the capture of the biodiversity data from their countries and regions. Institutions holding data

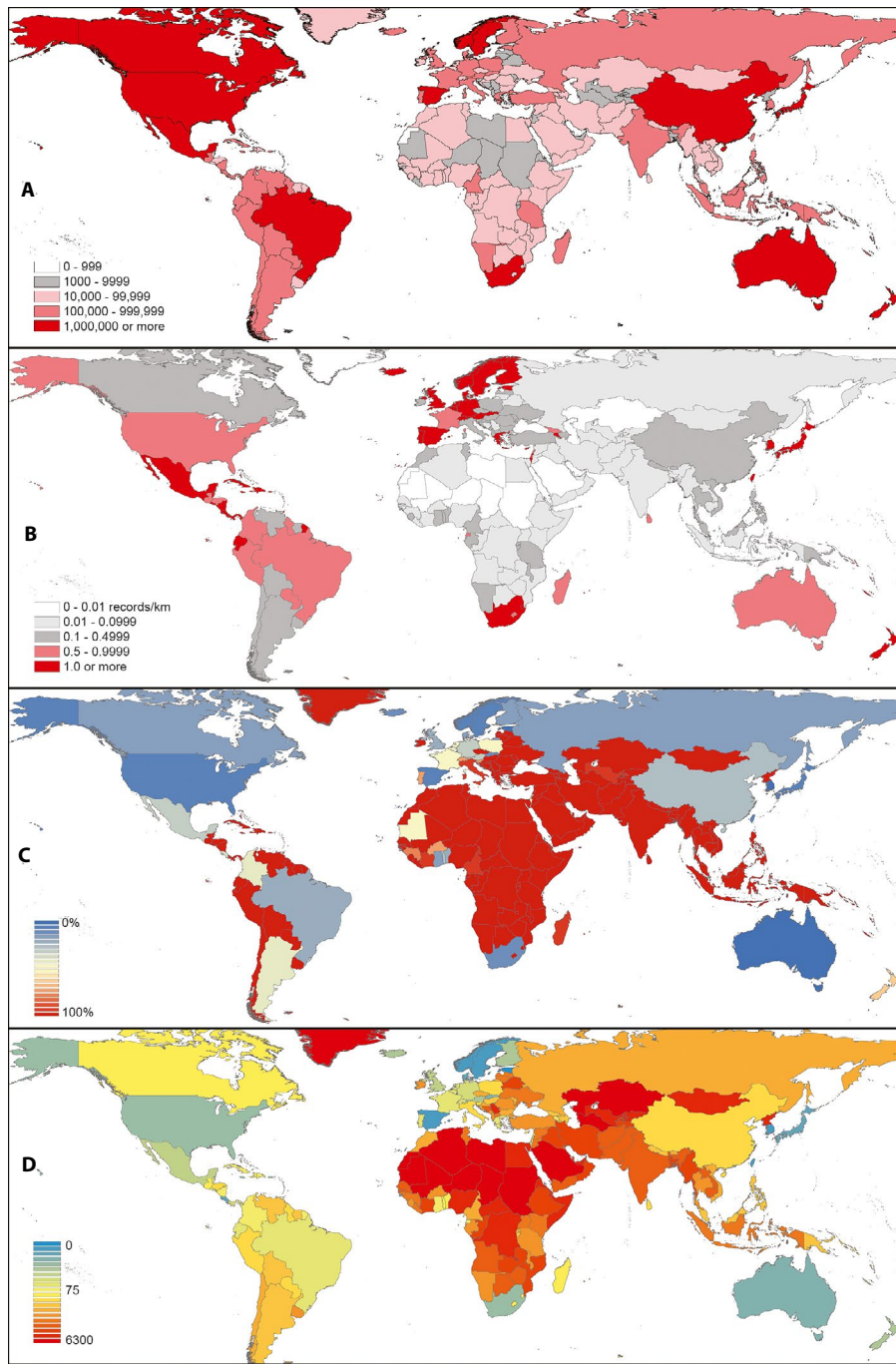


FIGURE 5. Four aspects of a global view of the density of digital accessible knowledge based on plant specimens. (A) Number of records in terms of raw numbers. (B) Record density on a per-km² basis. (C) 100× the proportion of a country's plant records that are made accessible by a different country. (D) Ratio between percent external and record density as an index to the need of a given country for mobilization of data records from outside countries.

