Building a global observing system for biodiversity

Robert J Scholes¹, Michele Walters¹, Eren Turak²,³, Hannu Saarenmaa⁴, Carlo HR Heip⁵, Éamonn Ó Tuama⁶, Daniel P Faith⁷, Harold A Mooney⁸, Simon Ferrier⁹, Rob HG Jongman¹⁰, Ian J Harrison¹¹, Tetsukazu Yahara¹², Henrique M Pereira¹³, Anne Larigauderie¹⁴ and Gary Geller¹⁵

The Group on Earth Observations Biodiversity Observation Network (GEO BON) has been in formal existence for three years, following several years of design and discussion. It is the realisation of the biodiversity societal benefit area envisaged in the GEO System of Systems (GEOSS). GEO BON links together existing networks, each covering particular aspects of biodiversity or parts of the world, and takes steps to help fill important gaps in the system. GEO BON focuses on coordination and harmonisation of the existing and emerging systems; advocacy and action to sustain the observing systems and to fill the identified gaps; and understanding and servicing user needs for biodiversity observations, particularly in the policy-making domain.

Addresses
¹ CSIR Natural Resources and Environment, PO Box 395, Pretoria 0001, South Africa
² NSW Office of Environment and Heritage, Department of Premier and Cabinet, PO Box A290, Sydney South, NSW 1232, Australia
³ Australian Museum, Sydney, NSW 2000, Australia
⁴ University of Eastern Finland, 80101 Joensuu, Finland
⁵ Netherlands Institute of Ecology, Center for Estuarine and Marine Ecology, PO Box 140, 4400 AC Yerseke, The Netherlands
⁶ Global Biodiversity Information Facility, Universitetsparken 15, Dk-2100 Copenhagen Ø, Denmark
⁷ The Australian Museum, 6 College St., Sydney, NSW 2010, Australia
⁸ Department of Biology, Stanford University, Stanford, CA 94305, USA
⁹ CSIRO Ecosystem Sciences, GPO Box 1700, Canberra ACT 2601, Australia
¹⁰ Alterra, Wageningen UR, PO Box 47, 6700AA Wageningen, The Netherlands
¹¹ Conservation International and IUCN Biodiversity Assessment Unit, 2011 Crystal Drive, Suite 500, Arlington, VA 22202, USA
¹² Center for Asian Conservation Ecology, Department of Biology, Kyushu University, 6-10-1 Hakozaki, Fukuoka 812-8581, Japan
¹³ Centro de Biología Ambiental, Facultad de Ciencias da Universidade de Lisboa, Campo Grade, 1749-016 Lisboa, Portugal
¹⁴ DIVERSITAS, Museu Nacional d’Histoire Naturelle (MNHN), 57 Rue Cuvier – CP 41, 75231 Paris Cedex 05, France
¹⁵ Jet Propulsion Laboratory, California Institute of Technology, MS171-264, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA

Corresponding author: Scholes, Robert J (bscholes@csir.co.za)

Introduction

There is a broadly based emerging understanding of the need for a comprehensive and integrated observation system for biodiversity at several scales, from the subnational to the global for the purpose of protecting and improving biodiversity and human well-being [1,2,3,4]. The system should help to compare the status of biodiversity at different places and track changes in biodiversity at a given place over time. The main stakeholders in this enterprise include national biodiversity and natural resource communities; nature conservation and management agencies; national government departments responsible for biodiversity-related international treaty obligations; the treaty secretariats and United Nations organisations; non-governmental organisations (both national and international) involved with the monitoring and conservation of biodiversity; and a variety of mostly science institution-based biodiversity information centres.

The Johannesburg World Summit on Sustainable Development in 2001 identified the urgent need for globally shared, adequate information in support of the collective management of the environment. This led, following a period of negotiation and design, to the formation of the Group on Earth Observations (GEO) in 2005, a voluntary and growing partnership of (currently) some 87 governments, the European Commission and 64 international organisations, dedicated to improving the availability and interoperability of information relating to the global environment. For its initial 10-year implementation plan, GEO is organised into nine Societal Benefit Areas (SBAs), of which one is biodiversity [5]. The GEO Biodiversity Observation Network (GEO BON) was formed in February 2008 to give effect to actions in this domain.

