

ASSESSMENT OF BIODIVERSITY DATA HOLDINGS AND USER DATA NEEDS FOR GHANA

ALEX ASASE^{1*} AND GLADYS O. SCHWINGER²

¹*Department of Plant and Environmental Biology, University of Ghana, P. O. Box LG 55, Legon, Ghana.* ²*Institute for Environment and Sanitation Studies, University of Ghana, P. O. Box LG 209, Legon, Ghana.* *Email: aasase@ug.edu.gh.

Abstract.—Data on biodiversity are important to addressing the challenges of sustainable development, and for decision-making about natural resources and environments. Biodiversity information, when mobilized and shared openly, has the potential to impact science and conservation positively. However, biodiversity data mobilization is expensive, such that data mobilization and sharing activities must be prioritized to meet the needs of the user community. In this study, we undertook a detailed assessment of biodiversity data holdings and user needs in Ghana through semi-structured questionnaire interviews, and focus-group discussions in the form of a workshop. Most biodiversity data-holding organizations were at preliminary stages of digital biodiversity data mobilization and sharing. Taxonomic, checklist, and geographic data on plants and animals were identified as most needed. Priority thematic needs were as regards protected areas, invasive alien species, threatened species, economic species (timber and non-timber forest products), and pathogens and diseases. Human and infrastructural capacities, and sustainable coordination were identified as the major challenges to biodiversity data management. This study provides a detailed case study of how assessing biodiversity data holdings and user data needs can be used to strategize biodiversity data mobilization, data publication, and data use activities.

Key words.—biodiversity, primary biodiversity data, data mobilization, digitization, Ghana

Biological diversity may be defined as the full variation of living organisms from all sources including, *inter alia*, terrestrial, marine, and other aquatic ecosystems and the ecological complexes of which they are a part. This term thus includes diversity within species, between species, and of ecosystems (CBD, 2001). It includes genetic diversity, species diversity, ecosystem diversity, and associated evolutionary and ecological processes. Biodiversity is a compound word derived from biological diversity, and therefore is considered to have the same meaning. Biodiversity is important for human wellbeing: it provides tangible benefits, such as food, clothing, and shelter, as well as intangible benefits such as climate amelioration and clean water (Brauman et al., 2007). However, biodiversity is being lost at unprecedented rates owing to a plethora of factors: deforestation, agricultural expansion, habitat loss, timber extraction, firewood collection, and mineral extraction (Norris et al., 2010). Data on biodiversity are crucial to addressing the challenges of sustainable development and decision-making about natural resources and environments (Chapman, 2005; Sousa-Baena et al., 2013).

Biodiversity data include data on species inventories, distributions, images, sounds, specimens,

and ecological interactions, as well as descriptions of datasets (i.e., metadata) (Costello et al., 2013). Biodiversity data are basically of two kinds (i.e., primary and secondary), and can be numerical, categorical (e.g., species or place names), or media-based (Costello et al., 2013). Primary biodiversity data are data records that document the occurrence of a particular species at a place at a point in time. In contrast, secondary biodiversity data represent summaries, interpretations, or syntheses of primary biodiversity data. Secondary biodiversity data, such as species atlases and range maps from field guides, often include subjective elements that reduce their utility when compared to primary biodiversity data. As such, primary biodiversity data have many applications: documenting basic biodiversity patterns (Guralnick and Hill, 2009), identifying priority areas for conservation efforts (Myers et al., 2000), providing baseline information for detection of biotic change (Peterson et al., 2015), and supporting modeling efforts that anticipate biotic responses to local and global change (Ehrlén and Morris, 2015).

Sources of primary biodiversity data are many, including labels associated with specimens in research collections of natural history museums and herbaria, and data from field studies and

observations made by scientists and researchers (Peterson et al., 2011), as well as data from field observations made by citizen scientists (Asase and Peterson, 2016). Types of primary biodiversity data include primary occurrence data that document presence (and absence) of organisms such as those on labels on herbarium sheets (Peterson et al., 2011); sample-based data that have information on species occurrence and their abundance, such as those from ecological plot inventories; and multimedia data such as sound, images and videos. Biodiversity data offer greatest information when they are integrated and used with other data types, such as environmental data and socioeconomic data, to address the most pressing questions in biodiversity and sustainability science (Faith et al., 2013). Nonetheless, the importance of natural history collections has been demonstrated clearly (e.g., Ariño et al., 2013).

At this point in time, biodiversity data must be digitized and shared openly (Costello et al., 2013). Certainly, past decades have witnessed massive progress in digitization of biodiversity data thanks to advances in information technology, development of efficient data digitization workflows, and changes in policies of owners of primary biodiversity data (Asase and Peterson, 2016). Biodiversity data digitization and sharing of natural history collections include several stages: pre-digitization preparation, advance curation, image capture, processing and storage, capture of data records from either images or specimens, georeferencing, data cleaning, and data publication (Nelson et al., 2015). It is expensive to produce and share biodiversity data, and not all data are fit for all uses (Hills et al., 2010). Consequently, tasks of digitization and sharing of primary biodiversity data must be prioritized to meet the needs of the user community. Surveys of biodiversity data user needs are important to understanding data needs across diverse user communities (Gaiji et al., 2013), and to measuring scientific and policy contributions of data mobilized (Ariño et al., 2013).

