

BRIDGING BIODIVERSITY DATA GAPS: RECOMMENDATIONS TO MEET USERS' DATA NEEDS

DANIEL P. FAITH (1)*, BEN COLLEN (2), ARTURO H. ARIÑO (3), PATRICIA KOLEFF (4), JOHN GUINOTTE (5), JEREMY KERR (6) AND VISHWAS CHAVAN (7)

(1) *Australian Museum, 6 College Street, Sydney, NSW 2010, Australia.*

(2) *Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, United Kingdom*

(3) *Department of Zoology and Ecology, University of Navarra, 31080 Pamplona, Spain.*

(4) *National Commission for the Knowledge and Use of Biodiversity (CONABIO), Mexico*

(5) *Marine Conservation Biology Institute, 2122 112th Ave NE, Suite B-300, Bellevue WA 98004, USA.*

(6) *University of Ottawa, Canada*

(7) *Global Biodiversity Information Facility Secretariat, Universitetsparken 15, DK 2100, Copenhagen, Denmark*

*corresponding author

Abstract – Freely available high quality, data on species occurrence and associated variables are needed in order to track changes in biodiversity. One of the main issues surrounding the provision of such data is that sources vary in quality, scope, and accuracy. Publishers of such data must face the challenge of maximizing quality, utility and breadth of data coverage, in order to make such data useful to users. With the Global Biodiversity Information Facility (GBIF), we recently conducted a content needs assessment survey to consolidate and synthesize major user needs regarding biodiversity data. We find a broad range of recommendations from the survey respondents, principally concerning issues such as data quality, bias, and coverage, and ease of access. We recommend a candidate set of actions for the GBIF that fall into three classes: 1) addressing data gaps, data volume, and data quality, 2) aggregating data types that are relatively new to GBIF, to support emerging new applications, and 3) promoting ease-of-use and providing incentives for wider use. Addressing the challenge of providing high quality primary biodiversity data potentially can serve the needs of national and international biodiversity initiatives. These include the “flexible framework” for addressing the new 2020 biodiversity targets of the Convention on Biological Diversity, the global biodiversity observation network (GEO BON) and the new Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). Each of these presents opportunities for countries to define appropriate actions and corresponding data needs, with links from local to global scales.

INTRODUCTION

Biodiversity is the variety of life, extending from the level of genes, to species, to ecosystems (Norse et al. 1986; Wilson 1988; Wilson 1999; Margalef 2000). The extent of the biodiversity crisis – the potential loss of much of this living variation – was highlighted again recently by the Third Global Biodiversity Outlook (Secretariat of the Convention on Biological Diversity 2010). It documented a decline in the measures of biodiversity, and an increase over the past 30 years in the pressures that cause biodiversity loss (Butchart et al. 2010). In 2012, the United Nations

Conference on Sustainable Development (UNCSD or “Rio+20”) marked 20 years since the original conference that gave birth to the Convention on Biological Diversity. The major outcome document from the UNCSD conference (UNCSD 2012) reinforced the critical links between biodiversity conservation and sustainability. The recent call by scientists for new efforts to “map the biosphere” (Wheeler et al. 2012) also highlighted the importance of biodiversity data for addressing many basic research question in biodiversity science. Effective strategies for addressing biodiversity loss, and progressing biodiversity

science, require that biodiversity-relevant information is made available in a useful form. The range of existing available data must be extended by strategically filling biodiversity knowledge gaps (Collen et al. 2008). Further, biodiversity data (data relevant to variation at the genes, species, and ecosystems levels) will be of limited use on their own. Such data must be integrated at various spatial scales, not only with other environmental data, but also with a wide variety of types of socio-economic data, in order to address the most pressing questions in biodiversity and sustainability science. All of these considerations highlight the fact that useful biodiversity databases must go beyond lists of species, and provide integrated data of a quality required for a range of research and applications.

Biodiversity data typically may have been generated for one specific intention, but then used subsequently for other, multiple, purposes (Chapman 2005). Biological specimen data originally collected by museums and herbaria have contributed to spatial models describing broader biodiversity patterns, to species richness estimation, and to ecological/ environmental studies on attributes of species, including population size, geographic distribution, habitat and behaviour (Pyke and Ehrlich 2010; Gropp 2012). Increasingly, such data have helped to assess the impacts of threats to species, including pollution, disease, and climate change. If we consider one of the most prominent of those challenges, climate change, it is clear that assessment of climate change impacts is limited by lack of species-occurrence data, including limitations in mechanisms for data discovery and access (Tews 2006; Chavan and Ingwersen 2009). The shortfall in available data highlights the ongoing need for a mechanism to facilitate sharing of existing and future biodiversity data, both within and among countries (Gaikwad and Chavan 2006; Chavan and Ingwersen 2009).

The Global Biodiversity Information Facility (GBIF), an inter-governmental organization, facilitates a discovery and publishing mechanism for free and open exchange/sharing of existing (and future) biodiversity data, and is the predominant international, publically funded resource for species occurrence data. Through GBIF, institutions and countries are discovering

and publishing their data resources online with common exchange standards, and are part of a growing global network of shared biodiversity data. More than 388 million primary biodiversity data records are accessible (as of October 2012) through the GBIF global networks data portal (<http://data.gbif.org/>).

There are many studies highlighting the variety of uses for shared biodiversity data (e.g. Chapman 2005; including examples of its potential for new scientific research and for decision making related to natural resources management). However, deficiencies in species occurrence data coverage also have been highlighted (e.g., Yesson et al. 2007; Collen et al. 2008; GBIF 2010a; Boakes et al. 2010; Gaiji et al. 2013, this volume). Deficiencies in coverage and data quality (Bortolus 2008) reduce the magnitude and utility of available data (Escobar et al. 2009).

Data concerns raised in the routine feedback from GBIF users have included: lack of sufficient data (volume, depth, and density), lack of fitness-for-use (precision, accuracy and authenticity), and lack of mechanisms for data discovery, publishing, and accessing data (Chavan and Ingwersen 2009; Ariño et al. 2012; GBIF 2010c). A common concern is that applications may not provide defensible conclusions because of the largely opportunistic way in which existing biodiversity data have been collated, and published (Pyke and Ehrlich 2010).

Content Needs Assessment

To address these issues in a holistic manner, GBIF established the Content Needs Assessment Task Group (CNA TG) composed of the authors of this paper (co-chairs: Faith & Collen; GBIF senior programme officer: Chavan) with a remit to identify major areas of opportunity to mobilize data in a way that better considers users' needs (GBIF 2009a). The objective of CNA TG therefore was to investigate user needs regarding biodiversity data within the broad arena of biodiversity research. The CNA TG was mandated also to provide recommendations to GBIF that would (a) determine the priority questions that GBIF mobilised data should be able to address for various areas of science and policy in the near, medium, and long-term, from local to global level including thematic areas, (b) for the identified

priority questions, evaluate the content needs (volume, depth, and density) and data fitness-for-use (precision, accuracy, authenticity) for specific uses, (c) assess what unique scientific and policy contributions of GBIF mobilised data that cannot be met easily through other mechanisms, (d) identify gaps in accessible data mapped against data needs, and (e) recommend strategies and priorities for data discovery and publishing through GBIF network. In this paper, we draw on the work of CNA TG to propose some guidelines to GBIF for addressing user needs. However, given the broad importance of primary biodiversity data to biodiversity science, our findings may be more broadly applicable.

The sections below present and discuss a candidate set of recommendations resulting from the CNA TG analysis and discussion of the results of an extensive survey among biodiversity information stakeholders. The many opinions and comments provided through the survey have resulted in a variety of detailed suggestions, and we have refined these to produce recommendations falling under several different themes or contexts. As a complement to this list of recommendations, we also have drawn upon the current relevant scientific literature, including that relating to international and national biodiversity initiatives, in order to provide additional broad context for these recommendations (see also GBIF 2011).