An example of the need for better biodiversity information is provided by the United Nations Convention on Biological Diversity (CBD). The parties to the CBD — which include virtually all countries — set themselves a target of reducing the rate of loss of biodiversity by the year 2010. Twenty-one subtargets were identified,
aimed at achieving 11 principal goals for reducing the rate of biodiversity loss [6: Table 1]. Nevertheless, reliable information on many of the proposed indicators was unavailable at global scale [7,8]. From the data that were compiled, it was recognised that despite some local successes, many of the subtargets were not met at the global scale and the selected indicators show that there has been no reduction in the overall rate of biodiversity loss [6: Tables 1 and 2]. An analysis on a broader set of 31 indicators of the state of biodiversity also shows a decline in the state of, and an increase in pressures on, biodiversity [9]. The projections for the remainder of the century are alarming [10]. At its Conference of the Parties to the convention in Nagoya in 2010, the CBD agreed to the new strategic plan for biodiversity including the twenty ‘Aichi targets’ for the period 2011–2020, and asked GEO BON to help advise on how the datasets can be assembled. The initial step in this process was an Adequacy Report produced by GEO BON in 2011 [11].

GEO BON does not start from a blank slate. Biodiversity observations are among the most numerous and longest recorded observations of the environment. There are vast collections of plants and animals in museums and herbaria around the globe (one estimate suggests 2–3 billion such records [12]); hundreds of millions of observations in the field by professional and lay experts; and terabytes of remotely sensed images and maps of the changing cover on the land surface. Most observations, however, are local, uncoordinated, in situ observations and very few are globally harmonised. Moreover marine, freshwater and terrestrial observation systems are not connected and many differences in observation approaches and taxonomy exist. To complicate matters further, remotely sensed observations are often used to describe land cover and not ecosystems.

A range of national and international organisations have arisen to deal with these data and bring it into the interconnected, digital age. Prominent examples are: first, the Global Biodiversity Information Facility (GBIF), a ‘megascience’ initiative to share digital information, initially largely addressing museum and herbarium collections in participating developed countries, but now with field observation data predominating among the more than 300 million primary biodiversity records accessible via their data portal, and including participants from around the world, and showing huge potential for growth (in a state-of-the-network review in 2010, one third of GBIF Participants reported that while nearly 2.5 billion records were available within their domain only some 820 million were in digital form [13]); second, the Red List of Threatened Species assembled by the International Union for the Conservation of Nature (IUCN); third, NatureServe’s Explorer and InfoNatura databases (on biodiversity in the US, Canada, and Latin America); fourth, the global maps of freshwater, marine, and terrestrial ecoregions assembled by the World Wildlife Fund (WWF) and partners; fifth, the maps of protected areas maintained by UNEP’s World Conservation Monitoring Centre; and sixth, the International Nucleotide Sequence Database Collaboration formed by the European Molecular Biology Laboratory (EMBL), GenBank Data Libraries (GenBank) and DNA Data Bank of Japan (DDBJ).

While each of these initiatives meet some of the information needs required for establishing overall spatial and temporal trends in biodiversity status for every corner of the world, they exist for particular purposes, each with their own scope and mandate. This means that gaining an integrated, comprehensive, geographically and taxonomically balanced view is extremely difficult without a mechanism that links them to an overarching goal. Bisby [14] commented on the richness of electronic databases but the ‘maddening difficulty in knowing what is where, or of comparing like with like.’ Several authors [e.g, 15] noted the importance of ensuring biodiversity databases are regularly updated and linked to ensure maximum applied value for conservation and management. This is the challenge for GEO BON — not to supersede the existing networks, but to efficiently connect them to one another and to users, and to help identify and fill the gaps.

Two meetings of interested parties were called between 2006 and 2008, and a concept document and initial implementation plan were developed [16,17]. An interim steering committee was formed, leading to a full steering committee and a detailed implementation plan [18,19]. There is currently an active programme of work, based around eight, largely self-organising, working groups, in the areas of: gene-level observations; terrestrial species observations; terrestrial ecosystem observations; freshwater ecosystem observations; marine ecosystem observations; ecosystem services; model-based integration of in situ and remotely sensed data; and data integration and interoperability.