Here, we undertook a detailed assessment of biodiversity data holdings and data needs across the user communities of scientists, researchers, curators of natural history collections, non-governmental organizations (NGOs), and policy-makers in Ghana. We explored how detailed assessment of biodiversity holdings and user needs of a country can be used to guide biodiversity data mobilization, data sharing, and data use activities. We used Ghana as a case study in view of its active involvement in biodiversity data

mobilization activities and biodiversity informatics initiatives, such as the Global Biodiversity Information Facility (GBIF).

METHODS

Data were collected via a combination of two common survey methods: semi-structured interviews, and focus group discussions. Interviews were achieved via distribution of a questionnaire to major biodiversity stakeholder organizations in Ghana for completion. Stakeholders included university departments, government agencies, research institutions, NGOs, and other groups. Questionnaires were administered either in person or were sent to organizations for them to complete. We did not attempt online questionnaires owing to unreliable Internet connectivity.

The questionnaire was designed to assemble information on three main areas: (1) the profile of organization, (2) its data holdings, and (3) its biodiversity data needs. For data holdings, our focus was on the status of the holdings, strategies towards data mobilization, the digitization landscape, and attitudes about data publication, as well as data preservation and archiving. We focused on data needs regarding five broad categories: checklist, taxonomic data, geographic data, ecological data, and educational data (i.e., biodiversity data for public education, and training materials). Within each broad category, we assessed level of importance, availability, and sustainability of the data source for major taxonomic groups (plants, animals, fungi, microorganisms, algae). In total, we received 22 fully completed questionnaires from stakeholder organizations out of >50 sent out. The stakeholders were 50% from academia and research institutions, 40% from governance and policy institutions, and 10% from NGOs and other groups, and thus were broadly representative of major biodiversity stakeholder groups in Ghana.

Focus group discussions were in the form of a national biodiversity stakeholders' workshop. The aim of this workshop was to arrive at national consensus on biodiversity information needs for Ghana, including identifying challenges. The workshop included three breakout sessions among different stakeholder types: (a) academic and research, (b) governance and policy, and (c) NGO and others. About 61% of the stakeholders that completed the questionnaire attended the workshop, and 31 stakeholder groups in total attended. Each of the three

groups was tasked to brainstorm on data needs, opportunities, threats, and challenges in biodiversity data mobilization for Ghana. After initial group discussions, the groups reconvened and discussed major findings of each group. The results of the discussions were summarized in the form of a “Strengths, Weaknesses, Opportunities and Threats” (SWOT) analysis, as well as with summaries of priority data needs and data challenges. Workshop participants were well-trained people with capacity to answer questions and verify answers regarding biodiversity information in Ghana; each participant presented views of their respective institutions.

Data were entered into Microsoft Excel, checked for consistency, and cleaned for errors such as duplications and name variants. Using the PivotTable function in Excel, we analyzed responses to survey questions in terms of frequencies. Summary statistics and frequencies data were compiled, and presented using appropriate visual aids.

RESULTS

Data holdings, digitization, and publication

Of the 22 organizations interviewed, 15 had biological collections, 4 had only field data, and 3 were data users only. About 89.0% of the 19 organizations with data had clearly defined purposes for mobilizing data, and 36.8% had already assessed the scope and extent of biodiversity data digitization (Fig. 1). Similar proportions (47.4%) of the organizations had already achieved pre-digitization activities or had not undertaken pre-digitization activities; the remaining 5.3% did not know about pre-digitization activities. Institutional data policies (e.g., as regards data sharing) were present in 26.3% of the organizations; most (63.2%) of the organizations had no such policies. Most (68%) of the organizations had no data management systems in place.

About 26% of the organizations had a digitization workspace, whereas 68.4% had no digitization workspace, while 5.3% were not sure (Table 1). Ten of the organizations were knowledgeable about the proportions of their holdings that had been digitized, whereas 47.4% did not know. Pre-curation digitization activities had been carried out by only 36.8% of organizations, whereas 57.9% of organizations had not pursued such activities. Digitization processes and technologies existed for 21.1% of organizations, whereas a majority (73.7%) of organizations had no such processes and technologies in place. Out of the 19 data holders, only

42.1% had well-defined workflows for digitization. Most (73.7%) organizations did not have staff adequately trained and equipped for digitization. More than half (57.9%) of organizations had no specifications for data quality control and standards. However, most (73.7%) organizations had selected a data platform for digitization, and formats for digital data storage were available for 42.1% of organizations.

Regarding data publication, 37.1% of organizations had considered end-users and web publication needs, whereas the rest had not made such considerations. About 74.1% of organizations had not selected any data publication tool or data licensing option, and 68.3% have no online platform for sharing data. Access to long-term archival repositories and safeguards against obsolescence of data formats and applications were not available in most of the organizations.

Data needs and challenges

The questionnaire interviews were helpful in identifying biodiversity data needs of the organizations interviewed. Data on plants were most frequently cited as needed, followed by those on animals; data on fungi were least mentioned as needed (Fig. 2). Data types most frequently mentioned as needed were taxonomic data (26%), followed by checklists (25%) and geographic data (25%), then ecological data (18%), and lastly educational data (6%). At the same time, the most available data type was taxonomic data, while checklists were the least available (Fig. 3). Details on data needs according to data type and the major taxonomic groups are presented in Table 2. About 49% of data sources were regarded as sustainable; 44.4% were unsustainable; for 6.6%, it was unknown.

