Research, part of a Special Feature on Scale and Governance

Databases, Scaling Practices, and the Globalization of Biodiversity

Esther Turnhout [1] and *Susan Boonman-Berson* [1]

ABSTRACT. Since the Convention on Biological Diversity in 1992, biodiversity has become an important topic for scientific research. Much of this research is focused on measuring and mapping the current state of biodiversity, in terms of which species are present at which places and in which abundance, and making extrapolations and future projections, that is, determining the trends. Biodiversity databases are crucial components of these activities because they store information about biodiversity and make it digitally available. Useful biodiversity databases require data that are reliable, standardized, and fit for up-scaling. This paper uses material from the EBONE-project (European Biodiversity Observation Network) to illustrate how biodiversity databases are constructed, how data are negotiated and scaled, and how biodiversity is globalized. The findings show a continuous interplay between scientific ideals related to objectivity and pragmatic considerations related to feasibility and data availability. Statistics was a crucial feature of the discussions. It also proved to be the main device in up-scaling the data. The material presented shows that biodiversity is approached in an abstract, quantitative, and technical way, disconnected from the species and habitats that make up biodiversity and the people involved in collecting the data. Globalizing biodiversity involves decontextualization and standardization. This paper argues that while this is important if the results of projects like EBONE are to be usable in different contexts, there is a risk involved as it may lead to the alienation from the organizations and volunteers who collect the data upon which these projects rely.

Key Words: *biodiversity; classification; databases; scaling practices; standards; statistics; volunteer recorders*

GLOBAL BIODIVERSITY, GLOBAL SCIENCE

It has been suggested that biodiversity is the main issue of the 21st century (Wilson 2000, Bowker 2005). Since its emergence, biologists and conservationists started to use it as the main concept to describe the phenomenon they study and argue for its protection (Takacs 1996). It proved very useful because it was able to connect various concerns and issues, including endangered species, the preservation of wilderness areas, and nature conservation in one term (Takacs 1996). Indeed definitions of biodiversity are generally broad, referring to the totality of life on earth. The formal definition of the Convention on Biological Diversity (CBD) (Article 2 of the Convention text) serves as an illustration:

Biological diversity means the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems (Secretariat of the Convention on Biological Diversity 2006).

If biodiversity, as the CBD suggests, is a global phenomenon, its degradation is a global issue. Global environmental problems are of relatively recent origin. It is commonly argued that the first picture of the earth taken from space was a crucial element in their invention (e.g., Hajer 1995, Escobar 1996). Apparently, seeing the earth from space created awareness of the planet's wholeness and fragility and of the importance of joining efforts to

---

[1]Forest and Nature Conservation Policy Group, Wageningen University, The Netherlands

conserve it. The global nature of the issue also implies that a global scientific effort is needed to address it. The fact that the amount and kinds of biodiversity are currently unknown is generally seen as a problem. It is commonly agreed that global biodiversity needs to be known and counted before it can be effectively managed and protected. According to the CBD, for example, current deficiencies in taxonomic knowledge impact on "our ability to conserve, use and share the benefits of our biological diversity" (Secretariat of the Convention on Biological Diversity 2007). It is in this sentiment that E.O. Wilson (2000:2279) emphasizes the importance of finishing "the Linnaean enterprise" and "describe and classify all of the surviving species of the world".

The generation of global knowledge about global biodiversity faces two major problems. The first is the generation of reliable data. Biodiversity recording requires not only highly specialized knowledge about species identification, but also a huge amount of resources in terms of time and money. Part of this involves the recruitment and training of volunteer recorders. Science alone will never be able to achieve a global census of biodiversity and thus, these volunteers are absolutely vital. Ensuring that these volunteers remain motivated is an issue that deserves more attention than it currently receives (Bell et al. 2008). Many volunteer recordings are motivated by a love of nature and a concern for its conservation (Ellis and Waterton 2004, Lawrence and Turnhout 2010). Although they are also motivated by contributing to science and want their data to be used in policy, there may be tensions involved when volunteer data are to be used in scientific and policy projects and have to meet strict quality and reliability standards (Lawrence and Turnhout 2010). The second major problem is the usability of biodiversity data. Many biodiversity data are not globally available, either because they are not digitized and inscribed in standardized databases, or because the biodiversity databases that do exist are incompatible, leading to "maddening difficulties in knowing what is where and comparing like with like" (Bisby 2000:2309). Consequently, existing biodiversity databases are unable to form a global whole.