**Characteristics of the desired system**

A global biodiversity observation system has certain desirable characteristics. It should be:

- On the basis of the GEO Data Sharing Principles [20], calling for the full and open exchange of data, metadata, and products, while contributing to data architecture, standards and interoperability efforts. It should recognise relevant international and national efforts for the appropriate protection of sensitive data, for example, the precise location of threatened species.
- Relevant to the goal of determining the status of, and change in biodiversity, to help protect and improve biodiversity alongside human well-being.
- Orientated toward the needs of users who share this goal and easily accessible to these users. GEO is explicitly a user-focused organisation, trying to balance the
widespread tendency for observation systems to be driven by technology-push rather than user-pull.

- Focussed on **tracking change at broad spatial scales**. This dictates the way in which baselines are established and future observations are planned. On the other hand, a global sampling frame that scales to local, national and regional scales is required to provide a statistically sound basis for repeated measurements of biodiversity. It is also essential to coordinate and manage **in situ** data that are collected by disparate institutions and individuals for different purposes. Similarly such a system would need to be partial toward future monitoring and favour observations yielding outputs that accurately reflect the overall status of biodiversity among different places at different times.

- **Global in coverage**, but with sufficient resolution and accuracy at subnational scales to be useful to the main decision-makers at this scale, who typically have national or subnational jurisdiction. Currently, biodiversity information holdings are concentrated in developed, temperate countries, while biodiversity itself is concentrated in developing, tropical countries.

- Sufficiently **comprehensive in terms of taxonomic coverage**. There are approximately 1.9 million scientifically described species [21], but the actual number is likely to be much higher, with estimates of the total number of eukaryotic species most commonly falling between 5 million and 30 million [22]. Currently biodiversity observations focus disproportionately on certain popular and easily observed groups: birds, mammals and higher plants especially (see, for example, the taxonomic breakdown of data shared on the GBIF network [13]), whereas phylogenetic variation is greater in many of the less well surveyed groups such as insects and microbes, for instance.

- **Quality controlled**. At a minimum, this means traceability of the observation to its place and time of origin, including the techniques used to make the observation and subsequently modify the data. Apart from filtering out obviously incorrect information (impossible locations or dates, non-existent species), the objective is to be able to associate the data with a known degree of accuracy, and then let the user decide on its fitness-for-use (see e.g., GBIF [23]).

- **Cost efficient**. Replication is a scientific virtue, but duplicative work in recording or analysing the same observations for the same time period, and highly uneven sampling intensities, are both examples of inefficiencies that waste scarce observational resources.

- **Sustained**. Data continuity and comparability over time is essential for many applications, not least the monitoring of biodiversity loss, mitigation and recovery. Observation systems must nevertheless be responsive to new technical possibilities and emerging societal needs, or they quickly become irrelevant.

- **Interoperable**. This means that data from one part of the system must not only be available to (and discoverable by) other parts of the system, but it must be possible to analyse them together. Achieving this characteristic requires that the data be accompanied by adequate metadata, describing the data. It also requires harmonisation of observations, analysis and data exchange standards and protocols. The need for interoperability extends beyond biodiversity data itself, to the associated biophysical or social databases needed for its interpretation. Data integration and interoperability concerns are addressed in the recommendations for a GEOSS Common Infrastructure (GCI) [24] and inform development and implementation of the GEO BON informatics infrastructure (see chapter 8 in GEO BON [19]).