and specimens continue to hold those materials and to curate and care for them, as they have in many cases for more than a century, assuring permanency of the resources. These specimen-rich institutions provide access to persons and institutions from the regions of origin to promote their digitization. All involved—both

the countries of origin and the institutions holding the specimens—cooperate in refinement and “publication” of the resulting data, and in the scientific interpretation and analysis of the new data resources.

Note that this model is the reverse of the traditional model, in which the institutions holding the specimens create the information resources that are used by the rest of the world. Here, instead, the motivation and implementation is carried out by scientists often based in the developing world, which heretofore have been rich in biodiversity but poor in biodiversity information. Given dramatic processes of globalization of scientific knowledge and ability (Dong et al., 2017), this new model has considerable promise to accelerate and improve the process of generating biodiversity information.

ACKNOWLEDGMENTS

We dedicate this paper to the impressive and foundational work that has been achieved and the trails that have been blazed by the Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). CONABIO provided a stellar, decades-long success story that has illustrated the concept presented herein, and inspired the present work. We highly appreciate work done by our data capturers: Anthony Adu-Gyamfi, Sabastian W. Yunkavi, Sodjinou E. Komlan, and Emmanuel Chukwuma. We are grateful to Kim Watson and Yannick Klomberg for training project interns and data capturers in specimen imaging and record capture using BRAHMS, respectively. Daniel Jiménez-García, Ali Khalighifar, Qiao Huijie, Vijay Barve, and Marlon E. Cobos all provided invaluable help in managing big data sets. The JRS Biodiversity Foundation provided support for the project in the form of a generous research grant to the University of Ghana. Thanks also to the European Union for financial support through the GBIF-BID program (to A.A.) for biodiversity informatics research.

DATA AVAILABILITY

The data resources generated by this project are available as online resources via the Global Biodiversity Information Facility (<https://doi.org/10.15468/e8rhqm> [Asase, 2018]; <https://doi.org/10.15468/9zczig> [Ugbogu et al., 2019]). Other data sets, yet to be published, are available from the authors upon reasonable request.