METHODS

The survey (see Ariño et al. 2013, this volume for full details), was launched in 2009 in six languages (English, Spanish, French, Chinese, and Russian) (GBIF 2009b; GBIF 2009c; and GBIF 2009d). The survey consisted of 21 questions covering (a) respondent profile, (b) uses of primary biodiversity data, (c) access to primary biodiversity data, (d) data quality and quantity requirements, (e) species level data requirements, and (f) usefulness of GBIF mobilised data.

SurveyMonkey (<http://www.surveymonkey.com>) was used to administer the survey and retrieve responses. The survey was widely circulated using biodiversity-related lists and portals. Resulting raw answers were retrieved from *SurveyMonkey*, demographic information was collected, and individual responses anonymised before proceeding to analysis. The raw files were

converted into a purpose-made database (Ariño et al., this volume) amenable for deeper analysis and cross-referencing. Responses sent in languages other than English were translated, and all answers coded homogeneously while retaining the language demographics. The coded database was cross-tabulated and analyzed variously for frequencies, correlations, and statistically summarized over several dimensions (Ariño et al., this volume) to produce frequency tables, plots, maps and summaries that helped discover trends and steer our discussions.

Responses to the CNA TG survey formed the basis for deliberations by the CNA TG and consultation with colleagues from the GBIF and other related communities. This process resulted in 14 main recommendations.

RESULTS

The survey resulted in responses from more than 700 individuals, providing more than 48,000 individual answers and nearly four-thousand individual verbatim comments. Our analysis of the responses revealed a vast array of uses of primary biodiversity data, underlining the importance of its provision. Importantly, the survey highlighted some new types of data in high demand, particularly data used to add value to the geographic distribution and taxonomic data already mobilized by GBIF (Ariño et al. 2013, this volume). At the same time, the survey highlighted a common lack of awareness about the extent of availability of accessible primary data. Of great concern were the responses that identified unreliable or incomplete data as the reason for not using data portals such as GBIF.

RECOMMENDATIONS

We note that these recommendations do not nominate any specific actor, as many of these recommendations require collective, simultaneous, and/or coordinated actions by multiple actors. We discuss the recommendations under three sections below. First, we examine important issues regarding data gaps, data volume, and data quality. In the second section, we focus on the visions for new kinds of data, and new potential applications. We extend that section by considering some of the related national and global challenges raised by international biodiversity initiatives, including the new

biodiversity targets (<http://www.cbd.int/sp/targets>) of the Convention on Biological Diversity (UNEP 2010a), the global biodiversity observation network (GEO BON) and the new Inter-governmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). Finally, in the third section we consider issues and recommendations that relate to strategies for improving ease of use of primary biodiversity data, over the wide range of potential applications, and with particular reference to the data mobilization by the GBIF network.

Data gaps, data volume, and data quality

The CNA TG survey revealed that a major concern for biodiversity data users is the quality and coverage of the available data (Ariño et al. 2013, this volume). Highest among their concerns were geographic and taxonomic gaps in data coverage, and the need for data quality assurance. Given the existing trends of publishing data predominantly by the data rich countries as opposed to biodiversity rich countries, some biases

towards particular taxa, places or periods of time, are inevitable (Yesson et al. 2007; Robertson 2008). Our recommendations focus on ways to address bias by identifying and filling data gaps. The problem of data quality raises a number of issues. It is critically important that sources of error can be accounted for. The recommendations made under this section (Table 1) suggest strategies to address these issues, in order to ensure that GBIF mediated data are of the highest possible quality. The recommendations relating to data quality issues in particular, would help deliver an improved process for user feedback and reporting of errors (including the ability to pin-point malicious records). A key outcome from these recommendations would be increased credibility and increased confidence in use of GBIF mobilised data. Further, data producers and primary publishers will be in better control of assessment and improving ‘fitness-for-use’ of GBIF mobilised data prior to its discovery and publishing over the network.

Table 1. Recommendations relating to data gaps, data volume, and data quality

1. Recognise that geographic, temporal, taxonomic, and ecosystem gaps exist in currently mobilised data, develop tools to spot and identify biases at multiple geographic scales, and expedite efforts to bridge these gaps from local to global scale.
2. Facilitate ways to overcome the inevitable biased nature of the data discovered and mobilised through the GBIF network, promoting more uniform spread of primary biodiversity data (geographic, taxonomic, and temporal).
 - a. Develop best practice guidelines to help Participants and potential Data Publishers to prioritise data discovery, capture, digitization, and publishing on a demand-driven basis, so that such proposals make business sense for donor agencies.
 - b. Draw ‘local-to-global’ scale strategy and action plans for discovery and publishing of primary biodiversity data.
3. Encourage and increase investment in retrospective discovery, digitization and publishing of historical and time series datasets.
 - a. Expedite digitization of natural history collections data, especially for type specimens.
4. Ensure increased access to authoritative taxonomic catalogues and expedite progress towards a ‘Global Names Architecture’.
5. Add and enhance features in the GBIF data portal, promoting improved ‘fitness-for-use’ of data discovered and accessed through the portal.

6. Initiate the following steps to enhance the trust-worthiness of GBIF mobilised data:

- a. Expedite the tagging of data to highlight possible errors, data quality, and uncertainty levels.
- b. Develop standards for annotations and data user feedbacks.
- c. Strengthen data quality assessment and quality enhancement mechanisms.
- d. Develop distributed, decentralized, yet coordinated (inter-connected) annotation and feedback mechanisms.
- e. Develop a user-friendly standard vocabulary for types of errors and types of quality issues.
- f. Develop mechanisms for users to document the ‘use confidence rating’ at data set and data record level.
- g. Improve pathways for data publishers to provide warnings about biases or errors in the data at an early stage of discovery and publishing process.
- h. Make provision for inclusion of data problems (lack of fitness-for-use) relevant annotations or feedbacks together with data, not only at dataset level but at the data record level.
- i. Develop contributor specific best practices and/or mechanisms for quality assurances and quality control.
- j. Expedite efforts in improving taxonomic and geo-spatial quality of GBIF mobilised data. This task includes attention to geo-referencing.
- k. Improve fitness-for-use of data at the data producer and/or primary publisher stage.
- l. Establish links to the original datasets to enhance the verifiability of the data.

There are approximately 388 million primary biodiversity data records published through the GBIF network (as on October 2012). Among the key recommendations in this section are those calling for efforts to bridge gaps (taxonomic and geographic, from local to global scales) in the coverage of these records, and to overcome the inevitable biased nature of the data discovered and mobilised through the network. The type of information collated by GBIF is being used in many applications, including models of species diversity status and change and tracking progress in conserving biodiversity (Boakes et al. 2010). One of the main advantages of species occurrence data is that it has the potential to give more than a simple snap shot of biodiversity status and distribution (e.g. Mace et al. 2010). A sampling that is representative and balanced will greatly promote such analyses. Therefore strategies to create a balanced spread of data (geographically, taxonomically, and temporally) are essential to facilitate meaningful analysis and interpretation of biodiversity as a whole, as opposed to “sectorial”

biodiversity (for example, biodiversity of birds or biodiversity in Fennoscandia). GBIF mobilised data are currently dominated by terrestrial ecosystems and higher animals (GBIF 2010a; Gaiji et al. 2013, this volume). A balanced spread of records across ecosystems and taxa provides a stronger basis for broad generalisations about biodiversity patterns that are not limited to narrow taxon groups, or highly-sampled geographical areas. The patterns of use revealed by the survey, spanning monocots to algae, and lower animal phyla, reflect the lack of data on such taxa. GBIF should investigate taxonomic, geospatial, and temporal biases and errors in data at the level of individual records (rather than just at the higher data-set level).