### Integration in many dimensions

An enormous amount of time and money already goes into biodiversity observation systems, mostly at the national level, and mostly with public money. Without this continued investment, none of the rest of the system can operate. The value addition that a system such as GEO BON brings is that for a relatively small additional investment in integration, much greater value can be realised. This integration takes many forms:

- **Across the many taxonomic forms of biodiversity and between the gene, species and ecosystem levels into which it is organised**. Given that we are probably decades from having a sufficient observation system in many taxa — some functionally very important — to what degree can well-observed taxa act as proxies for poorly observed taxa? Can butterflies, for instance, represent the sensitivity of all insects to insecticides? A specific challenge is an observation system for genetic diversity, covering many taxa. Proxies again may help — for example, observations on changes in range extent at the species level may allow some inferences about corresponding changes in genetic diversity. Although many genetic, species and ecosystem databases have been developed independently, there is a growing body of opportunities to link them for integrated analyses of biodiversity states and changes. First, rapid technical advance in genome science has enabled us to apply genetic approaches to adaptive variation of species and ecosystem function of both microbes and macro organisms [25]. Second, accumulation of georeferenced distribution records of species has opened the way to modelling distributions of a large number of species and projecting changes of ecosystem composition [26,27]. Third, development of a large database of plant functional traits (TRY; see Figure 1) has linked biodiversity research with other earth system science components [28].

- **Across environmental realms**. It is well known that terrestrial runoff has large effects upon coral ecosystems [29]. Nutrient transfer from sea to land is often
critical to maintain terrestrial vegetation [30]. More generally, the importance of linking biodiversity data and conservation planning across marine, freshwater and terrestrial realms is recognised [31,32]. The logistic, institutional and political constraints that stand in the way of such integration [33] need not present a major problem for the development of a biodiversity observation system but will require a concerted effort to bring together the experts and implementation partners operating in each of these realms. A major strength of GEO BON as a system is its efforts toward linking these realms.

- **Integrating ecosystems, habitats and species.** The system should recognise links between ecosystems and habitats and let them act as proxies for species, in order to make use of global observation systems, such as remote sensing, and to facilitate the regular, cost effective, assessment of biodiversity change.

- **Between biodiversity and its abiotic drivers and correlates, such as climate and habitat.** Biodiversity is not randomly distributed around the globe. If we understand and model non-random distribution patterns of species, functional groups and ecosystems, monitoring and protection become much more efficient. This, for instance, requires integrating biodiversity data, spatially and temporally, with rainfall, temperature, elevation and soils data in terrestrial environments, and with temperature, light and water depth, and primary productivity in the oceans.

- **Between social and natural sciences.** The drivers of biodiversity loss relate to human activities such as land use change, harvest pressures, fisheries, pollution, wildlife trade, and introduction of exotic species. The interventions to protect biodiversity involve laws, economic measures and institutions and requires a general framework for social-ecological systems.
Biodiversity has not only ecological and intrinsic value, but cultural and economic value as well. These connections cannot be fully revealed unless the role that biodiversity plays in maintaining ecosystem functions are identified, and the value of the consequent services delivered by those functions are defined. This requires the integration of natural science data with political, social and economic data, thus providing an essential resource for the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) and strengthening the use of science in policy making. Traditionally, these databases have been entirely separate, and even incompatible in terms of spatial and temporal resolution.

- **Between collection data, direct (in situ) observations, and remotely sensed observations.** Each of these main biodiversity data types has unique properties, advantages and disadvantages. Bringing them together makes for a more useful package than keeping them separate. For instance, marine surveys are greatly assisted by remote sensing, yet remote sensing covers only a very limited depth range and therefore needs to be supplemented by *in situ* observations. Existing collections data, such as those mobilised by GBIF, provides a rich source of information for spatial biodiversity models when integrated with environmental observations. However, such data on its own typically do not provide the time series needed to assess biodiversity change. Integration of these models nevertheless can act as a biodiversity ‘lens’ to interpret the time series information on changes in land condition derived from remote sensing observations.

- **Between national systems, independently developed.** Since the main historic drivers for biodiversity observations were focussed at the national or subnational level, a large number of independent systems arose, with their own standards and in their own language. They will continue to exist, because the mandate, funding and stakeholders require them to do so. However, sharing their observations (many of which relate to biodiversity found in other countries as well) has enormous collective benefits, and some local benefits, such as the security of having the data separately archived, and the quality control provided by making the data holdings widely visible. Since it is unreasonable to expect existing systems to completely redesign their databases for this purpose, the effort mainly goes into adopting common data exchange standards and protocols, such as Darwin Core [36], ontologies and semantic mediation layers for data harmonisation, and establishing data sharing agreements such as those championed by GEO and GBIF. For countries establishing new systems, the benefits of using existing designs are great.