The intended users of global knowledge about global biodiversity are rarely clearly identified (Ellis et al. 2010). The implicit assumption is that this knowledge is of use to the global community at large. What it envisions resembles what Haraway (1988:581) has called the "god-trick of seeing everything from nowhere". The GBIF (Global Biodiversity Information Facility) for example, aims to open up the world's biodiversity data to all (http://www.gbif.org/index.php?id=269, accessed Sept.27, 2010). However, this "nowhere" is not nowhere, nor is it everywhere. It is in practice highly localized and situated. Just as any other scientific practice, biodiversity science is done by individuals in their offices behind their computers, in conference rooms, in their laboratories, or in the field. In addition, it is centralizing. The GBIF has the ambition of becoming the gateway to biodiversity, particularly species, data, and a global network that connects and integrates the various databases and biodiversity scientists and recorders all over the world. The GBIF is not the only initiative (see Hine 2008 and Bowker 2005 for an overview). Hine (2008:187) refers to a dance of initiatives, a parade of interlinked initiatives and acronyms that have emerged "in part by the increasing significance of biodiversity on the global political stage". Although the different initiatives differ in important ways in terms of structure, organization, and objectives, each aims to integrate biodiversity data and make them available at a higher level. In other words, each aims to become a "center of calculation" (Latour 1987:215) or an "obligatory passage point" (Callon 1986:204). Thus, the emergence of biodiversity as a global phenomenon and its degradation as a global issue have coincided with the emergence of a global science that is dedicated to knowing and representing global biodiversity, or, in other words, to the globalization of biodiversity.

Despite its global character, this science can be understood as a situated scientific practice; the global is created locally. Two key factors that go into the globalization of biodiversity are databases and scaling. Biodiversity scientists create databases to store biodiversity data and they use scaling techniques in order to transform these data into scaled-up representations of biodiversity. Thus, to understand how biodiversity is globalized, it is important to investigate the databasing and scaling practices in which this is achieved. Consequently, we address the following research questions:

- How is biodiversity globalized in practices of databasing and scaling?

- Which dilemmas, negotiations, and techniques are involved in these practices?

- How do these practices influence how biodiversity is globalized and represented?

Our analysis is based on empirical material from the EBONE project: the European Biodiversity Observation Network. EBONE is a project funded by the European Union's Seventh Framework Program, which aims to develop a coherent and cost-effective system of biodiversity assessment. Although it focuses on the European level, admittedly only a small subsection of the global, it is part of ongoing attempts to globalize biodiversity as it aims to analyze and integrate biodiversity data and make them available. Before introducing the EBONE project and presenting the findings, relevant theoretical background of databases, scientific practice, and scale are presented.

DATABASES, STANDARDS AND CATEGORIES

Databases are an essential part of biodiversity science. Representing biodiversity on a global level requires data about global biodiversity that is stored in databases. How these databases are constructed impacts on how biodiversity is known and represented. Databases are often considered to be neutral carriers or storage devices. Like closets in which you put away clothes or linens, databases are generally not seen as influencing the items they carry. However, on closer scrutiny, this falls apart. Again using the storage cabinet analogy, the things that go into it have to fit and the cabinet has to be organized in such a way that order and retrieval are possible. Then things get complicated: should it have doors or drawers, how many, and what size? Should the drawers have subcompartments like cutlery drawers, and if so, how many and what size? A final answer to these questions can only be given if it is known what it is that needs to be stored. However, a storage cabinet needs to anticipate the future to prevent the necessity of designing and creating a new facility every time a new piece of silverware or clothing is purchased. Thus, databases always involve bootstrapping problems; you need to classify the data before designing the databases and you need to know about database before you can classify the data (Bowker 2000). This combination of meeting existing requirements and anticipating possible new future ones makes the design of databases very complex.

