CrossMark

From ecological records to big data: the invention of global biodiversity

Vincent Devictor[1,2] · Bernadette Bensaude-Vincent[1]

Abstract This paper is a critical assessment of the epistemological impact of the systematic quantification of nature with the accumulation of big datasets on the practice and orientation of ecological science. We examine the contents of big databases and argue that it is not just accumulated information; records are translated into digital data in a process that changes their meanings. In order to better understand what is at stake in the 'datafication' process, we explore the context for the emergence and quantification of biodiversity in the 1980s, along with the concept of the global environment. In tracing the origin and development of the global biodiversity information facility (GBIF) we describe big data biodiversity projects as a techno-political construction dedicated to monitoring a new object: the global diversity. We argue that, biodiversity big data became a powerful driver behind the invention of the concept of the global environment, and a way to embed ecological science in the political agenda.

Keywords Big data · Biodiversity · Ecology · Foucault · Politics

1 Introduction

The term "big data" has become widely used to describe large datasets in the social, political, commercial, and scientific domains, resulting from the development of computer networks and the Internet. The rhetoric is particularly fervent in the field

✉ Vincent Devictor
  vincent.devictor@univ-montp2.fr

[1] CETCOPRA (Centre d'Etudes des Techniques, des Connaissances et des Pratiques), Université Paris 1 Panthèon Sorbonne, Paris, France

[2] Institut des Sciences de l'Evolution, Université Montpellier, CNRS, IRD, Place Eugène Bataillon, 34095 Montpellier, France

🙋 Springer

of life science. In ecology, big data was welcomed and embraced all possible biodiversity data (Edwards 2000). More recently, the fields of "bioinformatics" and "biodiversity informatics" have emerged, and address the storage and processing of large databases in the life sciences (Michener and Jones 2012). A new community of practitioners, called "eco-informaticians", is actively involved in environmental science (Sarkar 2009) and has developed specific technical skills in the handling and analysis of mega-databases, although sometimes with no knowledge of ecological theory, experimental methods, or experience in field studies (Soberón and Peterson 2004). Some authors have gone further, and announced a "silent revolution", driven by the field of biodiversity informatics (Bisby 2000, p. 2309).

Historians of science have already analyzed a number of epistemic issues raised by the introduction of big data into ecology. While Aronova et al. (2010) argue that massive and systematic data accumulation resulted from the promotion of "big science" in ecology after World War II, Strasser (2012) questions the novelty of the big data project as data accumulation has been central to the production of knowledge in the life sciences since the Renaissance. Callebaut (2012) emphasizes that the data accumulated in large datasets are not theory-free "matters of fact"; rather they are theory-laden and should be considered as the output of an inter-subjective network that addresses biological complexity and multiscale modeling. Accordingly, Callebaut argues that "scientific perspectivism" provides a suit-able philosophical framework to understand the consequences of big data for the life sciences. In addition, Hallam Steven's recent work on the data-driven history of bioinformatics suggests that this new form of knowledge production supports the claim that "biology has adapted itself to the computer, not the computer to biology" (Stevens 2013, p. 39).

Among the societal impacts of big biodiversity data Anna Lawrence (2006) claims that many data collectors (notably in so-called "participatory science" projects) are not driven by the conviction that biodiversity is in itself an object worth of being investigated, measured and classified. Instead, data collection is meaningful for amateur scientists as an expression of their attachment to a locality. Their sense of place, however, disappears in big biodiversity databases since globalization blurs local particularities. Ellis et al. (2007) also describe the social life of biodiversity software as creating a tension between the will to establish strong connections with policy makers and the possibility to reinforce individual relations and the network of civic associations.

With the exception of Geoffrey Bowker (2000a), most epistemological studies of big data in the domain of biodiversity have focused on its accumulation, rather than on the transformation of ecological records into digital data. Bowker points out that big data projects dedicated to the description of biodiversity create a systematic disarticulation between data accumulation and knowledge production, while the database becomes the end product (Bowker 2000a, p. 644).

In the second section of this paper, we use the same line of argument to shift attention from the accumulation of information to the actual practice of data production, and unpack the process whereby records are transformed into data. Furthermore, we clarify how the intensive work of standardization, quantification, abstraction, maximization, and visualization results in knowledge gains and losses.

What we call here the "datafication" of biodiversity is much more than just information simplification. Rather, it corresponds to a shift in priorities in the ecological sciences—from concerns about localities and interaction milieu—to a focus on the emerging concept of global biodiversity.

To fully understand the datafication of biodiversity, one should also question the motivations of its promoters. In the third section of this paper we will consider how the aspiration to quantify the global biodiversity became so important that it radically changed the epistemological profile of ecological science. A historical perspective, focused on the creation of the global biodiversity information facility (GBIF) in the 1980s, sheds light on the decisive role of a new category of actors: policy advisers, who deeply transformed research orientations and the daily practices of environmental scientists. We argue that the concept of global biodiversity is a techno-political tool partly shaped by the GBIF and similar initiatives looking for aligning life science with a model of big science.

Datafication prompted a new knowledge production regime in ecological science focused on the creation of specific "indicators" that could be used to make "rational interventions". In the final section, we discuss the implications of this technoscientific regime aimed at monitoring global biodiversity. While endorsing Sabina Leonelli's suggestion that the gains and losses of data accumulation should be carefully examined (Leonelli 2014) we also critically assess how big data undermines established standards in both life and social sciences (Kitchin 2014).

To clarify our line of argument, let us present a few definitions. First, we consider the impact of big data on practice and knowledge building in scientific ecology. We refer to ecology as the science of interactions among living entities, and between living entities and their environments. Ecology is thus a specific branch of the broader field of life science, which integrates several other fields dedicated to the study of the environment (e.g. biology, zoology, oceanology, geology).

Biodiversity is a very broad term referring to the variability of life forms on earth. It is usually divided into various categories such as species diversity (defined as the number and abundance of different species that occupy a location), genetic diversity (the amount of variation in genetic material within a species or within a population) and ecological diversity (variation in ecosystems at various scales, including the complexity and variability of ecological processes). Of these various categorizations of biodiversity, species diversity is the most commonly-used descriptor, and is measured as the spatial and temporal distribution of individual species. Understanding the diversity and composition of species assemblages in space and time is also the central objective of ecology. We therefore use "biodiversity big data" to refer to the intensive data accumulation of digitalized information on biodiversity, corresponding to a spatial and temporal description of species distribution.

A straightforward definition of the phrase "big data" is difficult to find. It is part of the vocabulary in fields such as computer science, physics, economics, mathematics, political science, bioinformatics, sociology, public health, and agronomy. For our purpose, however, it is important to note that the term emerged in the language of e-commerce when private companies were searching for new ways to develop and control large quantities of data, principally to improve

performance (McAfee and Brynjolfsson 2012). A common definition of big data comes from the report of a leading information technology company on the data management challenges faced by e-commerce. The challenges are summarized by the three Vs: Volume, Velocity and Variety (Laney 2001). The three Vs have been adopted as a way to define big data, sometimes with the addition of two other Vs: Variability and Veracity, together with Complexity. More than a purely quantitative issue, big data is also often considered as the result of the interplay between (i) the maximization of computer power, (ii) the collection, analysis, and linking of large datasets, (iii) the identification of relevant economic, social, technical, and legal patterns, and (iv) the assumption that these data can generate new types of true, accurate, and objective knowledge (Boyd and Crawford 2012, p. 663). Therefore, while there is no a priori definition of big data tailored to biodiversity, this paper characterizes it as techno-political tool to manage the distribution of biological species.