- ** Seamlessly from data providers to data users and back.** Biodiversity observation is a system, not a simple linear flow from primary observation to use. What should be observed? How often and where? How should the raw observations be summarised and analysed to make them maximally useful? These questions can only be answered in dialogue between the many components of the system. It should not be necessary for users to navigate multiple, arcane interfaces and negotiate access to all datasets individually. From their perspective, it should all seem like one, unified system.

**An achievable vision**

The information technology tools, the political environment, and the biodiversity informatics community are all ready for a new approach that will provide a quantum leap in the amount, reliability and usefulness of biodiversity information that is available. The main advances will come from two sources: the synergies unlocked by taking a systems view of the issue; and the increase in effort and efficiency that will result from focussing on removing data impediments and filling data gaps.

A preliminary version of the ‘systems view’ that GEO BON envisages is shown in Scholes et al. [4]. An elaborated version is shown in Figure 1. It consists of a rich, but not infinite, set of biodiversity data types, logically linked to one another in such a way that a wide range of user needs can be addressed — probably including user needs which are not yet fully articulated. A key element of the linkage involves ‘semantics’ — shared data fields with defined meanings that allow records to be linked within and across databases. A good example in the biodiversity field is the recognised scientific name (and its synonymy and history of changes) as a unique species identifier, allowing databases covering genetics, preserved specimens, populations status, and geographic range and distribution to be connected. Without efforts such as Species 2000/ITIS Catalog of Life, the Catalog of Fishes and World Register of Marine Species (WoRMS), which created catalogues of species names, these links would not be possible. Knowledge Organisation Systems will provide the semantic foundations for an integrated system for biodiversity observations (cf. [37]). Once in place, the ease by which data can be queried and manipulated becomes greatly expanded.

A key part of this vision is the role which ‘models’ play as the mechanism for connecting disparate *in situ* and remotely sensed data types [38], and especially for bridging data gaps until such time as they can be filled with observations. By combining state-of-the-art models with the notion of reporting the degree of uncertainty associated with both data and modelled predictions, it is not necessary to wait until all aspects of biodiversity are equally perfectly observed before useful products can be generated. Spatially explicit mapping of such uncertainties then provides a guide as to where the observation system most needs strengthening. Thus, robust models simultaneously make best-possible use of currently available observations and also point to the observations gaps
that are needed to improve the models and inferences about biodiversity status and change.

Models also create value by adding new insights to the primary observations. Typically in the information system there are three layers: the primary observations themselves; a layer of indicators formed by combinations of the primary observations; and a decision-making layer of reduced detail and greater integration, directed at informing the policy objectives. While there should be as much stability over time in the primary observations, the sets of indicators and decision interpretations which are built on them can evolve. It serves the dual interests of efficiency and service to a broad user group, that there are multiple interpretive logics built on the same primary observations. Each is oriented toward a particular purpose.

Following recommendation XV/1 (Indicator framework for the Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets) of the Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA) of the CBD [39], GEO BON is developing the concept of Essential Biodiversity Variables (EBVs) as the anchor of the primary observation level. This is analogous to the Essential Climate Variables (ECVs), required to support the work of the UN Framework Convention on Climate Change (UNFCCC) and the Intergovernmental Panel on Climate Change (IPCC), that have served the Global Climate Observing System (GCOS) well. The number of biodiversity indicators being suggested in a recent report [40] by the CBD’s Ad Hoc Technical Expert Group on Indicators and the GEO BON Adequacy report [11] is well over one hundred. GEO BON’s first attempt at a list of EBVs that could address these indicators, however, suggests a much smaller number of EBVs (a couple dozen) necessary to derive operational indicators for the Aichi targets. Selecting EBVs involves posing the question: what minimal set of repeatable and broadly applicable observations is needed in order to serve the majority of present and future user needs? Thus, defining a set of EBVs could enable the establishment of priorities for biodiversity monitoring, resulting in strong participation from national parties to the CBD in assessing progress toward the 2020 targets at different scales, from the subnational to the global. GEO BON has adopted the development of EBVs as a high priority for the year 2012. The EBV development process will be moved forward through a series of workshops and reviews leading to a publication. Key stakeholders in the data provider, data user and research communities will be involved. The timing meshes well with efforts of the Committee on Earth Observation Satellites (CEOS) and its space agency members to define the remote sensing needs of the biodiversity community for the next decade.