The stakeholders' workshop was important in highlighting biodiversity data needs at the national level. Results of the SWOT analysis revealed strengths (8 points), opportunities (6 points), weaknesses (5 points), and threats (2 points) regarding biodiversity data in Ghana (Table 3). The data requirements of the different user-groups indicated above varied (Table 4). In general, stakeholders identified by consensus 7 priority areas of data need areas: (a) up to date biodiversity inventory data on protected areas such as forest reserves and national parks; (b) geographic data on biodiversity under land-use and climate change scenarios; (c) national red lists of biodiversity for

major groups such as plants, mammals, birds, and insects; (d) pathogens and microbes, and their effects on economic crops and livestock; (e) harvested biodiversity particularly timber and Non-Timber Forest Products (NTFPs); (f) economic and useful biodiversity such as medicinal plants and edible mushrooms; and (g) invasive alien species, and their presence and distribution. Major challenges to biodiversity data included poor human and infrastructural capacity, lack of funds for data acquisition, and lack of sustainable coordination.

DISCUSSION

Biodiversity data mobilization landscape

The need for biodiversity data to be made accessible, discoverable, and integrated cannot be overstated, as biodiversity research is rapidly becoming a data-intensive science (Kelling et al., 2009). Biodiversity data mobilization is an expensive enterprise, but once mobilized, the data can be of great value (Borgman, 2007). The results of this study show that the biodiversity data mobilization landscape in Ghana is at preliminary stages of operation, as most organizations have not yet begun strategizing for data mobilization. It is always important that proper strategies and planning be in place before embarking on data mobilization (Frazier et al., 2008) to assure effective implementation of project tasks, and ultimately the success of projects. Such strategies should consider the various project stakeholders, and their roles at the different phases of the project implementation. Long-term policies and a strong vision towards gathering and using biodiversity data are required for any country such as Ghana. Currently, no such policies and vision have been articulated for Ghana, which is worrying.

The present study has been helpful in characterizing the major biodiversity data holdings in Ghana. Biodiversity data-holders may be either cooperative or non-cooperative. Cooperative biodiversity data-holders are those willing to collaborate and share their data resources openly, whereas non-cooperative biodiversity data-holders do not want to participate and share their data resources. It is unclear which of the many biodiversity data-holders identified in Ghana belongs to each of the two types; as such all data-holders must be encouraged, especially those that prove to be non-cooperative.

Information that is shared and accessible has the potential to impact science and conservation, as well as the care and curation of specimens (Asase and

Peterson, 2016). Barriers to sharing biodiversity data could be psychological and behavioral (including legal barriers), or may relate to describing information and data, or may spring from inadequate strategies and resources. Non-cooperative data-holders would most likely become at least intermediate biodiversity-data holders willing to collaborate and share their data resources after the barriers have been identified, and they have been highly motivated and assured of incentives.

Another area of importance in terms of biodiversity data mobilization is the large amounts of data associated with significant collections of Ghanaian biological and paleontological specimens held elsewhere in the world. Data repatriation from European and North American organizations with large collections from Ghana is an important potential source data on Ghanaian biodiversity. For example, European and North American institutions, such as Naturalis Biodiversity Centre in the Netherlands, Missouri Botanic Gardens in the United States, and the Royal Botanic Gardens in the United Kingdom have digital images and data records of botanical collections from Ghana that they could make openly to the appropriate institutions in Ghana. Indeed, Naturalis Biodiversity Centre has already provided large series of digital images of botanic collections from Ghana to the Ghana Herbarium at University of Ghana.

Digitization of biodiversity data refers to capture of information in electronic form from checklists, field notebooks, or specimens, or may be extracted from publications, documents, or other media. It may refer to electronic capture of an image of an object, or it can also refer to capture of textual information about an object or extracted from an object that contains text (Frazier et al., 2008). The advantages of biodiversity digitization are many: broad dissemination of data via open and accessible platforms; enabling natural history collections to be studied in different ways, including from outside of the museum or herbarium; enhancement of curatorial activities. This step also reduces future time spent on transcription of data records from specimens, and enhances visibility of institutions sharing data.

Biodiversity digitization workflows and protocols have been developed to maximize rates of specimen digitization without sacrificing the most useful information on each specimen (Tulig et al., 2012). According to Nelson et al. (2015), efficient workflows provide the foundation for successful digitization of

biodiversity collections and foster mobilization of increased quantities of specimen data for scientific research, natural resource management, education, and policy-making. Application of available workflows for digitization of natural history collections in Ghana will lead to better refinement, and additions that will increase availability of mobilized biodiversity data, and enhance specimen-based research in the country. The Ghana Herbarium at the University of Ghana has started using some of these protocols. It is highly encouraged that other biodiversity data holding institutions in Ghana, particularly those currently involved in biodiversity informatics projects e.g., Plant Genetic Resource Research Institute of the Council for Scientific and Industrial Research (CSIR-PGRRI), A Rocha Ghana and Conservation Alliance will apply such protocols in their digitization programmes.