Finally, we characterize technoscience as a regime of knowledge production whereby the idea of pure and autonomous scientific research is replaced by a policy oriented science which is shaped by economic, societal and technological promises (Bensaude-Vincent 2009; Bensaude-Vincent et al. 2011).

2 From records to data

Initiatives to generate, store, and connect biodiversity databases have proliferated in past decades. For example, the National Biodiversity Network Gateway (http://data.nbn.org.uk/) is an organization that collects, sorts, analyses, and disseminates data for biodiversity in the United Kingdom. In Europe, the Biodiversity Data Centre, (http://www.eea.europa.eu/themes/biodiversity/dc) has been established with a similar purpose, together with (for example) national initiatives in Ireland and Belgium (see http://www.biodiversityireland.ie/ and http://data.biodiversity.be/). More ambitious international infrastructures seek to link databases across countries and continents. The Global Biodiversity Information Facility (GBIF) is an international, open data infrastructure for aggregating local and disparate sources of information on biodiversity. It "allows anyone, anywhere to access data about all types of life on Earth, shared across national boundaries via the Internet" (http://www.gbif.org/). The techniques associated with biodiversity databases and their applications are also developing rapidly (Curry and Humphries 2007), and a scientific journal, the *Biodiversity Data Journal*, was created in November 2012 to "make small data big" and accelerate the publication, dissemination, merging, and sharing of all forms of biodiversity data (http://biodiversitydatajournal.com/).

In principle, all of these initiatives share the same objectives: the accumulation of very different sources of data, their transparency, and their dissemination. Most are collections of species occurrences (a species presence at a given point in space and time) or characteristics. For instance it is possible to establish, from among about 570,000,000 records stored in the GBIF, that a blackbird (*Turdus merula*) was observed in Finland at Latitude 60°50′N and Longitude 24°80′E in January 2015, or that perennial rye-grass (*Lolium perenne*) was observed in Australia (at 34°89′S/

138°73′E) in February 2014 (accessed in September 2015). Furthermore, the database shows that these two occurrences were drawn from two different sources, respectively BirdLife, an ornithological non-governmental organization (NGO), and the Department of Environment, Water and Natural Resources of South Australia. In juxtaposing individual data from disparate sources into one or few data systems, such infrastructures introduce radically new information into ecological science. What exactly is their epistemological impact?

Previous studies of this question have tended to highlight the "maddening difficulty in knowing what is where, or of comparing like with like" when trying to build a global biodiversity information system (Bisby 2000, p. 2309). Such studies clearly demonstrate that data classification and integration is much more than just a process of quantification; it requires an abstract concept of biodiversity measurement. Some authors even claim that the way biodiversity is measured should be adapted in order to align data collection with progress in biodiversity informatics (Deans et al. 2012). Indeed, measuring biodiversity requires an intense effort to homogenize disparate sources. Bowker (2000a) convincingly demonstrated that the production of big data related to biodiversity presents a challenge to classification systems and the categories used to describe diversity in living forms. He argued that not only are entities (e.g. species, landscapes, communities) difficult to classify (to the point that they are excluded from the final database), but also that it is difficult to retain the context of data collection (e.g. measurement techniques or data localization).

Bowker also noted that aggregating or extracting biodiversity data tends to obfuscate the specificity of scientific disciplines with respect to data collection (e.g. genetics, ecology, paleontology). For instance, two communities (ecology and biogeography) that work at very different scales can upload their data to the same database, despite the fact that the spatial resolution of interest varies from 1 m^2 to the entire globe. He thus concluded that the uniformity of databases concealed a diversity of spatial and temporal framings (Bowker 2000a, p. 668). On the other hand, (Turnhout and Boonman-berson 2011) argued that scaling techniques made it possible to select the data best able to reliably represent biodiversity at a large scale, rather than providing an objective description of ecological phenomenon.

The process of combining multiple data from disparate sources into unified and uniform databases makes the accumulation of biodiversity big data possible. But it requires ontological adjustments to harmonize heterogeneous information into a uniform data infrastructure. In order to explore in more detail the consequences of the transformation of singular and heterogeneous records into aggregated and unified biodiversity databases, we further consider how the juxtaposition of heterogeneous records alters the information itself (what), the motives for collection (why), and its possible usage (how).

The first point is that information (the 'what') deposited in integrative biodiversity databases loses part of its ecological meaning. "Records" are converted into "data". Records are defined as the spatial and temporal delineation of biophysical realities motivated by specific objectives. Data are collections of individual pieces of quantitative or qualitative information. This distinction is particularly straightforward for the life sciences and even more specifically for

ecology. Indeed, early ecologists paid great attention to the collection of meaningful information about the natural world. Charles Elton, in particular, dedicated a book to it (Elton 1966). According to Elton, relevant ecological information is embodied exclusively in individual "records" rather than datasets. Building ecological records is an active process of collecting information not only *from* but also *in* the field. In the late 19th century, botanists promoted ecology as a scientific discipline in order to distinguish their interest in plant adaptation from the narrower physiological studies conducted in the laboratory, and to go beyond descriptive inventories and museum-based taxonomic work. Ecology was a call to develop a new science devoted to investigating the relations between organisms and their environment through experiment and field observations. It was intended to replace the "intellectual barren practice of collecting plants without paying attention to their surroundings". In this context "the word *survey* implied just this modern approach: a rounded analysis that included an understanding of the life history of the plant and its ecological context" (Kingsland 2005, p. 69).

The holistic approach to recording interacting objects is thus crucial for ecological science, which is primarily concerned with the interrelations between living organisms and their environment. As George Clarke pointed out in one of the first academic textbooks dedicated to modern ecology, ecology must be viewed as the study of interrelations of plants and animals with their environment, which include "the influences of other plants and animals present as well as those of the physical features" (Clarke 1954, p. 2). Ecology is thus devoted to the description and understanding of the interaction milieu, i.e. how local natural communities are shaped by the relationships between biotic and abiotic elements.

Furthermore, ecology is primarily concerned with whether and how species interact between them and with their environment, as these inter-relations are key components of the definition of the concept of the ecological "niche". This concept was adopted early in the development of the discipline to describe either species' requirements or their impacts in a given environment. For instance, Elton defined the niche (sometimes called the functional or trophic-niche) as the impact of species in the environment rather than on its response to particular resources (Elton 1927). Joseph Grinnell, another influential ecologist contemporary to Elton describes the niche as the response of species to a given set of resources (Grinnell 1917). The Grinnellian niche was extended by Hutchinson (1957), who considered all biotic and abiotic resources and added a distinction between the fundamental and realized niche. The fundamental niche is considered to be the niche that results from the intrinsic attributes of species, while the realized niche is a contingent property dependent on each local environment. These distinctions have been—and are still— central in almost all aspects of ecological and biogeography thinking (Chase and Leibold 2003, p. 2).

In practice, focusing on a particular aspect of the niche largely determines what and how ecological records are constructed (Devictor et al. 2010). Measuring the Eltonian niche of a given species depends on the ability to quantify the role of the species with respect to others (e.g. by quantifying trophic interactions), while measuring the Grinnelian niche necessitates a fine-grained description of each species' resource use. The distinction between the fundamental and realized niche

also necessitates measuring what interactions are determined by species' integral attributes (e.g. their optimal diet when all resources are available) or when influenced by competition due to the presence of other species. Regardless of which view of the niche is adopted, a record is typically intended to keep track of the ecological context surrounding organisms observed in a given space and time. A record is produced by an "ecological survey", which Elton considered "by no means just a matter of registration and enumeration. It is not a static task but an exciting study of processes in nature" (Elton 1966, p. 25). In other words, a record provides not only a rich description of the living form of interest (e.g. a plant occurring in a given place and time) but also information about its biotic and abiotic surroundings and/or the process of recording itself (the name of the field observer, the duration of observation, etc.).