LITERATURE CITED

- Ariño, A. H., V. Chavan, and N. King. 2011. The biodiversity informatics potential index. *BMC Bioinformatics* 12(Suppl 15): S4.
- Arita, H. T., J. A. Christen, P. Rodríguez, and J. Soberón. 2008. Species diversity and distribution in presence-absence matrices: Mathematical relationships and biological implications. *American Naturalist* 172: 519–532.
- Asase, A. 2018. Plants of Ghana. Ghana Herbarium. GBIF occurrence dataset. <https://doi.org/10.15468/e8rhqm> [accessed 10 January 2020].
- Asase, A., and A. T. Peterson. 2016. Completeness of digital accessible knowledge of the plants of Ghana. *Biodiversity Informatics* 11: 1–11.
- Balmford, A., P. Crane, A. Dobson, R. E. Green, and G. M. Mace. 2005. The 2010 challenge: Data availability, information needs and extraterrestrial insights. *Philosophical Transactions of the Royal Society B* 360: 221–228.
- Bini, L. M., J. A. F. Diniz-Filho, T. F. Rangel, R. P. Bastos, and M. P. Pinto. 2006. Challenging Wallacean and Linnaean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions* 12: 475–482.
- Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-Qing, N. E. Clark, K. O'Connor, and G. M. Mace. 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology* 8(6): e1000385.
- Canhos, V. P., S. de Souza, R. D. Giovanni, and D. A. L. Canhos. 2004. Global biodiversity informatics: Setting the scene for a “new world” of ecological forecasting. *Biodiversity Informatics* 1: 1–13.
- Chapman, A. D. 2005. Uses of primary species-occurrence data, version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. Available at <http://www.gbif.org/document/80545> [accessed 26 December 2019].
- Collen, B., M. Ram, T. Zamin, and L. McRae. 2008. The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science* 1(2): 75–88.
- Comisión Nacional Para el Conocimiento y Uso de la Biodiversidad (CONABIO). 2012. Dos décadas de historia 1992–2012. Comisión Nacional Para el Conocimiento y Uso de la Biodiversidad, Mexico City, Mexico.
- Dauby, G., R. Zaiss, A. Blach-Overgaard, L. Catarino, T. Damen, V. Deblauwe, S. Desein, et al. 2016. RAINBIO: A mega-database of tropical African vascular plants. *PhytoKeys* 74: 1–18.
- Dong, Y., H. Ma, Z. Shen, and K. Wang. 2017. A century of science: Globalization of scientific collaborations, citations, and innovations. In *Proceedings of Knowledge Discovery and Data Mining (KDD) 17*, 13–17 August 2017, Halifax, Nova Scotia, Canada.
- Ganglo, J. C., and S. B. Kakpo. 2016. Completeness of digital accessible knowledge of plants of Benin and priorities for future inventory and data discovery. *Biodiversity Informatics* 11: 23–29.
- Global Biodiversity Information Facility. 2017a. GBIF occurrence download. <https://doi.org/10.15468/dl.ujfhsu> [7 November 2017].
- Global Biodiversity Information Facility. 2017b. GBIF occurrence download. <https://doi.org/10.15468/dl.yjr4tf> [accessed 10 January 2020].
- Guralnick, R. P., J. Wieczorek, R. Beaman, R. J. Hijmans, and the BioGeomancer Working Group. 2006. BioGeomancer: Automated georeferencing to map the world's biodiversity data. *PLoS Biology* 4: 1908–1909.
- Kearney, M. R., B. A. Wintle, and W. P. Porter. 2010. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters* 3: 203–213.
- Koffi, K. J., A. F. Kouassi, C. Y. A. Yao, A. Bakayoko, I. J. Ipou, and J. Bogaert. 2015. The present state of botanical investigations in Côte d'Ivoire. *Biodiversity Informatics* 10: 56–64.
- Loyola, R. D., U. Kubota, and T. M. Lewinsohn. 2007. Endemic vertebrates are the most effective surrogates for identifying conservation priorities among Brazilian ecoregions. *Diversity and Distributions* 13: 389–396.
- Mannion, P. D., P. Upchurch, R. B. J. Benson, and A. Goswami. 2014. The latitudinal biodiversity gradient through deep time. *Trends in Ecology and Evolution* 29: 42–50.
- Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.
- Navarro-Sigüenza, A. G., A. T. Peterson, and A. Gordillo-Martínez. 2002. A Mexican case study on a centralised database from world natural history museums. *CODATA Journal* 1: 45–53.