For many biodiversity studies, the principal data-coverage limitations include the ‘Linnaean shortfall’ (we only know a fraction of the planet’s species) and the ‘Wallacean shortfall’ (even for described species, we know little about geographic distributions - for a review, see Brito 2010). For

example, in Canada, perhaps 2/3 of all species are known to science, and in Australia, perhaps only about 1/3 of all species are known. The fact that only a small fraction of known species have mobilised, high quality data, covering their full geographic distribution makes the task of biodiversity assessments even more daunting (Collen et al. 2009). Large geographic and taxonomic biases in biodiversity data from historical inventories have been widely documented (Hortal et al. 2008). Boakes et al. (2010) observed that the details of sampling biases (or validation) may be difficult to find in databases, and concluded that *“compensating for these biases will be important for any study aiming to draw conclusions about real trends in biodiversity over time and space. Accounting for biases in biodiversity samples depends on a clear knowledge of the source and nature of those biases.”* Boakes et al. (2010) investigated patterns of spatial and temporal bias using a database covering more than 200 years, for species in the avian order Galliformes. Among their various sources of species distribution data (museum collections, scientific literature, ringing records, ornithological atlases, and website reports from “citizen scientists”), museum data provided the best historical coverage of species' ranges. On the other hand, they concluded that such data were time-intensive to collect.

One of the quickest ways to enhance data coverage would be to mobilise additional natural history collections. Type specimens are valuable data resources that provide baselines for comparison, and historical context that is often missing from biodiversity trend data (Lotze & Worm 2009). However, historical context requires capacity to carefully re-examine “old” data. Although there is no readily available mechanism to back track to original datasets which form the index of the GBIF mediated data, and some of these resources have been either moved away from their original locations or are no longer online (and some may no longer exist), the GBIF index itself could eventually be used as a historical repository of published dynamic data (i.e., data published and thus usable within the time frame when they were on line) through its periodical snapshots. A persistent ID attributed to each specimen would greatly facilitate this task, and alleviate the limit to

the ability of users to verify the authenticity or validity of the resource.

One of the principal advantages of using occurrence data to evaluate changing biodiversity is that it can potentially provide time series information (e.g. Mace et al. 2010). Time series perhaps represents a more challenging gap than taxonomic or geographic imbalances in data. Boakes et al. (2010) argued that an historical context to current changes can be set into context by long-term trends that reveal major shifts in abundance and composition of biological communities. Unfortunately, many datasets do not seem to survive the typical 2-5 years worth of data support linked to funding cycles or graduate student tenure (Likens 1989), and if they do, another limit seems to be reached when the research career of particularly dedicated individuals end (Warner et al. 1995; Ariño and Pimm 1995). If data such as the type mediated by GBIF are to be used to evaluate biodiversity change as a whole, as well as securing historical data for the future and making it available for users, sampling biases need to be understood and addressed (Boakes et al. 2010).

Wheeler et al. (2004) argued that *“Some naively see the information technology challenge as liberating data from cabinets. The reality is that for all but a few taxa, much data is outdated or unreliable. Many specimens represent undescribed or misidentified species. Rapid access to bad data is unacceptable; the challenge is not merely to speed data access but to expedite taxonomic research.”* These arguments justify the dual focus, in the recommendations listed above, on filling taxonomic (and other) gaps, and, at the same time, reducing data errors.

Users of GBIF mobilised data can take advantage of several developments that help overcome limitations imposed by gaps and errors, including the development of robust biodiversity modelling approaches (e.g. providing surrogates for overall biodiversity patterns; e.g., Ferrier 2002), and related ‘gap’ analyses (GBIF 2010a; Gaiji et al. 2013, this volume; Otegui et al. 2013, this volume) which promote strategic growth in mobilised biodiversity data (both through new surveys and priorities for digitising collections data) (Berendsohn et al. 2010; Berents et al. 2010).

There is increasing support (e.g. see Juutinen et al. 2008) for a surrogates framework which uses models to combine environmental and primary biotic data, in order to provide best-possible use of available data to serve urgent needs of decision-making (e.g. Faith and Walker 1996). The approach uses primary species data as a starting point, but integrates this with environmental data to make better inferences about overall biodiversity (Faith et al. 2004).

An example of this perspective is found in Linke et al. (2011):

“With increasing availability of both geographic information system (GIS) data and new, user-friendly modelling techniques, it is rapidly becoming easier to produce modelled species surrogates or highly informed physical surrogates. Moreover, with the availability of predictive modelling techniques that are robust to data-poor inputs, the use of environmental rather than species surrogates is only necessary where species data are extremely limited and the area very large, for example in the Amazon.... A key advantage of systematic approaches is that they make the best use of existing data and can be applied where data are limited to generate reliable but coarse assessments.”

The surrogates approach is able to make effective use of even the small amounts of data available for some taxonomic groups, because the model depends only on the combination of species (across all groups) to produce an estimated pattern for overall biodiversity. This strategy greatly increases the information content for a broad range of GBIF data. Further, because the robust models integrate widely available environmental variables, the models based on the GBIF mobilised data can be applied in a region that has little GBIF mobilised data of its own.

Biodiversity surrogates strategies might suggest that limited existing data are all that are needed. However, better quality data will generally produce more accurate surrogate models. Importantly, initial surrogate models can be used to strategically fill in taxonomic and geographic gaps. One method, “Survey Gap Analysis” (see Ferrier 2002) indicates places most in need of new data

collection or integration into GBIF – so promoting the mobilisation of new data. Similarly, Brito (2010) suggests that directing surveys towards areas of known data deficiency will likely result in the discovery of species new to science, helping to address the Linnean shortfall (see also, Raxworthy et al. 2003). Another strategy suggests that priorities for data mobilisation should focus in part of how best to improve current surrogates models (Faith 2005). Of course, the potential gains from new data and knowledge are further enhanced when the possibility exists that the new knowledge will provide many future applications, not just one. These considerations all highlight issues linked to our recommendations to fill data gaps and improve data quality, even while using robust methods to make best-possible use of the currently available data.

These arguments again highlight the need to analyze errors and biases in existing data. The approaches suggested by Boakes et al. (2010) for coping with various biases (including differences in taxonomic efforts, differences in sampling efforts in different localities) are concordant with our recommendations for the GBIF network (see also Hortal et al. 2006). The comments also strongly support a strategy in which GBIF increases its’ capacity to discover and publish data from a wide range of sources, including extended data types (including historical or time series).

We note that uncertainties are relevant to other data types Parr et al. (2011) in reviewing “evolutionary informatics” highlight challenges for phylogenetic information including the adequate communication of uncertainties: “accommodating topological disagreement where necessary, would consolidate taxon names, phenotypic and geographical distributional data across clades.”

Extending data types and applications

Our CNA TG Survey raised issues relating to the needs for “new” kinds of data serving a broader range of applications (Ariño et al. 2013, this volume). These needs may be addressed here by four recommendations (Table 2). Again, a number of the recommendations have subsidiary recommendations (lettered entries).

Table 2. Recommendations relating to extended data types and applications

7. Develop a biodiversity landscape map depicting GBIF's position (including its Participants, and Data Publishers), role, unique advantages and collaborative strategies amidst the myriad biodiversity, biodiversity informatics initiatives of local to global scale.
 - a. Establish or initiate collaborations with initiatives, networks, and organizations outside of the sphere of biodiversity.
8. Develop initiatives for dealing with new or under-appreciated kinds of data.
 - a. Expand discovery and access to new data sources, types and themes and types (e.g. observations including absence only records, population richness data, and other data associated with different types such as multimedia objects and environmental impact assessment etc).
 - b. Focus attention on mobilising data that will be useful for multilateral environmental agreements, mainly for those closely related to biodiversity issues such as CBD and CITES, etc.
 - c. Develop mechanisms for linkages with other data types and data resources.
 - d. Provide functionalities for export, linkages and retrieval of data associated with GIS shape files and/or polygons.
9. Develop initiatives for enhancing applications, demonstrating that existence of GBIF mobilised data indeed makes a difference in biodiversity conservation and sustainable use.
 - a. Demonstrate use of GBIF mobilised data for applied environmental sciences as well as socially relevant issues, rather than merely for basic scientific purposes.
 - b. Promote publication of data addressing a wide spectrum of applications or usages of biodiversity data.
10. Conduct a 'content needs assessment' (local, regional, thematic, and global scale) at frequent intervals.
 - a. GBIF Participants could conduct multilingual content needs assessment exercise at regular intervals.
 - b. Ensure improved coordination while conducting content needs assessment exercise.
 - c. Ensure increased participation of stakeholders, policy makers, administrators, natural resources managers, representatives of civil society and non-governmental organizations into content need assessment activities.