When data from different surveys are aggregated, the ecological perspective of each survey is reduced to two-dimensional charts made up of rows and columns. This way to organize data assumes that a number of individual objects (rows) with certain characteristics (columns) can be juxtaposed in the same database. The database structure is dictated by checklists of minimum criteria that are endorsed by a community interested in information exchange, easy access, automated uploading, and interoperability. Agreed rules are adopted to describe, format, submit, and exchange data. Typically, data frames combining two different datasets require that one dataset is compatible with the other. In practice, new rows can be added only if existing columns are also meaningful for those rows, and for this purpose share specific attributes. This is how the blackbird can be found alongside the rye-grass in the same database, although they were recorded with different methods at different times, in different countries, by different people with quite different objectives.

Decontextualization converts the ecological niche into a "data niche". A given record has explicit and specific spatial and temporal scales (e.g. the name of a particular locality corresponding to a given spatial resolution sampled for a given purpose). On the other hand, in biodiversity big data each unit of data can be located in space and time using common and universal references (e.g. its longitude and latitude in a standard projection system). As Griesemer (2011) argue, "exogenous" concepts of time and space are added to the data to make it fit the specific format required by databases. But when applied to ecological records, exogenous space and time become detached from a specific survey and a particular view of the ecological niche. The recontextualization in exogenous space and time frames is thus necessary to ensure that data coming from different surveys can be grouped into a uniform dataset.

These artificial space and time frameworks are generated by the actual construction of data, which is an endless task. As Bowker noted, big data only covers a "thin slice of species and environments" (Bowker 2000a, p. 645). A large number of specimens stored in museum collections around the world await processing and have not yet been taken into account. This is also true for many records accumulated in tables published by the ecological literature. This means that exogenous space and time frameworks are open-ended and cannot really "frame" the biodiversity map. An additional problem raised by the aggregation of heterogeneous records is that the loss of information regarding sampling for each

individual dataset introduces significant biases in the database (Beck et al. 2014). Not only is the slice thin, it is also distorted as digital constraints have removed the ecological valence of information. In this process the various features of the ecological niche (Eltonian, Grinnellian, realized or fundamental) disappear. This loss has practical consequences when data are used in a geographic information system (GIS) to draw biodiversity maps. Like biodiversity objects, environmental variables are dynamic and cannot easily be fitted into quantitative or taxonomic categories (Bowker 2000b). Biodiversity data therefore become detached from relevant and reliable habitat characteristics. As a result, there is no "interaction" in the interaction milieu, and the resulting abstract entities can hardly be called "milieu" as they are in the middle of nowhere.

Turning our attention to *why* information is being collected and *how* it can be exploited reinforces the impression of big losses caused by biodiversity big data projects. As noted by Stevens (2013, p. 66), investigating life through unified databases no longer means testing hypotheses or theories by means of observations and experiments. New knowledge about life is supposed to emerge from the data that has been accumulated. However, knowledge generated by the accumulation of data is by no means theory-free. It is shaped and formatted by the constraints of simplification, standardization, and interoperability all of which depend on implicit theoretical assumptions. In ecology, a field survey is motivated by explicit objectives, which mirror specific hypothesis or applied perspectives about the specific records of interest. In particular, one research question at the core of any ecological survey concerns the identity of ecological units. Different approaches to the definition of ecosystems, community, or populations rely on different theoretical backgrounds, which influence survey design and record collection (Jax et al. 1998). For instance, the historical debate between Frederic Clements (1874–1945) and Henry A. Gleason (1882–1975), or the more contemporary debate between the influence of the niche and neutral processes in shaping natural communities, revolve around how individual species dynamics contribute to forming species assemblages. Clements took a more deterministic position and suggested that a particular community would develop in a given environment, while Gleason considered that stochastic processes mainly influenced species associations. More recent debates on the topic focus on the relative importance of environmental filtering and species interactions (niche processes), and random fluctuation in very basic processes such as dispersion, and random extinction or colonization events (neutral processes) (Chase and Myers 2011).

Such theoretical backgrounds are also reshaped when records are transformed into data. For instance, among the most advanced methods of processing environmental data coming from different sources, species distribution modeling (SDM) is based on the ability to make new, *ad hoc* assumptions about the relationships between species and their environment. SDM estimates the geographic distribution of species from the combination of occurrence data and environmental variables. For a given species, it relates sites of known occurrence (or absence) of the species with known predictor variables for these sites and all other sites, in order to produce quantitative estimates of the spatial distribution of the species. This method is widely applied to predict shifts in species ranges following climate

change (Guisan and Thuiller 2005) and is considered a major toolkit for the analysis of GBIF data (SDM is promoted as a key resource on the GBIF website, http://www.gbif.org/resource/81009). But SDM makes many assumptions that are imposed by the new ontology of the data niche. For instance, the influence of biotic interactions and physiological constraints are assumed to be constant in space and time. The specific view of the "niche" adopted in SDM is thus unclear and subject to debate (Jiménez-Valverde et al. 2008); furthermore a number of ecologists have criticized its lack of explicit theoretical grounding (Elith and Leathwick 2009).

In fact, the conversion of records into data is more than a loss of information. It also partly blurs the theoretical background, while data accumulation tends to become an end in itself rather than a means to an end, to the point where the scientific purpose of data accumulation becomes unclear. Data is accumulated on shelves and waits to be used. For example, the GBIF recently launched "The Ebbe Nielsen Challenge", designed to "inspire scientists, informaticians, data modelers, cartographers and other experts to create innovative applications of open-access biodiversity data" (http://gbif.challengepost.com/). This open competition, with a first prize of €20,000, is an a posteriori quest for relevant applications of data processing. It was created to:

> encourage innovative uses of the more than half a billion species occurrence records mobilized through GBIF's international network. These creative applications of GBIF-mediated data may come in a wide variety of forms and formats—new analytical research, richer policy-relevant visualizations, web and mobile applications, improvements to processes around data digitization, quality and access, or something else entirely. Judges will evaluate submissions on their innovation, functionality and applicability.

Remarkably the list of envisaged applications of GBIF mentions neither tests of scientific hypotheses on the spatial distribution of species, nor to find solution to environmental problems, nor even to provide a precise ecological survey. It invites responses that prove the "innovation, functionality and applicability" of big data. In order to match the specific standards imposed by big databases, data have been decontextualized to the point that they have lost most of their relevance from an ecological perspective. Once again, this has direct consequences when mapping species distribution at large spatial scales. In their comparison of GBIF maps and verified maps of species richness for the American tropics, Maldonado et al. (2015) remark that erroneous GBIF georeferences could generate faulty maps. Although they emphasize that the "uniformity" of big data is an advantage, they also acknowledge that a small error in the georeference of a given specimen "could imply wrong interpretation of the results. This would not only affect studies aiming at identifying general biodiversity patterns, but also others such as reconstructing the environmental niche of species and predicting their total distribution" (Maldonado et al. 2015). They conclude that such lack of accuracy compromises our knowledge of biodiversity and distribution patterns.