Data needs and data use

Surveys of biodiversity data needs of various organizations and user groups can be a useful means of identifying common data needs *versus* priority data needs. The present study provided insights into data needs of various organizations and user groups across Ghana. Most organizations interviewed needed data on the taxonomy and geography of taxa. High demand for taxonomic data, particularly as regards identifications, was not surprising because it is fundamental to communicating about biodiversity (Judd, 2008). Also, data on names (taxonomic data) and place (geographic data) are important to exploring joint efforts that relate directly to applications such as ecological niche modeling and species distribution modelling (Peterson et al., 2018). Data on plants were the most required probably because of high human dependence on plants and plant products.

Data needs identified as priority areas at the national level in this study concerns protected areas, invasive alien species, threatened species, economic species, and pathogens and diseases. These data needs priority areas fall within the areas of interest of international organizations and bodies such as GBIF, International Union of Conservation of Nature, United Nation Environment Programme–World Conservation Monitoring Centre, Global Earth Observation Biodiversity Observation Network, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), and

Convention on Biological Diversity. Developing programmes and collaborating with these organizations will be useful in mobilizing primary biodiversity data about Ghana. Other priority areas should include agricultural biodiversity, given the fact that Ghana is an agrarian country, and aquatic biodiversity, because it is less studied than terrestrial biodiversity. It is also important that data needs be aligned to meet national and international obligations such as the Clearing House Mechanism and Aichi Targets 2020 of the Convention of Biological Diversity, and the Sustainable Development Goals, to which Ghana is a signatory. Another area worthy of consideration are the so-called “essential biodiversity variables” for monitoring biodiversity change (Pereira et al., 2013).

Our SWOT analysis was useful in identifying strengths, weaknesses, opportunities and threats regarding biodiversity data in Ghana. This information can be used in formulating strategic management decisions concerning biodiversity data and ecosystem services as they relate to the societal needs for Ghana. For example, IPBES has underscored the importance of integrating biodiversity data with data on ecosystem services. Biodiversity data should be used in making policy decisions to support sustainable development; for research at universities, colleges, and research institutions; and to enable training at different levels of the educational ladder in Ghana. Mainstreaming biodiversity data into national policy decisions could provide insights about how data are used, and could also demonstrate the value of digital mobilization of biodiversity data. Unfortunately, as pointed out in the *Ghana Shared Growth and Development Agenda (GSGDA) II* (2014-2017) policy document, Ghana sees weak integration of biodiversity issues in decision-making, especially at the local level, in Ghana (NDPC, 2014). The UNEP-WCMC Connect project¹ in Ghana, Mozambique and Uganda is one such model project on mainstreaming biodiversity into national policy decisions. Awareness about the importance of biodiversity at schools could promote biodiversity conservation into the future.

Training and capacity enhancement

Limited human and infrastructural capacities were identified as challenges to biodiversity data mobilization in Ghana. For example, ~74% of stakeholder organizations in this study do not have

¹ <http://gef-connect.web-staging.linode.unep-wcmc.org/>.

staff adequately trained to digitize and share biodiversity data. This scenario is worrying, as it may lead to data quality issues, an issue of grave concern in biodiversity data management (Veiga et al., 2017). Adequate human expertise and skills are required to produce research-grade data for use: data must be properly captured, georeferenced, and cleaned before publication, such that data shared will have immediate applications (Peterson et al., 2018).

The biodiversity informatics community has developed tools and standards to support these challenges, such as the Botanical Research and Herbarium Management Systems (BRAHMS²) and SPECIFY³ for capturing data records; GeoLocate has been developed for geo-referencing data records (Guralnick et al., 2006); DarwinCore was developed as a standard for publishing and integrating biodiversity information (Wieczorek et al., 2012); and the Integrated Publishing Toolkit (IPT) is a tool for sharing biodiversity datasets via the Internet (Robertson et al., 2014). Training in use of these tools and standards is required to deliver *usable* DAK (Peterson et al. 2018). Another key area of human capacity is in data analysis—e.g., multivariate statistics, place-prioritization efforts, and ecological niche modelling—as such skills are necessary if data mobilized are to be used to inform national and regional decision-making.

Biodiversity informatics is a young science, with few or no approved textbooks or academic programs developed. However, many biodiversity informatics initiatives and resources exist that could be of help in addressing the biodiversity informatics capacity challenges for Ghana: e.g., the Biodiversity Informatics Training Curriculum (BITC) (Peterson and Ingenloff, 2015), and various opportunities available through GBIF, GBIF-Africa, and Biodiversity Information Standards (TDWG). A long-term solution to the human capacity challenge is to develop sustainable biodiversity science programs in Ghanaian universities and colleagues. In Africa, a pilot biodiversity informatics programme has started at the University of Abomey-Calavi in Benin, and another is planned at the University of Western Cape in South Africa.

Mobilizing biodiversity data to ensure maximum access and use requires a robust and easily usable infrastructure (Robertson et al., 2014). It is therefore important that the various Ghanaian biodiversity

organizations should consider investing in this area of science. Basic equipment and software for data digitization, data storage, and data analysis are useful in achieving desired outcomes in biodiversity data management. Perhaps a long-term solution to this problem is for the national government to commit to establishment of a sustainable central organization with the needed facilities, more or less following the example set by the South African Biodiversity Institute in South Africa.