GBIF is regarded as having an important role in enabling a 'research infrastructure' based on the discovery and publishing of the world's primary biodiversity data (Bridgewater et al. 2010; Peterson et al. 2010). This role can create complementary linkages with initiatives within and outside of the sphere of biodiversity. These linkages will work particularly well when there is seamless access to both biodiversity and non-biodiversity data resources. This calls for extended

data types from disciplines such as genetics, ecology, fisheries, agro-biodiversity, and environmental impact assessment. These will enhance the potential use of GBIF mobilized data and will demonstrate scientific, ecological, social and economic relevance of GBIF network (Sutherland et al. 2009).

Ariño et al. (2013, this volume) noted several key data needs suggested by the Survey. For example, genetics/genomics data are needed for

integrated biodiversity studies, suggesting benefits will be gained from GBIF links to existing molecular sequence databases, including GenBank and Barcode of Life. Page (2008) notes that genomics and phylogenetics research call for identifiers, such as specimen codes and GenBank accession numbers, as companions to conventional taxonomic names. These genomics and phylogenetics databases in turn call for integration with other auxiliary data that supports applications. The importance of such variables is highlighted in “Genomic Observatories” (GOs), which focus sequencing efforts in locations that are already well-studied and rich in associated data (e.g. data on environmental variables) (Davies et al. 2012).

A wide range of environmental and ecological data will complement standard biodiversity data in multi-disciplinary biodiversity science. Mace and Baillie (2010), for example, argued that “available data also tend to emphasize taxonomic diversity, rather than ecosystem functions and services”. These arguments are reinforced by survey respondents’ call for mobilised data relating to species’ trait data and corresponding functions for species. Biodiversity studies that integrate species data, genomics, functional traits, and other data will be supported by emerging tools for research and policy that can incorporate all these kinds of data (e.g. Faith et al. 2009). Equally, greater integration of different data types can be accompanied by a shared analytical toolbox. For example, the Barcode of Life Data Portal (Sarkar and Trizna 2011) currently makes available a range of tools specifically for analysis of DNA barcoding data, but also could take advantage of tools commonly used for other data types.

Arguments for expanded biodiversity-related data sometimes are reflected in an expanded definition of “biodiversity”, beyond its core interpretation as “living variation”. For example, Diaz et al. (2009) consider biodiversity as “*the number, abundance, composition, spatial distribution, and interactions of genotypes, populations, species, functional types and traits, and landscape units in a given system*”. Each of these aspects defines potential new data types that need to be considered in an integrated way by the GBIF network and its partners (see also Hardisty et al. 2013). To expedite such integration, GBIF might link to progress already made in digitizing

some of these types of biodiversity data. Other useful progress is found in compiled databases such as that for phylogenetic relationships (see Page 2008).

Important new data links of this kind include, for example, observations data from the Global Mountain Biodiversity Assessment (GMBA), a cross-cutting network of DIVERSITAS. GMBA has produced a metadata catalogue (<http://gmba.unibas.ch/index/index.htm?output=printable>) linked to GBIF, that describes “*the "who, what, where, when, why and how" pertaining to the collection of a given ecological dataset on mountain biodiversity*”.

Another example is found in the extensive information contained in existing atlas data bases, which provide collections of spatially explicit data on species occurrences. Robertson et al. (2010) advocate expanded databases of this kind: “*The most useful atlases have a good measure of sampling effort; include data collected at a fine enough resolution to link to habitat variables of potential interest; have a sufficiently large sample size to work with in a multivariate context; and offer clear, quantitative indications of the quality of each record to allow for the needs of users who have specific demands for high-quality data.*” When these requirements are met, ecological biogeography and other inferences can be drawn, as demonstrated by exemplary cases such as Escala et al. (1997).

GBIF users would benefit from links to the rapidly expanding information on ecosystem services (any processes or benefits derived from ecosystems; Tallis and Polasky 2009). While conservation of ecosystem services does not necessarily provide support for conservation of biodiversity, it does represent a key way to reduce the opportunity costs of retaining intact localities (Faith 2011). The need to integrate ecosystem services into regional-scale biodiversity assessments is a prime example of the gains to be made by integrated data. Indeed, this fundamental requirement for effective trade-offs and synergies among biodiversity conservation and other needs of society points to the need for integration of a wide variety of different data types.

“Unaccounted-for ecosystem services” have been recognised as a major information challenge

(Tallis and Polasky 2009; Eigenbrod et al. 2010). At the same time, biodiversity data gaps must be addressed in order to properly find synergies and trade-offs among these different aspects of human well-being. The biodiversity measures used in many recent trade-offs studies do not yet combine ecosystem services information with the available primary biotic data that would produce effective surrogates or indicators for overall (“wholesale”) biodiversity patterns (Faith et al. 2010). The Millennium Ecosystem Assessment (2005) called for “*a ‘calculus’ of global and regional biodiversity. This would allow global biodiversity gains and losses to be quantified in a unified way over various response strategies, thereby clearly identifying the trade-offs involved at a regional level. Such a calculus of biodiversity depends on effective biodiversity surrogates. These will be based upon the best possible use of a combination of environmental and species (for example, museum collections) data.*” A calculus of biodiversity therefore must use surrogates information to provide estimates of useful measures relating to biodiversity change, in order for biodiversity to effectively be “on the table” for decision-makers (for examples, see Faith et al. 2008). International initiatives (discussed below) make the comprehensive use of mobilised primary biodiversity data even more urgent.

National and international biodiversity initiatives

Our recommendations covering “extending data types, applications” can support links between GBIF and important national and international biodiversity initiatives. Because these initiatives require integration of biodiversity conservation goals and other factors (ecosystem services, socio-economic costs, etc.), they not only reinforce our recommendations about “making a difference” in biodiversity conservation, but also link to the need to consider “new” types of data and applications. Below, we particularly highlight some of the linkages that build on the capacity for GBIF mobilised data to develop a calculus of biodiversity.

1. *GEO BON*

The Group on Earth Observations (GEO; www.earthobservations.org) and its Global Earth Observing System of Systems (GEOSS) seek to improve the coordination of new and existing “Earth observation” data sets. One of the GEOSS systems (addressing to one of GEO’s nine designated societal benefit areas) is the new Biodiversity Observation Network (GEO BON; Andrefouet et al. 2008). GEO BON is to provide capabilities for sharing observations regarding biodiversity, from regional to global scales.

GBIF is already established as a key partner in pursuing these goals (Andrefouet et al. 2008). In accord with our recommendations, future efforts could pursue links between existing GBIF mobilised data and the other data types that are required within GEO BON (including remote sensing, ecosystem services, vegetation maps, and genomics data). Case studies, and early GEO BON products, could highlight the value added by such integration of GBIF mobilised data with other information. For example, GBIF mobilised data on species distributions can be used to add value to existing vegetation/ecosystem classifications (by estimating biotic overlaps among types), so providing improved surrogates (proxy) information for global biodiversity patterns (Faith and Walker 1996). One important new context is the Red List for Ecosystems (Rodríguez et al. 2012), which may use GBIF mobilised data to interpret these assessments at the species level (Faith 2012). GEO BON may present many opportunities for such value-adding using GBIF mobilised data (see also Couvet et al. 2012).