The Ebbe Nielsen challenge reveals that, contrary to ecological records collected according to a priori research questions the accumulation of data is considered an end in itself. The cognitive value of biodiversity data is expected to emerge a

posteriori because it is assumed that the accumulated data will generate new research questions. As we argue in the next section the most suitable methods and results to make the data meaningful and useful are determined from a governance perspective. In this respect, big biodiversity data are accumulating matters of (potential) interest, rather than raw matters of fact.

In summary, this section emphasizes the stark contrasts between the data niche and the ecological niche. First, the big biodiversity databases aggregate records derived from quite diverse scientific programs with specific purposes (e.g., describing species distributions, the monitoring of temporal trends, the listing of endangered species, the test of scientific hypothesis, the interest of people for their biological surrounding). The initial drivers for recording bits and pieces of ecological information are diluted, if not totally lost, in the creation of a global data niche. And the process of globalization is supposed to be the main source of knowledge gain.

Second, the data niche filters out part of the ecological niche. Most ecological records contain more detailed information than what is required by a big biodiversity database. Due to the minimum data format that can be uploaded the data niche can deliver only basic information (a species name and a country) although the initial ecological records are often coupled to rich metadata (e.g. the sampling conditions, environmental variables co-occurring species, animal behaviour).

Third, a most striking difference between the ecological and data niches concerns their respective space and time frames. Ecological records are linked to a biological space in which species are either attached to environmental characteristics or to other species. The biophysical space is an essential component in all definitions of the ecological niche (Devictor et al. 2010). In the data niche most, if not all, traces of this biophysical environment are swiped away as it is delineated by a country name or universal coordinates. The time frames are even more dissimilar. Many ecological records proceed from routine periodical surveys which make sense because they are conducted with the same protocol and the same method. When absorbed in the data niche this temporal continuity is broken. The format of the infrastructure used to construct the concept of global diversity means that data are likely to be locked into the traditional, pre-ecological ontology used by the "collectors of the past" (Kingsland 2005).

Epistemological studies of data accumulation as a method of knowledge production in life science in general have noticed how quickly big data have been incorporated in daily practices of biological research. For instance, bioinformatics generated specific standards to ensure the circulation and retrieval of facts among users. As Sabina Leonelli argues, online databases are precisely meant to reach a sharp equilibrium between de-contextualizing local specificities while allowing their re-contextualization elsewhere and for new purposes (Leonelli 2011). Hallam Steven claims that bioinformatics modified not only the scope of biological questions investigated by biologists but their practices and careers as well. Biologists are indeed increasingly called to justify their skills dedicated to bioinformatics, data collection, handling and interpretation launching new battles for credits and jobs (Stevens 2013, p. 43). Regarding climate change, Paul Edwards

stresses that the idea of global climate change has been materialized in graphics, maps, and earth characteristics calculated globally and that there is a continuous effort to standardize data collection and communication (Edwards 2010, p. 251). We argue that the case of big biodiversity data is special. Although the quest for globalization is similar to the case of climate change, the integration of big data in the corpus of ecological knowledge seems more difficult. There is so far a divorce between the process of data accumulation and the basic concepts of ecological science (such as the notion of ecological niche). The specific interest of ecology regarding the study of interactions among organisms and their environment is also buried in the quantity of non-interacting data.

Obviously, not all of ecological science has been altered by the process of datafication. We rather claim that merging and unifying multiple records in global datasets have contributed to the up scaling of specific concerns, entities or processes. For instance, this abstraction allowed the calculation of a living planet index (Loh et al. 2005), summarizing, in one estimate, the global temporal trend of the world's vertebrates. Along with this abstract notion, the state and fate of individual populations or species recede into the background and leave the floor to the image of a living planet. Similarly, in ecosystem ecology the study of local ecosystem dynamics was extended to the potential drivers of planetary-scale transitions and lead to a call for "guiding the biotic future", conceived as a "vital task if the goal of science and society is to steer the biosphere towards conditions we desire, rather than those that are thrust upon us unwittingly" (Barnosky et al. 2012). The creation, unification and interoperability of big datasets have thus contributed to the emergence of globalized ecological entities and concerns.

Overall, transforming heterogeneous records into uniform and commensurable data requires several operations—such as quantification, standardization, and interoperability—that deprive them of substantial ecological meaning and objectives. Ecological records and biodiversity big data are ontologically different. For the data infrastructure to be operational, a data niche, defined in terms of computer language, replaces the ecological niche. Local and heterogeneous living forms are the victims sacrificed to the grandiose ideal of global biodiversity that we further characterize below.

3 The invention of global biodiversity

The previous section stressed the ontological loss generated by the process of datafication. On the other hand there is a gain, or the expectation of a gain, that drives the process. In this section, we argue that the major gain is the invention of the concept of global biodiversity itself. From a traditional ecological perspective, the idea of global biodiversity is not really straightforward. Each individual, and each species, experiences the environment on a unique range of scales. Rather than global mapping and numbers, ecology is more interested in how heterogeneous patterns are distributed across multiple spatial scales. As Levin (1992, p. 1959) puts it, "no description of the variability and predictability of the environment makes

sense without reference to the particular range of scales that are relevant to the organisms or process being examined".

However, the notion of global biodiversity—and more broadly that of the global environment—is more than a scientific output; it was crafted through megascience projects aimed at providing a global survey, a panopticon of biodiversity, before eventually becoming one of the United Nations' Development Programs. How and why did a global approach to plant and animal species become a priority? The approach is partly the result of the extension of the Big Science model (initiated by the Manhattan project) to biology. More precisely, as Aranova et al. (2010) argue, the International Biological Program (1964–1974) was inspired by the data-driven style of research used in geophysics. Data collection was prompted by the ambition to survey life phenomena on a transnational scale. Such initiatives viewed the globe as a single system (Kwa 2005; Bocking 2013).

A historical view of the emergence of global mapping projects (such as the GBIF) suggests that the concept of global biodiversity is, above all, a policy-making platform. The GBIF data infrastructure is designed to coordinate and aggregate species occurrence data and digitize natural history collections into a single, global-scale resource. Although there are other similar initiatives, the GBIF is a particularly useful example of how science and politics have interacted to promote the notion of a global environment. Slota and Bowker convincingly argue that the GBIF's global data portal "renders amenable not the species occurrence data itself but the concept of global biodiversity mapping, the usefulness of having a global-scope resource, and the value of working globally on global systems" (Slota and Bowker 2015, p. 2). In other words, the GBIF mainly served to experiment and promote global mapping as a legitimate research style. But what are the origins of the relevance of such an approach?

In tracing the origins of the GBIF, immediately a crowd of heterogeneous actors – such as the Council of the Organization for Economic Co-operation and Development (OECD)—take center stage. The GBIF emerged from the Mega-science Forum, which was established on 1 June 1992 by the OECD, as a subsidiary body of the committee for science and technological policy (CSTP). The aim of this Forum was to bring together well-known scientists and members of OECD governments to enable the exchange of information and develop large-scale research projects. The Forum was created following a proposal by David Allan Bromley (1926–2005), an American–Canadian physicist who chaired the United States Office of Science and Technology Policy (OSTP) and was Scientific Advisor to President George H. W. Bush from 1989 to 1993. The OSTP was itself established in 1976 by Congress to connect policymaking and politics with the most advanced science.