Biodiversity data and expertise are currently unevenly distributed in Ghana such that most institutions with data and expertise are in the southern half of the country. For example, of the six well-developed herbaria in Ghana, only the recently established Savanna Herbarium at the University for Development Studies (Nyankpala) is situated in northern half of the country. This situation is not surprising, as Ghanaian biodiversity is richest in the southern part of the country, especially in the southwest, where well-preserved forest remnants can be found, whereas much of northern Ghana is covered with savanna (MES, 2002). The geographic distribution of biodiversity holding data institutions and expertise is important to how decisions are made about biodiversity information management in Ghana.

Conclusions

This study presents the first detailed assessment of biodiversity holdings and user needs for Ghana. Although a number of biodiversity stakeholders did not participate in the survey, this study has highlighted pertinent issues about the biodiversity data landscape in Ghana. Most biodiversity data-holding organizations in Ghana are at preliminary stages of data mobilization, and human capacity and infrastructure, as well as sustainable coordination, are the major challenges to data mobilization, data publication, and data use. We did not solicit responses regarding biodiversity data for different ecosystems, or regarding ethnobiological and molecular data, which can be considered in future studies. A next logical step will be to undertake data gap analyses, to identify discrepancies between current ideas and the state of the entire biodiversity science enterprise in Ghana. An analysis of completeness of digital accessible knowledge exists for the plants of Ghana (Asase and Peterson, 2016), but completeness of

²<http://herbaria.plants.ox.ac.uk/bol/brahms/software>.

³<http://www.sustain.specifysoftware.org>.

digital accessible knowledge for other taxa has not been assessed. This study illustrates how assessing biodiversity data holdings and user data needs can be used to strategize for biodiversity data mobilization, data publication, and data use activities, for a country and/or for a taxon. Our findings are relevant to diverse biodiversity stakeholders, such as researchers, museum curators, and policy makers, in formulating strategic ideas and policies concerning their biodiversity data resources.

ACKNOWLEDGMENTS

The authors are thankful to all the biodiversity stakeholders in Ghana who participated in the study. We are most grateful to our funders: GBIF, Biodiversity Information for Development (Contract #: BID-AF2015-0032-NAC; supported by the European Commission), and the JRS Biodiversity Foundation. We are grateful to two anonymous reviewers of this manuscript for their insightful comments, and to Town Peterson for edits to improve this manuscript.

REFERENCES

- Ariño, A.H., V. Chavan, and D.P. Faith. 2013. Assessment of user needs of primary biodiversity data: Analysis, concerns, and challenges. *Biodiv. Inf.* 8:59-93.
- Asase, A., and A.T. Peterson. 2016. Completeness of digital accessible knowledge of the plants of Ghana. *Biodiv. Inf.* 11:1-11.
- Borgman C.L. 2007. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press, Boston.
- Brauman, K.A., G.C. Daily, T.K.E. Duarte, and H.A. Mooney. 2007. The nature and value of ecosystem services: An overview highlighting hydrologic services. *Annu. Rev. Environ. Resour.* 32:67-98.
- CBD. 2001. *Handbook of the Convention on Biological Diversity*. Earthscan Publications Ltd., London.
- Chapman, A.D. 2005. *Principles and methods of data cleaning: Primary species and species occurrence data, version 1.0*. Global Biodiversity Information Facility, Copenhagen.⁴
- Costello, M.J., W.K. Michener, M. Gahegan, Z.-Q. Zhang, and P. Bourne. 2013. Biodiversity data should be published, cited and peer-reviewed. *Trends Ecol. Evol.* 28: 454–461.
- Ehrlén, J., and W.F. Morris. 2015. Predicting changes in the distribution and abundance of species under environmental change. *Ecol. Lett.* 18:303-314.
- Faith D.P., B. Collen, A.H. Ariño, P.O. Koleff, J. Kerr, J. Guinotte, and V. Chavan. 2013. Bridging the data gaps: Recommendations of the GBIF Content Needs Assessment Task Group. *Biodiv. Inf.* 8:41-58.
- Gaiji, S., V. Chavan, A.H. Ariño, J. Otegui, D. Hobern, R. Sood, and E. Robles. 2013. Content assessment of the primary biodiversity data published through GBIF network: Status, challenges and potentials. *Biodiv. Inf.* 8:94-172.
- Guralnick, R., and A. Hill. 2015. Biodiversity informatics: Automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics* 25:421-428.
- Guralnick, R.P., J. Wieczorek, R. Beaman, and R.J. Hijmans. 2006. BioGeomancer: Automated georeferencing to map the world's biodiversity data. *PLoS Biology* 4:e381.
- Hill, A.W., J. Otegui, A.H. Ariño, and R. Guralnick. 2010. GBIF Position Paper on Future Directions and Recommendations for Enhancing Fitness-for-Use Across the GBIF Network, version 1.0. Global Biodiversity Information Facility, Copenhagen.
- Judd, W.S. 2008. *Plant Systematics: A Phylogenetic Approach*, 3rd Edition. Sinauer Associates, Sunderland, Mass.
- Kelling, S., W.M. Hochachka, D. Fink, M. Riedewald, R. Caruana, G. Ballard, and G. Hooker. 2009. Data intensive science: A new paradigm for biodiversity studies. *Bioscience* 59:613-620.
- MES. 2002. *National Biodiversity Strategy for Ghana*. Ministry of Environment and Science, Ghana.
- 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.
- Nelson, G., P. Sweeney, L.E. Wallace, R.K. Rabeler, D. Allar, H. Brown, J.R. Carter, M.W. Denslow, E.R. Ellwood, C.C. Germain-Aubrey, G. Ed, E. Gillespie, L.R. Goertzen, B. Legler, D.B. Marchant, T.D. Marsico, A.B. Morris, Z. Murrell, M. Nazaire, C. Neefus, S. Oberreiter, D. Paul, B.R. Ruhfel, T. Sasek, J. Shaw, P.S. Soltis, K. Watson, A. Weeks, and A.R. Mast. 2015. Digitization workflows for flat sheets and packets of plants, algae, and fungi. *Appl. Plant Sci.* 3:1500065.
- National Development Planning Commission (NDPC). 2014. *Ghana Shared Growth and Development Agenda (GSGDA) II, 2014-2017. Volume 1. Policy Framework*. 242.
- Norris, K., A. Asase, B. Collen, J. Gockowksi, J. Mason, B. Phalan, and A. Wade. 2010. Biodiversity in a forest-agriculture mosaic—the changing face of West African rainforests. *Biol. Conserv.* 143: 2341-2350.
- Pereira, H.M., S. Ferrier, M. Walters, G.N. Geller, R.H.G. Jongman, R.J. Scholes, M.W. Bruford, N. Brummitt, S.H.M. Butchart, A.C. Cardoso, N.C. Coops, E. Dulloo, D.P. Faith, J. Freyhof, R.D. Gregory, C. Heip, H.R.