Biodiversity surrogates models based on combining available environmental data and GBIF mobilised species data (Faith and Walker 1996; Ferrier 2002; Soberón and Peterson 2009) may provide “essential biodiversity variables” (Pereira et al. 2013) describing patterns of “overall” or wholesale biodiversity. These models can support a core GEO BON strategy in which the interpretation of remotely-sensed changes in the extent and condition of land (or water) localities are interpreted through the “lens” of these inferred spatial patterns in biodiversity (Andrefouet et al. 2008; Faith et al. 2009). Soberón and Peterson (2009) propose a simple variant of the lens

approach “based on the increased availability of raw data about occurrences of species, cutting-edge modelling techniques for estimating distributional areas, and land-use information based on remotely sensed data to allow estimation of rates of range loss for species affected by land-use conversion.”

An important point is that the developing toolbox for biodiversity models and inferences that make use of GBIF mobilised species data opens the door to integration with other kinds of data. For example, existing biodiversity surrogates modelling approaches may be applied also to genomics data (Faith et al. 2009). Such a common modelling framework highlights the fact that these previously separate data types need to be linked, supporting our recommendations related to new data types.

2. IPBES

While GEO BON focuses on observations systems, other relevant international initiatives focus on assessment – broadly, the task of bringing science to bear on policy. The new Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, IPBES (<http://www.ipbes.net>) is to identify key scientific information needed for policymakers “for the conservation and sustainable use of biodiversity, long-term human well-being and sustainable development” (UNEP 2010b). IPBES also will catalyze new research to fill knowledge gaps. Significantly, IPBES will respond to requests from individual nations and from other stakeholders.

In considering the gaps in the science-policy interface that might be addressed by IPBES, UNEP (2009) pointed to the need to build on GBIF’s increasing “*collaboration with a wide range of organizations in order to explore the value of the data available, and to seek to combine it with other data meaningfully.*” The challenges raised by IPBES provide important context for our recommendations. IPBES is to “*perform regular and timely assessments of knowledge on biodiversity and ecosystem services and their interlinkages, which should include comprehensive global, regional and, as necessary, subregional assessments*” (UNEP 2009). Further, the

assessments are to be based on “a clear and transparent process for sharing and incorporating relevant data.” (UNEP 2009). IPBES will identify “core variables” related to biodiversity to support ongoing assessments of biodiversity and ecosystem services at multiple scales (Van Jaarsveld et al. 2011). These are to provide “some degree of standardization in approach across regional and subregional assessments”. These core variables logically should include various indices of biodiversity based on the “core observations” provided through GBIF.

This process would benefit from an expanded GBIF global infrastructure for mobilising and sharing biodiversity information (at various geographic scales and across national boundaries), together with GBIF’s provision of information to common standards and open access principles. IPBES therefore provides an important opportunity to address our recommendations that call for biodiversity conservation applications and for greater integration of other types of data.

3. Convention on Biological Diversity

The Convention on Biological Diversity has adopted 20 targets as part of its new Strategic Plan (UNEP 2010a). The Strategic Plan calls for GBIF inputs towards implementation of the Plan:

“The following are key elements to ensure effective implementation of the Strategic Plan:

...Global monitoring of biodiversity: work is needed to monitor the status and trends of biodiversity, maintain and share data, and develop and use indicators and agreed measures of biodiversity and ecosystem change; - The GEO-Biodiversity Observation Network, with further development and adequate resourcing, could facilitate this, together with Global Biodiversity Information Facility and the Biodiversity Indicators Partnership.”

One goal of the new Strategic Plan is that “*the science base and technologies relating to biodiversity, its values, functioning, status and trends, and the consequences of its loss, are improved, widely shared and transferred, and applied*”(UNEP 2010a). GBIF clearly can help address this broad goal.

GBIF also can play a specific role in helping to

provide a biodiversity calculus, based on GBIF mobilised species data, as a basic building block for addressing the 20 new CBD biodiversity targets. This can help ensure that strategies for individual targets address overall biodiversity. As one example, Target 11 calls for conservation of “at least 17 per cent of terrestrial and inland water, and 10 per cent of coastal and marine areas, especially areas of particular importance for biodiversity and ecosystem services” (UNEP 2010a). A given percentage area protected can have large or small representation of biodiversity; consequently assessment requires some explicit measurement of biodiversity coverage by protected areas. GBIF data could help to achieve a balance among the sometimes competing goals of conservation of ecosystems, biodiversity, and

ecosystem services. Thus, GBIF mobilised data should play a key role in providing the information base to assess whether conserved areas include those of particular importance for biodiversity.

Promoting ease-of-use and incentives for wide-use

Our CNA TG Survey raised concerns relating to ease of use that may be addressed by four recommendations (Table 3; again, a number of the recommendations have subsidiary recommendations). These recommendations focus on enhancements of the GBIF global discovery and access portals (including useful outputs), on creating incentives for users and contributors, and on design aspects that better link different data sources within a de-centralised system.

Table 3. Recommendations relating to promotion of ease-of-use and incentives for wide-use

11. Implement a ‘data publishing framework’, including mechanisms for incentivizing efforts for data volume mobilization, as well as data quality enhancement.
 - a. Develop a ‘data citation mechanism’ to adequately credit and acknowledge the contributions of all players in the data ‘life cycle’, from data creation to dissemination.
12. Enhance the data portals (<http://data.gbif.org>, and Participant portals) to facilitate ready to use processed outputs such as maps as images for immediate use in publications and reports.
 - a. Explore production of customized maps as output options.
 - b. Provide more easy to use output formats and data retrieval processes.
13. Enhance the GBIF network with infrastructure, services, standards, and tools that facilitate rapid and cost-efficient discovery and publishing of ‘fit-for-use’ primary biodiversity data.
 - a. Implement persistent identifiers at dataset and data record level.
 - b. Establish the data hosting center infrastructure across the GBIF network.
 - c. Ensure that time required for indexing data sets is minimized. Similarly, the gap between updates should be minimised.
 - d. Develop and publish the GBIF internationalization strategy and action plans, enhancing the ability of the network to discover, publish and use multilingual data resources.
 - e. Enhance GBIF’s ability to discover and publish datasets in their entirety.
 - f. Promote more national, regional, and thematic portals (and access points) in addition to global data portal (data.gbif.org), with features and services that would satisfy the needs of the cross-sectional stakeholders.
14. Enhance GBIF’s role as a discovery service within a decentralised system, where users can discover data using data descriptions contained in distributed metadata catalogues.

In our recommendations, one of the key themes is incentives that can expedite progress in discovery and publishing of primary biodiversity data. Progress would be promoted if data publishing was regarded as having the same status as traditional scientific (peer-reviewed) scholarly publishing (Chavan and Ingwersen 2009). In accord with this, we recommend an early uptake of recommendations of its Data Publishing Framework Task Group (GBIF 2009e; Moritz et al. 2011) and incentivisation mechanisms proposed (Chavan and Penev 2011; Chavan and Ingwersen 2009; Ingwersen and Chavan 2011; Goddard et al. 2011).

Another key theme relates to the increasing decentralisation of data and the need for meta-data strategies to help users find the data they need. In recent years, the focus of the GBIF network has shifted from data access to broader data discovery (GBIF 2008). Therefore, further enhancements are essential to broaden the coverage of data types discovered through the GBIF Data Portal (GBIF 2009f; GBIF 2009g; GBIF 2009h; Kelling et al. 2008; Morris et al. 2013, this volume), and to promote meta-data catalogues. As mentioned earlier, local-to-global scale strategies and action plans for discovery and publishing of biodiversity data are essential (Berendsohn et al. 2010; GBIF 2010a; GBIF 2010b; Gaiji et al. 2013, this volume; Otegui et al. 2013, this volume). This requires that GBIF establish a network of distributed ‘data hosting centers’ to provide user-friendly infrastructure for data publishing (Goddard et al. 2011). In tune with this, GBIF needs to better facilitate indexing data, and its discovery through data portal. These efforts might be enhanced by appropriate training and capacity development (e.g. Coetzer 2012).