Bromley was the first policy advisor to hold the title of "Assistant to the President for Science and Technology". As a government consultant in the Nixon, Reagan and George H. W. Bush administrations, Bromley was a well-documented figure and an entire biographical memoir was dedicated to him in which George H. W. Bush personally underlined how Bromley understood his position:

to give the President the best objective advice on any policy matter that relates to science and technology (e.g., global warming and climate change), summarize the state of scientific understanding—including all the uncertainties, and when appropriate recommend policy options (Greiner and Lane 2009, p. 31).

Bromley was hired to prepare policy about emerging global problems. For him, addressing new global challenges required recognizing the mutual dependence of science, technology, and politics (Bromley 2002, p. 18). He advocated that solutions to issues of global change would "necessarily involve a complex mixture of science, of technology, and of politics, and for this to happen we simply must work together much more than we ever have in the past". This reflected his view that science and politics were, and should be, intimately interconnected, at least in support of big projects. Rather than emphasizing the specific historical role played by Allan Bromley and subsequent scientific advisors to the President (see Pielke and Klein 2010), here we trace the priority given to global mapping and data accumulation during Bromley's tenure in order to better understand the conditions and circumstances that changed the daily work of ecologists and turned them into environmental scientists.

This change can be traced to the proceedings of the United States Global Change Program developed by the Committee on Earth and Environmental Sciences that was established during Bromley's term as Scientific Advisor. The aim of this Committee was to put in place "the science of global change—to make the observations, develop the understanding, and form a predictive understanding of global change—and then hand that information to policy makers" (Kelmelis and Snow 1991). In the Committee's report, Bromley emphasized the importance that "the White House places on data management" and insisted on the role of global mapping. He argued that it made the data intelligible and valuable because "when the need arises, we'll be in great shape to digest, analyze and present these huge amounts of data in fashions that use the remarkable ability of the human to sense patterns and to develop hypotheses based on graphical information". But beyond inspiring scientific research, global mapping was intended to contribute to what Bromley call "the most important thing", i.e., "the translation of good science, good technology, and good economics into good policy" (Kelmelis and Snow 1991, p. 42).

Consequently, the reports of the OECD Megascience Forum aimed to develop policy recommendations, and this task included facilitating information exchange between governments and scientific communities involved in large-scale projects (OECD 1993, p. 3). However, the role of data collection and biodiversity in the interface between scientists and policy makers was far from clear at the time. Information that is collected by a variety of local scientific projects funded by all sorts of sponsors was not manageable. It gradually became clear that collection without datafication was useless. The formatting, storage, diffusion, and processing of data thus became the priority, and was viewed as the key to success in developing and coordinating the construction of a relevant interface for policy makers (OECD 1993, p. 30).

The Forum proposed an international collaboration to facilitate data exchange and standardization. It advocated the creation of a centralized management platform available to policymakers and scientists worldwide. Data would thus form the basis of scientific megaprojects:

> Data are the glue that holds distributed megaprojects together, and an ever-increasing component of such efforts, sometimes measured in terabytes. Effective data collection, management, evaluation and distribution are crucial to the success of such megaprojects. Challenges for the future include standardization and quality control for the data collected and processed, and assurance of access for qualified researchers worldwide (Ratchford and Colombo 1996, p. 216).

Although this quote suggests that megascience became an end in itself, the benefits of massive data collection had to be explained. Taking the example of bioinformatics, the report anticipated that the project would enable useful comparisons and correlations:

> This will be no less true for the rest of biological information; in fact, the broader the range of data types available, the wider the range of applications to which those data may be put. For example, agriculture will benefit from information on the habitat and evolutionary relationships of the wild relatives of crop species; health-related research will benefit from correlations among many types of neurological data (images, numerical results of surveys, etc.); and the pharmaceutical industry will benefit from access to information from biological collections (OECD 1999, p. 7).

This view of data accumulation suggests that what matters for the future data niche (as defined earlier) is its capacity to embrace as many fields as possible: "the broader the range of data types available, the wider the range of applications". Accumulating and coordinating data produced by environmental science was not primarily intended to better understand biological systems or biodiversity, while very large databases are most likely to hold spurious correlations of little interest (Calude and Longo 2015). But according to the Forum, accumulating information was less a way to gain scientific knowledge, than a tool that served economic and political agendas. The data niche was delineated by political demands. The aim was to facilitate applications by providing information about the locations of valuable resources. The Forum made no reference to ecological theory, but instead insisted on the potential benefits of accumulating biodiversity data for practical purposes (OECD 1999, p. 14).

The GBIF project was presented during the 1999 Megascience Forum. Technical aspects of biodiversity informatics and big data were based on established models from genetics and neuroscience as advances in computer technology had demonstrated applications in the storage and processing of data in these disciplines. A model of life science in which all biological entities are commensurable. It is remarkable to note that although the scale, complexity, and dynamics of genetic and ecological phenomena are far from commensurable, the GBIF anticipated a

straightforward alignment of biodiversity objects with genetic data (OECD 1999, p. 14):

> Great forward strides could be made in the understanding of the biological world, for instance, if informatics techniques were developed to make it possible to correlate historical information with newly collected satellite data; if molecular genetic datasets could be linked to species-documentation datasets such as those held by natural history collections; and if neurobiological, physiological, chemical, and other sorts of datasets could be correlated with taxonomic and ecological ones (OECD 1999, p. 7).

This brief historical description of the implementation of the GBIF suggests that globally, the systematic accumulation of data on biodiversity was seen as a tool for policymaking. However, the global environment is by no means a purely political object. It has been co-produced through the close interaction of converging interests (anticipated by Bromley). Scientists have learned to think globally, under the pressure of public environmental awareness that began in the 1970s. When increasing numbers of scientists and people realized the impacts of human activities on the environment, ecologists started to work on a bigger scale than that of the ecological niche. The upscaling of ecological studies has been developed through initiatives such as the International Geosphere-Biosphere Programme (Kwa 2005), when ecologists began to study the globe as a single system. Ecosystem research also played an active role in the construction of the notion of the global environment. As Bocking (2013) put it: "the ecosystem concept was compatible with the widely invoked metaphor of "Spaceship Earth", which combined notions of a finite and fragile global system with confidence in the capacity of science and technology to understand, and manage, this system". The example is similar to that of interactions between science and politics in climate science, where the role of the policy entrepreneur was held by science managers working for the government who had scientific and academic roots (Grundmann and Stehr 2012, p. 121).

Other movements also contributed to build the notion of global environment. The new "regime of ocularity" described by Fernando Elichirigoity emerged in the 1980s with an ensemble of practices revolving around machine vision and system thinking. This globalization was facilitated by satellite imaging and scanning as well as digitized interpretation, computer modelling and simulation. The result of these multiple representations of the global environment is that the monitoring of global biodiversity became a commonly accepted target to which, as we argue in the next section, a techno-political assemblage was progressively dedicated.

4 Monitoring global biodiversity: a techno-political assemblage

The previous section highlighted that the global datafication of biodiversity was the brainchild of policy advisors and research scientists and resulted in the promotion of the global environment as a research object and a matter of concern. But interestingly, the political dimensions of big biodiversity data rest neither on the process of data accumulation nor on the idea of capturing the "diversity" of life.

Indeed, Strasser (2012) has shown how accumulating vast quantities of data has been a common practice in life science from the Renaissance. Müller-Wille (2015) remarked that if collecting biodiversity data simply corresponds to counting and labelling taxonomic units as endangered, widespread, alien, etc. such practices date back to the late eighteenth century. Data visualization techniques such as global mapping are no more novel. In plant geography global mapping was also used to bridge popular and scientific cultures, which incidentally influenced ecology in the late nineteenth century (Güttler 2011). What seems unprecedented, however, is the disappearance of cognitive ambitions driving the process of datafication. The knowledge produced by big biodiversity data is instead expected to emerge, a posteriori, from the data niche itself.