⁴<http://www.gbif.org/resource/80528>.

- Hurttt, W. Jetz, D.S. Karp, M.A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J.R.W. Scharlemann, S.N. Stuart, E. Turak, M. Walpole, and M. Wegmann. 2013. Essential biodiversity variables. *Science* 339: 277–278.
- Peterson, A.T., A. Asase, D.L. Canhos, S. de Souza, and J. Wiecezorek. 2018. Data leakage and loss in biodiversity informatics. In review.
- Peterson, A.T., and K. Ingenloff. 2015. Biodiversity Informatics Training Curriculum, Version 1.2. *Biodiv. Inf.* 10:65-74
- Peterson, A.T., J. Soberón, R.G. Pearson, R.P. Anderson, E. Martínez-Meyer, M. Nakamura, and M.B. Araújo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Robertson, T., M. Döring, R. Guralnick, D. Bloom, J. Wiecezorek, K. Braak, J. Otegui, L. Russell, and P. Desmet. 2014. The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the Internet. *PLoS ONE* 9(8): e102623.
- 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. *Div. Distrib.* 20:369-381.
- Tulig, M., N. Tarnowsky, M. Bevans, A. Kirchgessner, and B.M. Thiers. 2012. Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209:103-113.
- Veiga A.K., A.M. Saraiva, A.D. Chapman, P.J. Morris, C. Gendreau, D. Schigel, and T.J. Robertson. 2017. A conceptual framework for quality assessment and management of biodiversity data. *PLoS ONE* 12: e0178731.
- Wiecezorek, J., D. Bloom, R. Guralnick, S. Blum, M. Doring, R. Giovanni, T. Robertson, and D. Vieglais. 2012. Darwin Core: An evolving community-developed biodiversity data standard. *PLoS ONE* 7:e29715.