Several of the key recommendations would help increase efforts to make it easy for a wide variety of users to discover GBIF. An example strategy that would address this is forging links to a new initiative through PLoS (Public Library of Science). PLoS argue that:

“Scientists are amassing details about the scope and status of life’s variation at an accelerating rate. This aids our understanding of species distributions and their interactions over time. However, if we are to address the consequences of

global environmental change for life’s future, biodiversity data must be integrated and synthesized to a much greater degree than they are at present, and this can be promoted by enhanced communication among the interested parties, and raising public awareness.”

PLoS have launched a Biodiversity Hub (<http://hubs.plos.org/web/biodiversity>) to promote such communication of biodiversity studies. The Biodiversity Hub is to add value to such studies in the form of data/images, etc. Thus, these added value contributions might serve to highlight uses of GBIF mobilised data.

CONCLUSIONS

The GBIF Content Needs Assessment Task Group investigated user needs relating to biodiversity data, and identified some major areas of opportunity to mobilize data serving these needs. The recommendations for GBIF fell into three classes: 1) data gaps, data volume, and data quality, 2) new kinds of data and new applications, and 3) promoting ease-of-use and incentives for wide-use. Addressing these challenges can serve the needs of international biodiversity initiatives, including the new biodiversity targets of the Convention on Biological Diversity, the global biodiversity observation network, GEO BON, and the new Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES). However, as is evident from the Content Needs Assessment Survey (Ariño et al. 2013, this volume), any user needs survey realistically provides just one glimpse at the evolving user landscape. One limitation of even the most up-to-date survey is that it cannot anticipate the full range of future applications and data needs. Further, any one survey will always have some bias in geographic coverage and/or scale. Future Content Needs Assessments no doubt will reflect the changing needs within what is a rapidly evolving multi-disciplinary biodiversity science. In that spirit, we finish by quoting from the recent call for major efforts to explore Earth’s species and map their distribution (Wheeler et al. 2012): *“There is ample evidence that clever scientists and advances in technology will continue to find new uses for museum specimens”*. This perspective no doubt applies across the whole spectrum of biodiversity data to be mobilized by GBIF.

LITERATURE CITED

- Andrefouet, S., M.J. Costello, D.P. Faith, S. Ferrier, G.N. Geller, R. Höft, N. Jürgens, M.A. Lane, A. Larigauderie, G. Mace, S. Miazza, D. Muchoney, T. Parr, H.M. Pereira, R. Sayre, R.J. Scholes, M.L.J. Stiassny, W. Turner, B.A. Walther, and T. Yahara. 2008: *The GEO Biodiversity Observation Network: Concept Document*. GEO - Group on Earth Observations, Geneva, Switzerland.
- Ariño, A.H., V. Chavan, and D.P. Faith (2013). Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges *Biodiversity Informatics*, 8, pp-pp
- Ariño, A.H., and S.L. Pimm (1995). On the nature of population extremes. *Evolutionary Ecology*, 9 (4): 429-443.
- Berendsohn, W.G., V. Chavan, and J. Macklin (2010). Recommendations of the GBIF Task Group on the Global Strategy and Action Plan for the Mobilization of Natural History Collections Data. *Biodiversity Informatics* 7: 67-71.
- Berents, P., M. Hamer, and V. Chavan (2010). Towards demand driven publishing: approaches to the prioritization of digitization of natural history collections data. *Biodiversity Informatics*, 7(2): 113-119.
- Boakes, E., 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*, 8:1-11.
- Bortolus, A. (2008). Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *AMBIO*, 37: 114-118.
- Bridgewater, P., S. Knapp, C. Prip, and M. MacDavette. (2010). *GBIF Review 2009. From Prototype to full operation: Managing expectations*. Copenhagen: Global Biodiversity Information Facility, May 3 (2010). 39pp.
- Brito, D. (2010). Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology* 11: 709–713.
- Butchart, S.H.M., M Walpole, B. Collen, A. van Strien., J.P.W. Scharleman, R.E.A. Almond, J.E.M. Baillie, B. Bomhard, C. Brown, J. Bruno, K.E. Carpenter, G.M. Carr, J. Chanson, A. Chenery, J. Csirke, N.C. Davidson, F. Dentener, M. Foster, A. Galli, J.N. Galloway, P. Genovesi, R. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington, J. Loh, M.A. McGeoch, L. McRae, A. Minasyan, M. Hernández Morcillo, T. Oldfield, D. Pauly, S. Quader, C. Revenga, J. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S.N. Stuart, A. Symes, M. Tierney, T.R. Tyrrell, J.-C. Vié, and R. Watson. (2010). Global biodiversity decline continues. *Science* 328: 1164-1168.
- 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 (available as a standalone PDF from <http://www.gbif.org>).
- Chavan, V. 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): S2, doi: 10.1186/1471-2105-10-S14-S2.
- Chavan, V. and L. Penev. (2011). The Data Paper: A mechanism to incentivise publishing in biodiversity science. *BMC Bioinformatics*, 12(Suppl 15):S2, doi:10.1186/1471-2105-12-S15-S2.
- Coetzer, W., M. Hamer, and F., Parker-Allie. (2012). A new era for specimen databases and biodiversity information management in South Africa. *Biodiversity Informatics* 8: 1-11.
- 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): 97-110.
- Collen, B., M. Ram, N. Dewhurst, V. Clausnitzer, V.J. Kalkman, N. Cumberlidge, and J.E.M. Baillie. (2009). Broadening the coverage of biodiversity assessments. In: *Wildlife in a changing world - an analysis of the 2008 IUCN Red List of Threatened Species* (ed. by J.-C. Vié, C. Hilton-Taylor & S.N. Stuart), pp. 67-76. IUCN, Gland.
- Couvet, D., V. Devictor, F. Jiguet, and R. Julliard. (2011). Scientific contributions of extensive biodiversity monitoring. *C. R. Biologie* 334: 370–377.
- Davies, N., C. Meyer, J.A. Gilbert, L. Amaral-Zettler, J. Deck, M. Bicak, P. Rocca-Serra, S. Assunta-Sansone, K. Willis, and D. Field. (2012). A call for an international network of genomic observatories (GOs). *GigaScience* 2012: 1:5.
- Diaz, S., A. Hector, and D.A. Wardle. (2009). Biodiversity in forest carbon sequestration initiatives: not just a side benefit. *Current Opinon in Environmental Sustainability*, 1: 55-60.