In this section, we argue that more than a motor of knowledge production, the transformation of records into data was the key process that made biodiversity manageable and actionable. Far from being a given that has to be carefully recorded, biodiversity came to be viewed as something that can be monitored, as an object of governance. What was the context for the emergence of the notion of the global environment as an object of governance?

Hamblin (2013) argues that this notion emerged in the aftermath of World War II and was driven by strategic imperatives and military objectives. In line with efforts by the Rand Corporation to anticipate the evolution of political regimes and future wars (Andersson and Rindzevičiūtė 2012), initiatives were launched to anticipate the evolution of biodiversity. Global surveys were intended to monitor both the spatial and temporal distributions of plants and animals. Beyond monitoring, the objective was to facilitate rational choices for action on the basis of sound evidence and measurements.

The creation of the Megascience Forum (1 June 1992) coincided with the United Nations Conference on Environment and Development (UNCED), also known as the Rio Summit (3 June 1992). Rio was the third "Earth Summit", after the Stockholm (1972) and Nairobi (1982) Conferences organized by the United Nations. The Rio Conference raised public awareness of the so-called "biological diversity crisis" and urged political action to be taken to stop its rapid decline. Charismatic writers and scientists, such as Edward O. Wilson, made a significant contribution to reinforcing the link between science and politics through promoting and connecting the concepts of "crisis" (Wilson 1985) and "biodiversity" (Wilson 1988). The phrase "biodiversity" (created from the contraction of "biological diversity") itself was coined during a scientific conference held in Washington which, according to Dan Janzen, a scientist who attended it, turned into a political event intended to alert Congress to the issue and put the environmental crisis on the political agenda in the United States (Takacs 1996, p. 37). At this stage, it was anticipated that biological data would play a significant role. It was even one of the top priorities noted in the conclusion of the Washington conference: "We need to place high priority on facilitating access to the world's conservation data bases stored on magnetic, compact disk/read-only Memory (CD-ROM), or laser disk media. Thought must now be given to management of the transfer, loan or sale of these data bases" (Wilson 1988, p. 293).

At the time of the Rio Conference, however, neither global warming nor the reduction in biodiversity was evidenced by "real data". Early alerts were based on rough estimates of species extinction and rates of deforestation. As Paul Edwards argued in relation to climate change, making data global, i.e., building complete, coherent, and consistent global datasets, was a top-down project (Edwards 2010, p. 281). This model only partly holds for biodiversity because biodiversity data were much more difficult to define. No single parameter (e.g. temperature or precipitation) could embody biological diversity, and a biosphere equivalent to the global perspective constructed from satellites for abiotic phenomena remained to be created. The "data diversity" related to biodiversity had to be thoroughly sorted, organized, and standardized prior to any conclusions and action on a global scale. The definition of a good data niche was a prerequisite for the good governance of global environmental issues.

The Rio Conference also defined the conditions for optimizing the coordination of environmental data accumulation and the development of biodiversity informatics. A convention on biological diversity (CBD) was signed by 192 states. The convention called on the parties to monitor the components of biological diversity through sampling and other techniques, and to maintain and organize the data derived from these techniques. In doing so, biological diversity was sliced into several layers with a particular focus on those of "social, scientific and economic importance" (Art 7). Biological diversity, as a concept, was therefore thoroughly organized and classified into specific and discrete sub-categories that facilitated environmental monitoring on a global scale. Decomposing the natural world into several "biodiversity components" (such as ecosystems and habitats, species and communities, genomes and genes) was a first move in homogenizing heterogeneous databases according to a standard format in order to be able to share both infrastructure and data.

The transformation of environmental big data into this techno-politic infrastructure can be precisely documented through the subsequent annual conferences of the parties (COP) that were responsible for implementing the Rio Convention's program. The archives of these annual meetings (available at http://www.iisd.ca/vol09/) make it possible to reconstruct the process of biodiversity data collection through the interplay of patterns, practices, knowledge, and people.

An initial impact of international concern for biodiversity management was the call to think globally about the diversity of living forms on earth. As a tool for designing an action strategy, the accumulated biodiversity data had to be treated as a collection of discrete units; therefore transparent, quantifiable, and measurable. This was an essential prerequisite for centralized control and coordination. A second major impact was that all research efforts were oriented toward the production of indicators. The agenda that consisted of monitoring a set of "indicators" of global environmental change already existed at the international level—it was a recommendation for action following the Stockholm conference in 1972 (recommendations 4 & 29 of chapter 10 of the conference text). Nevertheless, new indicators of policy-relevant objects for monitoring future biodiversity were adopted by scientists, NGOs and civic associations. Furthermore, the seventh COP proposed a list of biodiversity indicators to be monitored in order to assess whether parties

were able to reduce biodiversity loss. Some were derived from data gathered from many different sources, such as the Living Planet Index (Loh et al. 2005), while others consolidated existing indicators with a strong political valence (e.g. the Red List Index, which summarizes changes in conservation status observed for traditional "red lists" of threatened species).

A more general framework, proposed by the OECD and known as the "Pressure State Response" model (OECD 2003, p. 21) included indicators of biological diversity that took the form of a loop. The starting point is human activities ("Pressure") that exert pressure on the environment, affecting its quality and the quantity of natural resources ("State"), which in turn prompts societal responses to these changes through environmental, economic, and sectorial policies ("Response"). Biodiversity data and indicators therefore became embedded in a strategic model of problem solving, which was based on a heterogeneous network of actors. A paper published in *Science* co-signed by 28 scientists, members of the UNEP and NGOs instantiated the model. It called for the intensification of extensive data collection, classification, and integration. Biodiversity indicators were designed alongside economic indicators, and a consensus emerged regarding the potential to translate data and indicators from big databases in ways that were useful for policymakers (Balmford et al. 2005, p. 212). According to Robert Watson (a policy adviser/scientist to the United States President and co-chair for the Millennium Ecosystem Assessment), scientists "cannot simply talk about monitoring birds and butterflies" but they must rather "link the conservation and sustainable development use of biodiversity to the development issues that policy makers and the majority of the public care about" (Watson 2005). In 2010, it was agreed that biodiversity indicators would be used to inform "world leaders" in order to achieve the goal of reducing the rate of biodiversity loss (Balmford et al. 2005). At the same time the Pressure State Response model proposed by the OECD was adopted by scientists involved in the development of these indicators (Butchart et al. 2010).

We have argued that the accumulation, standardization, and coordination of biodiversity big data extended far beyond the technical issue of adaptation to the constraints of computer storage. The concept of global biodiversity emerged from an interplay of technical aspects and political ambitions of global governance. Nevertheless, it did not just result from a top-down initiative from policy advisors directed at active scientists and civil society. It also resonated with the concerns of people from various backgrounds and with different convictions, and we prefer to argue that it can be seen as an example of the emergence of a "strategy without a strategist", as described by Michel Foucault in his analysis of power.

According to Foucault, at specific historical conjunctures the combination of practices, knowledge, and institutions result in a coherent "strategy". These strategies are not due to any single individual social agent, but rather developed without any "strategist". They consequently form a "thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions" (Foucault 1980, p. 194). Foucault referred to these heterogeneous assemblages of elements that condition, maintain, and enhance a

form of power in the social body as '*dispositifs*' (Foucault 1980). Biodiversity big data is such a *dispositif*; a virtual version of Foucault's famous panopticon.