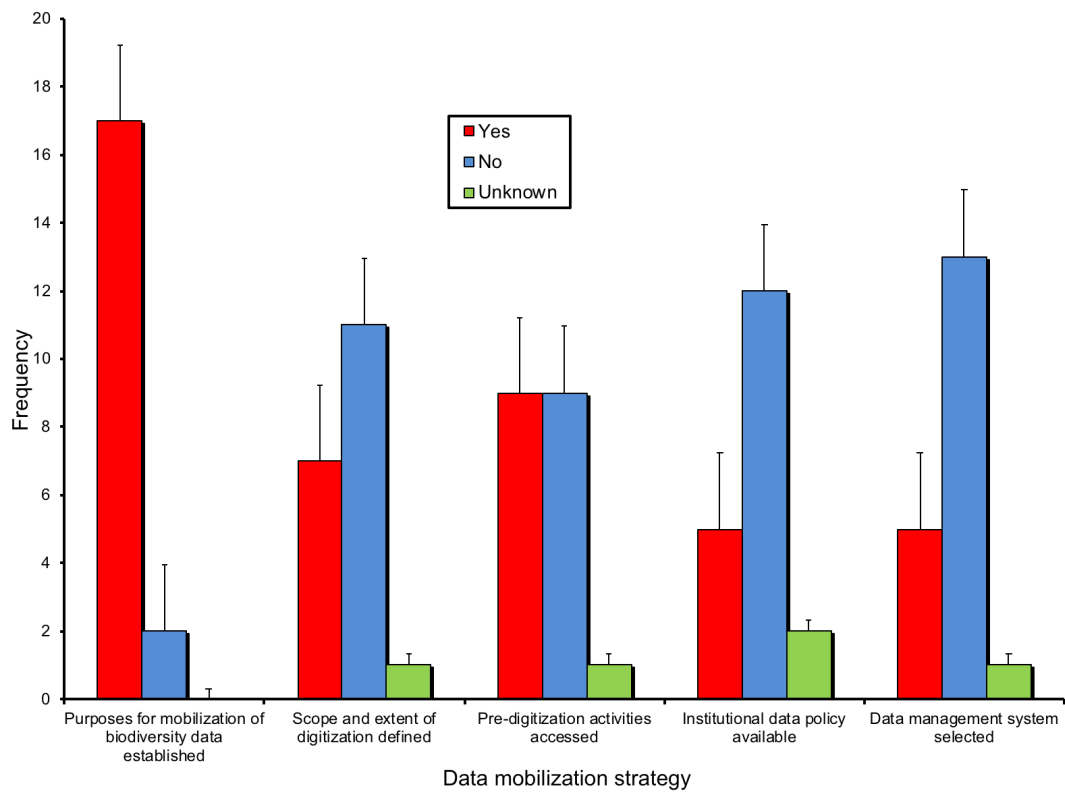


Figure 1: Status of biodiversity data mobilization strategies of data holders in Ghana. Error bars \pm standard error.

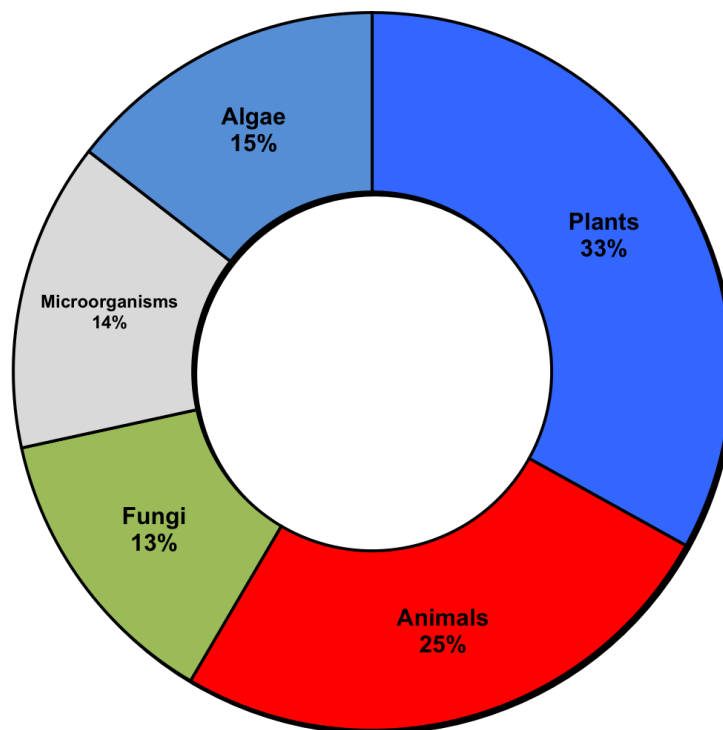


Figure 2: Biodiversity data needs in Ghana according to major taxonomic groups.