- Escala, M.C., J.C. Irurzun, A. Rueda, and A.H. Ariño (1997). Atlas of the Insectivora and Rodentia of Navarra. Biogeographical analysis. *Publ. Biol. Univ. Navarra, Ser. Zool.* 25: 1-79.
- Eigenbrod, F, B.J. Anderson, P.R. Armsworth, A Heinemeyer, S Gillings, D.B. Rob, C.D. Thomas, and K.J. Gaston. (2010). Representation of ecosystem services by tiered conservation strategies. *Conservation Letters*, 3(3): 184-191.
- Escobar, F., P. Koleff, and M. Rös. (2009). Assessment of the capacities for knowledge: the National Biodiversity Information System as a case study. In: CONABIO-UNDP (eds.) *Mexico: capacities for conservation and sustainable use of biodiversity*. National Commission for the Knowledge and Use of Biodiversity and the United Nations Development Programme, Mexico, pp 24-29. Accessible at http://www.biodiversidad.gob.mx/v_ingles/country/pdf/Capacities.pdf
- Faith, D. P. (2005). *Indicator taxa in support of the 2010 Biodiversity Target*. GBIF Seed Money Prioritization ECAT - Electronic Conference. (May 25 – June 1, 2005).
- Faith, D. P. (2011). Ecosystem Services and Biodiversity Option Values. *Science Online*. 10 February 2011. <http://www.sciencemag.org/content/330/6012/1745/reply>
- Faith, D. P (2012). Ecosystems on the EDGE: Ecologically Distinctive Globally Endangered. <http://australianmuseum.net.au/Ecosystems-on-the-EDGE-Ecologically-Distinctive-Globally-Endangered>
- Faith, D. P., and P.A. Walker. (1996). Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodivers. Conserv.* 5: 399–415.
- Faith, D.P., S. Ferrier, and P.A. Walker. (2004). The ED strategy: how species-level surrogates indicate general biodiversity patterns through an ‘environmental diversity’ perspective. *J. Biogeogr.* 31: 1207–1217.
- Faith, D.P., S. Ferrier, and K.J. Williams. (2008). Getting biodiversity intactness indices right: ensuring that “biodiversity” reflects “diversity. *Global Change Biology* 14: 207–217.
- Faith D.P., C.A. Lozupone, D. Nipperess, and R. Knight. (2009). The Cladistic Basis for the Phylogenetic Diversity (PD) Measure Links Evolutionary Features to Environmental Gradients and Supports Broad Applications of Microbial Ecology’s “Phylogenetic Beta Diversity” Framework. *Int. J. Mol. Sci.* 10:4723-4741.
- Faith, D.P., S.Magallón, A.P. Hendry, E. Conti, T Yahara, M.J. Donoghue. (2010). Ecosystem Services: an evolutionary perspective on the links between biodiversity and human-well-being. *Curr. Opin. Environ. Sustain.* 2: 66–74.
- Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology* 51: 331–363.
- Gaiji, S, V. Chavan, A.H. Ariño, J. Otegui, E. Robles, and N. King. (2013). Content assessment of the primary biodiversity data through GBIF network: Status, challenges and Potentials. *Biodiversity Informatics*, this volume.
- Gaikwad, J. and V. Chavan. (2006). Open access and biodiversity conservation: Challenges and Potentials for the developing world. *Data Science Journal*, 5: 1-17.
- GBIF. (2008). *GBIF Work Programme 2009-2010*. Copenhagen, 59 pp. Accessible at <http://www2.gbif.org/WP2009-10.pdf>
- GBIF. (2009a). *Content Needs Assessment Task Group*. Copenhagen: Global Biodiversity Information Facility. Accessible at <http://www.gbif.org/informatics/primary-data/task-groups/cna-tg/>.
- GBIF. (2009b). *GBIF Content Needs Assessment Survey (2009)*. Accessible at <http://www.gbif.org/communications/news-and-events/showsingle/article/gbif-content-needs-assessment-survey-2009/>. Last accessed 2010.12.26.
- GBIF. (2009c). *GBIF Content Needs Assessment Survey 2009 available in Chinese*. <http://www.gbif.org/communications/news-and-events/showsingle/article/gbif-content-needs-assessment-survey-2009-available-in-chinese/>. Last accessed 2010.12.26.
- GBIF. (2009d). *GBIF Content Needs Assessment Survey 2009 available in Russian*. <http://www.gbif.org/communications/news-and-events/showsingle/article/russkaja-versija-voprosnika-po-vyjasneniju-potrebnosti-v/>

- GBIF. (2009e). *Data Publishing Framework Task Group*. Accessible at <http://www.gbif.org/informatics/primary-data/task-groups/dpf-tg/>
- GBIF. (2009f). *Adoption of Persistent Identifiers for Biodiversity Informatics: Recommendations of the GBIF LSID GUID Task Group*. 23p. Accessible at <http://www2.gbif.org/Persistent-Identifiers.pdf>
- GBIF. (2009g). *Report of the GBIF GBRDS Stakeholders Planning Workshop*. 24p. Accessible at <http://www2.gbif.org/GBRDSWorkshopReport-Final.pdf>
- GBIF. (2009h). *Report of the GBIF Metadata Implementation Framework Task Group*. 38pp. Accessible at <http://www2.gbif.org/GBIF-MIFTG-Report.pdf>
- GBIF. (2010a). *State-of-the-Network 2010: Discovery and Publishing of the Primary Biodiversity Data through the GBIF Network*. Authored by Chavan, V. S., Gaiji, S., Hahn, A., Sood, R. K., Raymond, M., and N. King. (2010). Copenhagen: Global Biodiversity Information Facility, 36 pp. ISBN: 87-92020-13-5. Accessible online at <http://www.gbif.org>.
- GBIF. (2010b). *Best practice guide for 'Data Discovery and Publishing Strategy and Action Plans' version 1.0*. Authored by Chavan, V. S., Sood, R. K., and A. H. Ariño. (2010). Copenhagen: Global Biodiversity Information Facility, 29 pp. ISBN: 87-92020-12-7. Accessible online at <http://www.gbif.org>.
- GBIF. (2010c). *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 on-line at <http://www2.gbif.org/GPP-Final.pdf>.
- GBIF. (2011). *GBIF Strategic Plan 2012–2016 - Seizing the Future*. Copenhagen, Denmark: Global Biodiversity Information Facility.
- Goddard, A., P. Cryer, N. Wilson and G. Yamashita. (2011). Data hosting infrastructure for primary biodiversity data. *BMC Bioinformatics* 12(Suppl 15): S5.
- Gropp, R.E. (2012). Increasing access to biological collections. *Bioscience* 62: 703.
- Hardisty, A., Roberts, D. and The Biodiversity Informatics Community. 2013.. *BMC Ecology* 2013, 13:16. <http://www.biomedcentral.com/1472-6785/13/16>
- Hortal, J., A. Jiménez-Valverde, J.F. Gómez, J.M. Lobo, and A. Baselga. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117(6):847-858.
- Ingwersen, P. and V. Chavan. (2011). Indicators for Data Usage Index: An incentive for publishing primary biodiversity data through global information infrastructure. *BMC Bioinformatics* 12 (Suppl 15): S3.
- Juutinen, A., M. Mönkkönen, and M. Ollikainen. (2008). Do environmental diversity approaches lead to improved site selection? A comparison with the multi-species approach. *Forest Ecology and Management* 255: 3750-3757.
- Kelling, S., B. Ingole, B. Daly, B. Stein, D. Lepage, E. O'Tuama, J. Cooper, M. Jones, T. Lahti, and V. Chavan. (2008). *Recommendations of the GBIF Observational Data Task Group*. September 2008, Pp. 21.
- Likens, G. E. (Ed.) (1989). *Long-term studies in Ecology*. Springer-Verlag, New York.
- Linke, S., E. Turak, and J. Nel. (2011). Freshwater conservation planning: the case for systematic approaches. *Freshwater Biology* 56: 6–20.
- Lotze, H. K. and B. Worm. (2009). Historical baselines for large marine mammals. *Trends in Ecology & Evolution*, 24: 254–262.
- Mace, G. M. and J. E. M. Baillie. (2010). The 2010 Biodiversity Indicators: Challenges for Science and Policy. *Conservation Biology*, 21 (6): 1406–1413.
- Mace, G. M., B. Collen, R.A. Fuller, and E.H. Boakes. (2010). Population and geographic range dynamics: implications for conservation planning. *Philosophical Transactions of the Royal Society of London B*, 365; 3743-3751.
- Margalef, R. (2000). Diversidad y Biodiversidad. In: *La Diversidad Biológica en España*, F. Pineda, J.M. de Miguel, M.A. Casado, J. Montalvo (eds). Prentice Hall, Madrid.
- Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Responses*. World Resources Institute: Washington, DC, USA (2005). <http://www.millenniumassessment.org/en/Reports.aspx>
- Morris, R.A., A. Olson, G. Riccardi, G. Whitebread, V. Barve, G. Hagedorn, P. Leary, I. Teage, D. Mozzherin, C. Freeland, M. Carausu, J. Cuarda, and V. Chavan (2012). Discovery and Publishing of Primary Biodiversity Data associated with