Although big data does not embrace all species—as highlighted in the first section—it is clear that the ambition is to provide a global view, in other words a view from nowhere. With respect to biodiversity, it tends to see the observer as external to the planet, a perspective that emerged from growing public awareness that human technologies have an impact on the Earth. The realization that we have interfered with nature leads to the construction of a viewpoint located nowhere in particular in order to embrace what is left. Big data becomes a substitute for the Archimedean lever for moving the world, just as nineteenth-century chemists—Louis Pasteur in particular—used the laboratory as a fulcrum.

Is it possible to assess the performance of this *dispositif*? First and foremost it functions as an efficient alert apparatus. In turn, the heterogeneous assemblage creates multiple connections between heterogeneous actors: as much between scientific data related to different species or ecological levels as between institutions, scientists, the media, and policy makers. The technological constraints imposed by the process of data accumulation have favored what Turnhout et al. (2014) called "measurementality" in biodiversity governance (Turnhout et al. 2014).

It is also an educational tool. Once data accumulation had begun, the relevant processes were established between scientists, institutions, governments and NGOs, and the legal framework was in place, a social movement was created for promoting certain human interests and values. Biodiversity big data try to educate people to "think globally" about biodiversity issues and this social apparatus has become a symbol, intended to guarantee public trust in global numbers and phenomena. In raising environmental awareness at the global level, big data could be thought of as a practical tool for global monitoring of the environment. Institutionalization also contributed to conferring new social and political meaning on these data. Biodiversity issues have been endorsed by an Intergovernmental Panel on Biodiversity and Ecosystem Services (IPBES). Here, the objective is to launch a global biodiversity infrastructure capable of assessing, collecting, and summarizing knowledge about biodiversity trends, many of which will be derived from biodiversity big data. Interestingly, institutions such as IPBES reproduce the circularity at play in the wheel of big biodiversity data. As Esther Turnhout et al. (2016) argue: "Knowledge and power embrace tightly as globalized knowledge conditions the political imaginary of global environmental governance and vice versa: how one knows constrains how one governs and how one governs shapes what one needs to know" (Turnhout et al. 2016).

Biodiversity as understood by the IPBES, however, is closely related to the concept of ecosystem services and is therefore much more a reflection of neoliberal environmental governance than ecological knowledge (Turnhout et al. 2014). Yet, recent perspectives of ecosystem ecology, in which ecological units are no longer defined as orderly and stable systems but as non-linear complex phenomena, present an even greater challenge to the relevance of current big data initiatives which provide long lists of non-interacting species. Finally, although the emergence of "environmental monitoring" as a global issue may justify data accumulation, it may

also have unexpected consequences at the local scale. As Aronova (2015) points out, an intergovernmental global monitoring program can be used not only to motivate data collection by volunteers, but also to legitimate a lack of political action to protect local ecosystems.

5 Conclusions

This paper emphasizes the interplay between the ontological and political dimensions of big data projects related to biodiversity. It demonstrates that:

(1) The conversion of records into manageable data for the purpose of biodiversity interventions has a powerful impact on the scope of ecological science. These transformations have contributed to change the episteme of ecological science. When research priorities shifted from the interaction milieu to global diversity, ecological science became a kind of technoscience aimed at providing data for environmental management.

(2) Global diversity was co-constructed by scientists and policy advisors through international projects (such as the GBIF), and big data are the techno-political product. We do not suggest that ecology is a "pure science" that has been "contaminated" by politicians for societal reasons. We rather argue that ecological science has become a technoscience aimed at the management of biodiversity when it is locked in the endless process of big data accumulation.

(3) This paper also questions the relevance of global biodiversity management, based on big data measurements. One major value of the data accumulation described in this paper is to objectify the concept of global biodiversity with quantitative figures. Just like in the case of climate change, the concept of global biodiversity has been extremely efficient to convince the public that the biosphere is endangered. The abstract notion of global biodiversity provides whistle blowers with a powerful tool to raise the public sensibility regarding the future of the planet. While this is a powerful conceptual tool for alerting the public about the biodiversity loss it is less relevant for making the political decisions necessary to stop the erosion. The effort to gain political credibility is at the cost of the ecological significance of the data.

Overall, despite the promise of better decision-making based on the monitoring of the global biodiversity, the political objective to end the decline in biodiversity by 2010 has not been met and has been postponed to 2020 (Butchart et al. 2010). Incisive and insightful ways to understand and protect the multiplicity of complex, evolving, and interacting systems thanks to big biodiversity data are still missing. Rather, datafication is tightly linked to the litany of sustainable development. In this respect, GBIF data have already been used to analyze the benefits that ecosystems could bring to societies, adopting anthropocentric and utilitarian lenses to ecological problems and solutions (Schulp et al. 2014). Since the process shifts the focus from the empirical world to databases, it becomes possible 'to manage what is measured' while discarding the real world and its ecological or evolutionary specificities

(Sarrazin and Lecomte 2016). The risk is that big data will fail to address environmental issues and create an "inaction vortex" rather than better protection. In this context, the study of how ecological records can be used to restore the sense of place and the social values of ecological knowledge opens a new research agenda.

References

Andersson, J., & Rindzevičiūtė, E. (2012). The political life of prediction. The future as a space of scientific world governance in the Cold War era. *Les cahiers européens de Sciences-Po, 4*, 2–25.

Aronova, E. (2015). Environmental monitoring in the making: From surveying nature's resources to monitoring nature's change. *Historical Social Research, 40*, 222–245.

Aronova, E., Baker, K. S., & Oreskes, N. (2010). Big Science and Big Data in Biology: From the international geophysical year through the international biological program to the Long Term Ecological Research (LTER) Network, 1957–Present. *Historical Studies in the Natural Sciences, 40*, 183–224.

Balmford, A., Bennun, L., Brink, B., Cooper, D., Côté, I. M., Crane, P., et al. (2005). The Convention on Biological Diversity's 2010 Target. *Science, 307*, 212–213.

Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, J., Brown, J. H., Fortelius, M., et al. (2012). Approaching a state shift in Earth's biosphere. *Nature, 486*, 52–58.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics, 19*, 10–15.

Bensaude-Vincent, B. (2009). *Les vertiges de la technoscience: Façonner le monde atome par atome*. Paris: La Découverte.

Bensaude-Vincent, B., Loeve, S., Nordmann, A., & Schwarz, A. (2011). Matters of interest: The objects of research in science and technoscience. *Journal for General Philosophy of Science, 42*, 365–383.

Bisby, F. A. (2000). The quiet revolution: Biodiversity informatics and the internet. *Science, 289*, 2309–2312.

Bocking, S. (2013). The ecosystem: Research and practice in North America. *Web Ecology, 13*, 43–47.

Bowker, G. C. (2000a). Biodiversity datadiversity. *Social Studies of Science, 30*, 643–683.

Bowker, G. C. (2000b). Mapping biodiversity. *International Journal of Geographical Information Science, 14*, 739–754.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society, 15*, 662–679.

Bromley, D. A. (2002). Science, technology, and politics. *Technology in Society, 24*, 9–26.

Butchart, S. H. M., Walpole, M., Collen, B., van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., et al. (2010). Global biodiversity: Indicators of recent declines. *Science, 328*, 1164–1168.