Adequately and operationally defining user needs has long been a sticking point in this and other observational domains. When the researchers and data providers do so on behalf of users, the result unsurprisingly ends up looking rather like what people are doing anyway. On the other hand, if an unfocussed approach is taken to soliciting needs from the users, the result is usually a very long wish-list, some of which is unachievable, and which does not fully exploit new areas of technological possibility. Clearly, what is needed is a ‘co-development’ process where both sides are involved, and providing feedback to each other; that is, stakeholders are providing information on their specific needs for resource management, and researchers are providing information on the types of data that can be presented to meet the stakeholders’ needs. The indicator development process of the UN CBD for the 2020 targets has provided one such user engagement process. Other such processes include the Ramsar Convention on Wetlands of International Importance, which approached the GEO BON Freshwater ecosystems working group to take a leading role in the current initiative to develop a Global Wetland Observing System (a coordinated effort integrating realms, and involving a number of organisations, GEO BON working groups and GEO SBAs, all striving to develop a global system for detecting changes in the world’s wetlands), and the newly formed IPBES which should provide a broad and ongoing interface mechanism between the research, observation and policy-making domains.

A new way of organising?
The management textbooks you find on airport bookshelves tend to assume that organisations are hierarchical, with control and resources flowing from the top. That is not at all like the environment in which GEO BON and many similar organisations find themselves. GEO BON is a voluntary community of practice, a network-of-networks. The resources held at the centre are minimal—enough only for an executive officer, communications and occasional meetings. Most of the work is done in a distributed fashion, by partners in the network, on their own resources and in fulfilment of their own mandates. What keeps such a loose structure from disintegrating? Is it sustainable, and can it be effective? Is there a need for some parts to be centralised? The key incentive for organisations to belong to such a network, and align part of their effort with its workplan, is that doing so allows them to meet their own objectives more efficiently, or in ways which otherwise would not be possible. Alignment is by persuasion and agreement, rather than coercion. The working process involves the formation and functioning of working groups which have been critical in providing the depth and substance needed in GEO BON’s detailed implementation plan [19]. The memberships of the working groups and their networks also play key roles in engaging a wide range of implementation partners and in initiating programs. The working groups are largely self-organising once they have been
Building a global observing system for biodiversity

Scholes et al.

Catalysed, and persist only as long as they are seen as useful by their members.

One of the powerful drivers of the GEO BON strategy has been the spontaneous emergence of national and regional BONs. They have self-organised at national level (e.g. J-BON in Japan and ECOSCOPE in France), at regional, political community level (EBONE in the European Union [41] and Asia-Pacific BON), and around biophysical regions such as the Arctic. These structures provide a framework for achieving the combination of global reach but local relevance that GEO BON aspires to. The key strategic issue is striking an appropriate balance between regional autonomy and collective integration. The principle mechanism for the latter seems to be the establishment of practical standards, supported by targeted capacity building.

Conclusions

The framework for achieving an unprecedented level of availability and integration of biodiversity-related data is now beginning to be put in place. Many of the major players in this domain are collaborating through, among others, the mechanism provided by the GEO BON, which focusses on design, coordination, advocacy for gap-filling and continuity and interoperability standards. There are emerging innovations, such as the notion of EBVs and thematic global observing systems, which will help to create an efficient and sufficient system.

Acknowledgements

Some of the research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References and recommended reading

Papers of particular interest published within the period of review have been highlighted as:

- of special interest
- of outstanding interest


Briefly reviews the situation at the time and presents pragmatic suggestions to improve it.


Outlines a conceptual framework for combining different types of biodiversity data into an integrated system.


Documents the inadequacy of current observational data for tracking the 2010 targets.


Recent review of the state of biodiversity monitoring at global scale.


23. GBIF: GBIF Position Paper on Future Directions and Recommendations for Enhancing Fitness-for-Use Across the
8 Open issue