- Multimedia Resources: The Audubon Core Strategies and Approaches. *Biodiversity Informatics*, 8
- Moritz, T., S. Krishnan, D. Roberts, P. Ingwersen, D. Agosti, L. Penev, M. Cockerill, , and V. Chavan. (2011). Towards mainstreaming of biodiversity data publishing: recommendations of the GBIF Data Publishing Framework Task Group. *BMC Bioinformatics*, 12(Suppl.15):S1.doi:10.1186/1471-2105-12-S15-S1.
- Norse, E.A., K.L. Rosenbaum, D.S. Wilcove, B. A. Wilcox, W.H. Romme, D.W. Johnston, and M. L. Stout. (1986). *Conserving biological diversity in our national forests*. Washington, D.C.: The Wilderness Society.
- Otegui, J, A.H. Ariño, V. Chavan, and S. Gaiji. (2011). On the dates of the GBIF mobilized primary biodiversity data records. *Biodiversity Informatics*, 8
- Page, R.D.M. (2008). Biodiversity informatics: the challenge of linking data and the role of shared identifiers. *Brief Bioinform* 9: 345-354. doi: 10.1093/bib/bbn022
- Parr, C.S., R. Guralnick, N. Cellinese, R.D.M. Page. (2011). Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology & Evolution*, 27: 94-103.
- Peterson, A.T., D. Canhos, U. Gardenfors, R.J. Scholes, Y. Shirayama, M.S. Graham, and F. Pando. (2010). *Global Biodiversity Information Facility: 2010 Forward Look Report*. Copenhagen: Global Biodiversity Information Facility, 17 August (2010). 45pp.
- Pyke, G.H., and P.R. Ehrlich. (2010). Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews*. 85: 247 – 266.
- Raxworthy, C.J., E. Martinez-Meyer, N. Horning, R.A. Nussbaum, G.E. Schneider, M.A. Ortega-Huerta and A.T. Peterson. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837-841.
- Robertson, M.P., G.S. Cumming, and B.F.N. Erasmus. (2010). Getting the most out of atlas data. *Diversity and Distributions* 16: 363–375.
- Robertson, D.R. (2008). Global biogeographical data bases on marine fishes: caveat emptor. *Diversity and Distributions* 14: 891-892.
- Rodríguez, J.P.K., M. Rodríguez-Clark, D.A. Keith, E.G. Barrow, J. Benson, E. Nicholson, and P. Wit. (2012). IUCN Red List of Ecosystems, *S.A.P.I.E.N.S [Online]*, 5.2 | 2012, accessed 27 August (2012). <http://sapiens.revues.org/1286>
- Sarkar, I.N., and M. Trizna. (2011). The Barcode of Life Data Portal: Bridging the Biodiversity Informatics Divide for DNA Barcoding. *PLoS ONE* 6(7): e14689. doi:10.1371/journal.pone.0014689
- Secretariat of the Convention on Biological Diversity (2010). *Global Biodiversity Outlook -3*. ISBN: 92-9225-220-B, accessible at <http://gbo3.cbd.int/>, Montreal, 94 pp.
- Soberón, J. and A.T. Peterson. (2009). Monitoring biodiversity loss with primary species-occurrence data: towards national level indicators for the 2010 target of the Convention on Biological Diversity. *AMBIO*, 38(1): 29-34.
- Sutherland, W.J., W.M. Adams, R.B. Aronson, R. Aveling, T.M. Blackburn, S. Broad, G. Ceballos, I.M. Cote, R.M. Cowling, G.A.B. Da Fonseca, E. Dinerstein, P.J. Ferraro, E. Fleishman, C. Gascon, M. Hunter Jr, J. Hutton, P. Kareiva, A. Kuria, D. W. Macdonald, K. Mackinnon, F.J. Madgwick, M.B. Mascia, J. McNeely, E.J. Milner-Gulland, S. Moon, C.G. Morley, S. Nelson, D. Osborn, M. Pai, E.C.M. Parsons, L.S. Peck, H. Possingham, S.V. Prior, A.S. Pullin, M.R.W. Rands, J. Ranganathan, K.H. Redford, J.P. Rodriguez, F. Seymour, J. Sobel, N.S. Sodhi, A. Stoti, K. Vance-Borland, and A.R. Watkinson. (2009). One hundred questions of importance to the conservation of global biological diversity. *Conservation Biology* 23: 557-567.
- Tews, J. (2006). Biodiversity and Climate Change: A Modelling Perspective. In: *Focus on Biodiversity Research* (editor: Jan Schwartz), ISBN: 1-60021-372-3, pp. 15.
- Tallis, H. and S. Polasky. (2009). Mapping and valuing ecosystem services as an approach for conservation and natural-resource management. *Annals of the New York Academy of Sciences*, 1162: 265-283.
- UNCSO. (2012). *The Future We Want*. United Nations Conference on Sustainable Development Rio de Janeiro, Brazil. http://www.uncsd2012.org/content/documents/727_The_Future_We_Want_19_June_1230pm.pdf Cited 1 August 2012.
- UNEP. (2009). Gap analysis for the purpose of facilitating the discussions on how to improve and strengthen the science-policy interface on biodiversity and ecosystem services. UNEP/IPBES/2/INF/1.

- UNEP. (2010a). *UNEP/CBD/BS/COP-MOP/2/WG.1/CRP.1 Updating and revision of the Strategic Plan for the post-2010 period*. Convention on Biological Diversity, 2010. available online: <http://www.cbd.int/cop/cop-10/doc/advance-final-unedited-texts/advance-unedited-version-strategic-plan-footnote-en.doc>.
- UNEP. (2010b). Intergovernmental science-policy platform on biodiversity and ecosystem services. UNEP/GC.26/6.
- Van Jaarsveld, A.S. et al. (2011). *International Science Workshop on Assessments for IPBES*. United Nations University, Tokyo, Japan 25- 29 July 2011 Workshop report.
- Warner, S.C., K.E. Limburg, A.H. Ariño, M. Dodd, J. Dushoff, K.I. Stergiou, and J. Potts. (1995). Time Series Compared Across the Land-Sea Gradient. In: *Ecological Time Series*, T.M. Powell & J.H. Steele (Eds.). Chapman & Hall, New York. Pp. 242-273.
- Wheeler, Q.D., P.H. Raven, and E.O. Wilson. (2004). Taxonomy: Impediment or Expedient? *Science* 303(5656): 285.
- Wheeler, Q.D., S. Knapp, D.W. Stevenson, J. Stevenson, S.D. Blum, B.M. Boom, G.G. Borisy, J.L. Buizer, M.R. De Carvalho, A. Cibrian, M.J. Donoghue, V. Doyle, E.M. Gerson, C.H. Graham, P. Graves, S.J. Graves, R.P. Guralnick, A.L. Hamilton, J. Hanken, W. Law, D.L. Lipscomb, T.E. Lovejoy, H. Miller, J.S. Miller, S. Naeem, M.J. Novacek, L.M. Page, N.I. Platnick, H. Porter-Morgan, P.H. Raven, M.A. Solis, A.G. Valdecasas, S. Van Der Leeuw, A. Vasco, N. Vermeulen, J. Vogel, R.L. Walls, E.O. Wilson, and J.B. Woolley. 2012. Mapping the biosphere: exploring species to understand the origin, organization and sustainability of biodiversity. *Systematics and Biodiversity* 10:1, 1-20
- Wilson, E.O. (1988). The current state of biological diversity. In: *Biodiversity*, E. O. Wilson editor, pp. 3-18. National Academy Press, Washington, D. C. 521 pp.
- Wilson, E.O. (1999). *The diversity of Life* (Reissue edn.) W.W. Norton, New York,.
- Yesson, C., P.W. Brewer, T. Sutton, N. Caithness, J.S. Pahwa, et al. (2007). How Global is the Global Biodiversity Information Facility? *PLoS ONE*, 2(11): e1124 (doi:10.1371/journal.pone.0001124).