Callebaut, W. (2012). Scientific perspectivism: A philosopher of science's response to the challenge of big data biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*(1), 69–80.

Calude, C. S., & Longo, G. (2015). The deluge of spurious correlations in big data. In *CDMTCS Research Report Series* (pp. 1–13).

Chase, J. M., & Leibold, M. (2003). *Ecological Niches*. Linking classical and contemporary approaches: University of Chicago Press.

Chase, J. M., & Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society of Londonc B, Biological Sciences, 366*, 2351–2363.

Clarke, G. (1954). *Elements of Ecology*. New Jersey: John Wiley & Sons INC, Chapman & Hall LTD.

Curry, G. B., & Humphries, C. J. (2007). *Biodiversity databases: Techniques, politics and applications* (Vol. 485). Abingdon: Taylor & Francis.

Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2012). Time to change how we describe biodiversity. *Trends in Ecology & Evolution, 27*, 78–84.

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., et al. (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology, 47*, 15–25.

Edwards, J. L. (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science, 289*, 2312–2314.

Edwards, P. (2010). *A vast machine*. Cambridge MA: The MIT Press.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics, 40*, 677–697.

Ellis, R., Pacha, M., & Waterton, C. (2007). Assembling nature: The social and political lives of biodiversity softwares. Lancaster.

Elton, C. (1927). *Animal ecology*. London: Sidgwick and Jackson.

Elton, C. S. (1966). *The pattern of animal communities*. London: Methuen and Co Ltd.

Foucault, M. (1980). *The confession of the flesh. in power/knowledge: Select interviews and other writings 1972–1977* (p. 193). New York: Pantheon Books Edition.

Greiner, W., & Lane, N. (2009). David Allan Bromley 1926—2005. *National Academy of Sciences*, 1–49.

Grinnell, J. (1917). The niche relationship of the California Thrasher. *The Auk, 34*, 427–433.

Grundmann, R., & Stehr, N. (2012). *The power of scientific knowledge. From research to public policy*. Cambridge: Cambridge University Press.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters, 8*, 993–1009.

Güttler, N. R. (2011). Scaling the period eye: Oscar drude and the cartographical practice of plant geography, 1870s–1910s. *Science in Context, 24*, 1–41.

Hamblin, J. H. (2013). *Arming mother nature: The birth of catastrophic environmentalism*. Oxford: Oxford University Press.

Hutchinson, G. E. (1957). Cold spring harbor symposium. Quantitative biology. *Concluding remarks, 22*, 415–427.

Jax, K., Jones, C. G., & Pickett, S. T. A. (1998). The Self-Identity of Ecological Units. *Oikos, 82*, 253–264.

Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions, 14*, 885–890.

Kelmelis, J. A., & Snow, M. (1991). *Proceedings of the U.S. Geological Survey Global Change Research Forum*. Circular 1086.

Kingsland, P. S. E. (2005). *The Evolution of American Ecology, 1890–2000*. Baltimore: The Johns Hopkins University Press.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*, 1–12.

Kwa, C. (2005). Local ecologies and global science: Discourses and strategies of the international geosphere-biosphere programme. *Social Studies of Science, 35*, 923–950.

Laney, D. (2001). *3D data management: controlling data volume, velocity, and variety*. META Group Research Note 6.

Lawrence, A. (2006). 'No personal motive?' Volunteers, biodiversity, and the false dichotomies of participation. *Ethics, Place & Environment, 9*, 279–298.

Leonelli, S. (2011). Packaging small facts for re-use: Databases in model organism biology. In P. Howlett & M. Morgan (Eds.), *How well do facts travel? The dissemination of reliable knwoledge* (pp. 325–348). Cambridge: Cambridge University Press.

Leonelli, S. (2014). What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society, 1*(1), 2053951714534395.

Levin, S. A. (1992). The problem of pattern and scale in ecology: The Robert H. *MacArthur Award Lecture. Ecology, 73*, 1943.

Loh, J., Green, R. E., Ricketts, T., Lamoreux, J., Jenkins, M., Kapos, V., et al. (2005). The Living Planet Index: Using species population time series to track trends in biodiversity. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 360*, 289–295.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., et al. (2015). Estimating species diversity and distribution in the era of Big Data: To what extent can we trust public databases? *Global Ecology and Biogeography, 24*, 973–984.

McAfee, A., & Brynjolfsson, E. (2012). *Big Data. Harvard Business Review*, (October), 60–68.

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution, 27*, 85–93.

Müller-Wille, S. (2015). How the great chain of being fell apart: Diversity in natural history 1758–1859. *THEMA La Revue Des Musées de La Civilisation, 2*, 85–95.

OECD. (1993). *Megascience and its background*. Paris: OECD.

OECD. (1999). *Final report of the OECD megascience forum.Working group on biological informatics*. OECD: Paris.

OECD. (2003). *OECD Environmental Indicators. Development, Measurement and Use*. *OECD Reference paper* (Vol. 51).

Pielke, R., & Klein, R. A. (2010). *Presidential Science Advisors: perspectives and reflections on science, policy and politics*. New York: Springer.

Ratchford, J. T., & Colombo, U. (1996). *Megascience. UNESCO World science report*.

Sarkar, I. N. (2009). Biodiversity informatics: The emergence of a field. *BMC Bioinformatics, 10*(Suppl 1), 1–2.

Sarrazin, F., & Lecomte, J. (2016). Evolution in the Anthropocene. *Science, 351*, 922–923.

Schulp, C. J. E., Thuiller, W., & Verburg, P. H. (2014). Wild food in Europe: A synthesis of knowledge and data of terrestrial wild food as an ecosystem service. *Ecological Economics, 105*, 292–305.

Shavit, A., & Griesemer, J. (2011). Transforming objects into data: how minute technicalities of recording 'species location' entrench a basic challenge for biodiversity. In M. Carrier & A. Nordmann (Eds.), *Science in the context of application* (pp. 169–193). New York: Springer.

Slota, S., & Bowker, G. C. (2015). On the value of 'useless data': Infrastructures, biodiversity, and policy. iConference 2015 Proceedings. http://hdl.handle.net/2142/73663. Accessed 5 Sep 2016.

Soberón, J., & Peterson, A. T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 359*, 689–698.

Stevens, H. (2013). *Life out of sequence—A data-driven history of bioinformatics*. Chicago: University of Chicago Press.

Strasser, B. J. (2012). Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Biological and Biomedical Sciences, 43*, 85–87.

Takacs, D. (1996). *The Idea of Biodiversity: Philosophies of Paradise*. Baltimore: Johns Hopkins University Press.

Turnhout, E., & Boonman-berson, S. (2011). Databases, scaling practices, and the globalization of biodiversity. *Ecology and Society, 16*(1), 35.

Turnhout, E., Dewulf, A., & Hulme, M. (2016). What does policy-relevant global environmental knowledge do? The cases of climate and biodiversity. *Current Opinion in Environmental Sustainability, 18*, 65–72.

Turnhout, E., Neves, K., & De Lijster, E. (2014). 'Measurementality' in biodiversity governance: Knowledge, transparency, and the intergovernmental science-policy platform on biodiversity and ecosystem services (ipbes). *Environment and Planning A, 46*, 581–597.

Watson, R. T. (2005). Turning science into policy: Challenges and experiences from the science-policy interface. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 360*, 471–477.

Wilson, E. O. (1985). The biological diversity crisis. *BioScience, 35*, 700–706.

Wilson, E. O. (1988). *Biodiversity*. (N. A. of Science, Ed.).