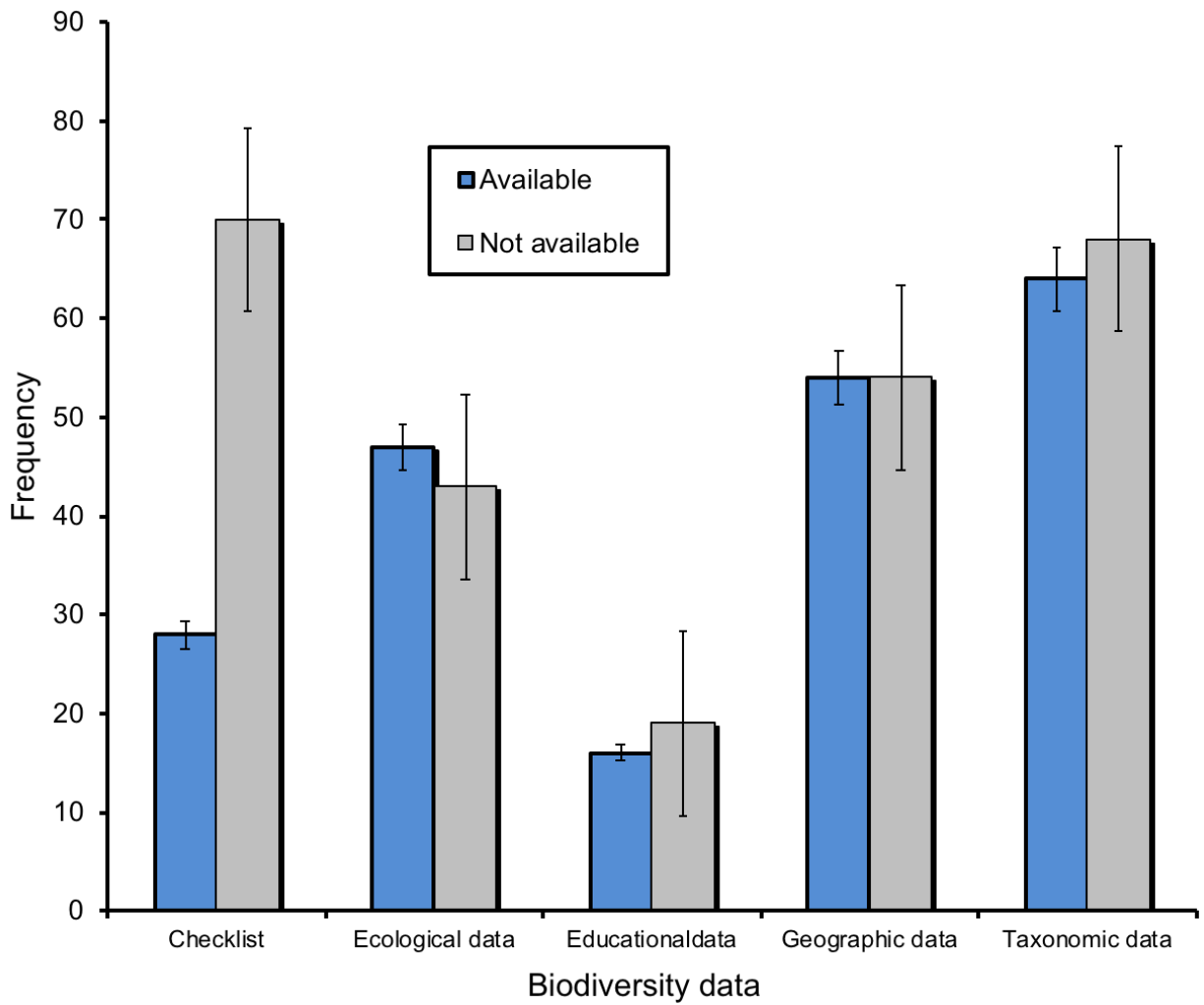


Figure 3: Survey responses regarding availability of biodiversity data in Ghana. Error bars \pm standard error.

Table 1: Summary of the biodiversity data digitization landscape in Ghana.

Items	Responses		
	Yes	No	Unknown
Digitization workspace available	5	13	1
Proportion of collections digitized	10	9	0
Pre-curation digitization carried out	7	11	1
Digitization processes and technologies defined	5	14	0
Digitization workflows defined	8	9	2
Staff adequately trained and equipped	4	14	1
Quality control and standards specified	7	11	1
Data platform for digitization selected	14	2	3
Format digital data stored selected	8	3	8

Table 2: Biodiversity data needs according to data types and major biodiversity groups.

Biodiversity data	Biodiversity groups / frequencies					
	Plants	Animals	Fungi	Microorganisms	Algae	Total
Checklist	84	65	28	36	37	250
Ecological data	72	59	36	36	31	234
Educational data	39	27	14	5	9	94
Geographic data	98	79	36	41	48	302
Taxonomic data	120	87	50	56	56	369
Total	413	317	164	174	181	1249

Table 3: SWOT analysis of biodiversity data needs landscape for Ghana.

<p>Strengths (internal)</p> <ol style="list-style-type: none"> Existence of a national legal framework pertaining to data use and management. High technical expertise on biodiversity science. Rich natural history collections on plants, mammals and insects. Both public and private organizations have data on Ghanaian biodiversity. Willingness of stakeholders to be trained in biodiversity informatics. Active participation of stakeholders in national biodiversity activities. Many biodiversity research programmes and projects across the country. Possibility to upgrading the National Biodiversity Committee into a Biodiversity Commission. 	<p>Weaknesses (internal)</p> <ol style="list-style-type: none"> Lack of motivation for biodiversity data-holders to share data. Poor infrastructural and human capacities in biodiversity informatics. Lack / inadequate funds to support biodiversity data mobilization. Weak institutional collaborations about biodiversity. Lack of a national sustainable and coordinating biodiversity information hub.
<p>Opportunities (external)</p> <ol style="list-style-type: none"> Biodiversity data about Ghana is available online through outlets such as GBIF etc. Repatriation of biodiversity data associated with collections of natural history museums in Europe and North America. Training opportunities in biodiversity informatics (e.g. BITC programme, GBIF, TDWG) Ghana is a signatory to international biodiversity conventions such as Convention on Biological Diversity, GBIF and IPBES. Availability of external funds for biodiversity science (e.g. UNE-WCMC Connect project, JRS Biodiversity Foundation, Biodiversity for Development (GBIF-BID) programme). Networking and collaborations with external partners. 	<p>Threats (external)</p> <ol style="list-style-type: none"> Non-cooperative biodiversity data-holders. Lack of funds to capture data in European / North America institutions.

Table 4: Biodiversity data needs or gaps, and challenges according to three broad user-groups in Ghana.

Biodiversity data user group	Data needs / gaps	Challenges
Academia and research	Plant distribution and phenology; aquatic biodiversity; data on pathogens and disease-causing organisms; medicinal plants; and nomenclatural changes.	Lack of motivation to share data, lack of taxonomists, minimal capacity for collection and curation of biodiversity, low expertise in biodiversity informatics, and lack of a sustainable biodiversity data-coordinating unit.
Governance and policy	Savanna biodiversity, agro-biodiversity, marine biodiversity, lower-taxon biodiversity (fungi, algae etc.), and invasive species.	Lack of motivation to share data, lack of taxonomists, inadequate financial resources, absence of policy on biodiversity data, and lack of law enforcement on data.
Non-governmental organizations and others	Herpetofauna, Red List of biodiversity in Ghana, species distribution maps, harvested biodiversity, and effects of climate and land use change on biodiversity	Lack of sustainable central biodiversity coordinating unit, and little collaboration between biodiversity data-holders.