- Navarro-Sigüenza, A. G., A. T. Peterson, and A. Gordillo-Martínez. 2003. Museums working together: The atlas of the birds of Mexico. *Bulletin of the British Ornithologists' Club* 123A: 207–225.
- Nelson, G., D. Paul, G. Riccardi, and A. R. Mast. 2012. Five task clusters that enable efficient and effective digitization of biological collections. *ZooKeys* 209: 19–45.
- Peterson, A. T., and K. Ingenloff. 2016. Biodiversity informatics training curriculum, Version 1.2. *Biodiversity Informatics* 10: 65–74.
- Peterson, A. T., and J. Soberón. 2017. Essential biodiversity variables are not global. *Biodiversity and Conservation* 27: 1277–1288.
- Peterson, A. T., S. Knapp, R. Guralnick, J. Soberón, and M. T. Holder. 2010. The big questions for biodiversity informatics. *Systematics and Biodiversity* 8: 159–168.
- Peterson, A. T., A. G. Navarro-Sigüenza, E. Martínez-Meyer, A. P. Cuervo-Robayo, H. Berlanga, and J. Soberón. 2015. Twentieth century turnover of Mexican endemic avifaunas: Landscape change versus climate drivers. *Science Advances* 1: e1400071.
- Peterson, A. T., A. G. Navarro-Sigüenza, and A. Gordillo-Martínez. 2016a. The development of ornithology in Mexico and the importance of access to scientific information. *Archives of Natural History* 43: 294–304.
- Peterson, A. T., A. G. Navarro-Sigüenza, and E. Martínez-Meyer. 2016b. Digital accessible knowledge and well-inventoried sites for birds in Mexico: Baseline sites for measuring faunistic change. *PeerJ* 4: e2362.
- Pignatelli, M., S. Romaniuc-Neto, S. D. Souza, S. Chagnoux, and D. A. L. Canhos. 2012. Saint-Hilaire virtual herbarium, a new upgradeable tool to study Brazilian botany. *Adansonia* 35: 7–18.
- Rands, M. R., W. M. Adams, L. Bennun, S. H. M. Butchart, A. Clements, D. Coomes, A. Entwistle, et al. 2010. Biodiversity conservation: Challenges beyond 2010. *Science* 329: 1298–1303.
- Rees, T., L. Vandepitte, W. Decock, and B. Vanhoorne. 2017. IRMNG 2006–2016: 10 years of a global taxonomic database. *Biodiversity Informatics* 12: 1–44.
- Sarukhán, J., T. Urquiza-Haas, P. Koleff, J. Carabias, R. Dirzo, E. Ezcurra, S. Cerdeira-Estrada, and J. Soberón. 2014. Strategic actions to value, conserve, and restore the natural capital of megadiversity countries: The case of Mexico. *BioScience* 65(2): 164–173.
- Siebert, S. J., and G. F. Smith. 2004. Lessons learned from the SABONET Project while building capacity to document the botanical diversity of southern Africa. *Taxon* 53(1): 119–126.
- Smith, G. F. 2004. The African Plants Initiative: A big step for continental taxonomy. *Taxon* 53(4): 1023–1025.
- Soberón, J., and J. E. Llorente-Bousquets. 1993. La Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). *Revista de la Sociedad Mexicana de Historia Natural* 44: 3–17.
- Sosef, M. S. M., G. Dauby, A. Blach-Overgaard, X. van der Burgt, L. Catarino, T. Damen, V. Deblauwe, et al. 2017. Exploring the floristic diversity of tropical Africa. *BMC Biology* 15: 15.
- Sousa-Baena, M. S., L. C. Garcia, and A. T. Peterson. 2013. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* 20: 369–381.
- Stropp, J., R. J. Ladle, A. C. M. Malhado, J. Hortal, J. Gaffuri, W. H. Temperley, J. O. Skoien, and P. Mayaux. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology and Biogeography* 25(9): 1085–1096.
- Thiers, B., M. Tulig, and K. A. Waston. 2016. Digitization of The New York Botanical Garden Herbarium. *Brittonia* 68(3): 324–333.
- Ugbogu, O., O. Osiyemi, E. Chukwuma, K. Adeniji, A. Odewo, T. Okanlawon, T. Soyewo, et al. 2019. Capture of primary biodiversity data for West African plants. Version 1.6. Forestry Research Institute of Nigeria. GBIF occurrence dataset. <https://doi.org/10.15468/9czcig> [accessed 10 January 2020].
- Wieczorek, J., D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7: e29715.
- Yesson, C., P. W. Brewer, T. Sutton, N. Caithness, J. S. Pahwa, M. Burgess, W. A. Gray, et al. 2007. How global is the global biodiversity information facility? *PLoS ONE* 2: e1124.