Databases need to be expandable and interoperable with other databases and the information stored in them needs to be retrievable and accessible for all sorts of purposes. Although the actual reuse of archived and stored information is rare, this potential memory function is enough to drive the continuous elaboration of databases (Bowker 2005). After the database has been created, its contents can be forgotten because the database promises the possibility of retrieval. In that way, forgetting is as much a motivation for the creation of databases as memorizing, albeit a more implicit one, and the pursuit of ignorance goes hand in hand with the pursuit of knowledge (McGoey 2007). A high-modernist, aesthetic ideal is present here that envisions a global network of completely interoperable and accessible databases containing data of all life on earth. As is also clear in the case of the CBD, complete and accessible information about the world is seen as a prerequisite for its management and control (Scott 1998, Turnhout 2010).

Standards and classification systems are indispensable for databases. Data can only be stored in the database if it meets certain standards and fits into the categories of the classification system. Biodiversity databases involve different kinds of categories. Perhaps the most important one of these is the category of species. Data are classified into the taxonomic classification of species. In addition, biodiversity databases involve tags and metadata. For example, each individual species record comes with tags that refer to the date of the observation, the habitat type in which the species was observed, the name of the observer, etc. Together, the classification system and the standards form the "filing system" of the database, which is key to data storage and retrieval. The main problem in connecting databases and making them interoperable is that databases are highly tailored to what they intend to store. Consequently, it is likely that different databases will use different standards and categories, that is, different filing systems (Bowker 2005). Although it is tempting to look at the categories of a classification system as preexisting empty containers with fixed boundaries to be filled with data that fit the categories, the reality is markedly different (Jones 2009). Categories and what gets classified in them are mutually constitutive; they bring each other into being. The making of a category involves defining standards and criteria and in the process, the items that go into it are named, labeled, and remade as belonging to

the category. Just as the development of classification systems requires preconceived ideas about the data that need to be classified, data production requires the existence of categories and classification systems; measuring or counting something always involves measuring or counting something as a member of a certain category (Stone 1988).

Biodiversity databases, the different taxonomic and other categories and standards they involve, and the data that they store are all human made. They are created in scientific practices. Bowker and Star (2000) argue that despite their human-made character, classifications and standards become increasingly normalized and accepted as appearing to stem from nature itself. This is certainly the case for common standards in everyday life: our gender classification tells us which toilet to use; and we sort our laundry before putting it in the washing machine (Bowker and Star 2000). However, under close scrutiny, the normality of these standards can always be disputed. Arguably, classifications of nature have never become fully normalized in the sense of being accepted as true (Turnhout 2009). While the distinctions between, let's say, plants and animals may appear to be obvious, biologists continuously face boundary problems trying to demarcate these two, or other categories of nature. Still, classifications of nature and biodiversity, even if their acceptance is incomplete and temporary, are not without consequences (Bowker 2000, Bowker and Star 2000). Abstract conceptualizations of the world such as classification systems are performative; they have the tendency to remake the world in their image. To quote Bowker (2005:659):

> [Databases] are set up so that a particular, skewed view of the world can easily be represented. With these [...] in place, it is easier to get funding and support for research that reproduces this view. [...] Thus, the world that is explored scientifically becomes more and more closely tied to the world that can be represented by [...] one's databases: and this world is ever more readily recognized as the real world.

In a similar vein, West and Brockington (2006:609) use the concept "virtualism", which is defined as "the attempt to make the world around us look like and conform to an abstract model of it".

This performativity means that the ways in which biodiversity is known, the ways in which this knowledge is archived, and the ways in which biodiversity comes to be represented based on this knowledge, have real implications for how biodiversity is treated in practice. We act on what we know and on what we have come to understand biodiversity to be. For biodiversity this means that only what is counted counts (Bowker 2000, Lawrence and Turnhout 2010). Only those species that are known, counted, and represented in databases can be actively protected. Reversely, when uncounted and unknown species that are not represented in databases go extinct, we will never know about it; it will be as if it never happened. We will now discuss exactly how biodiversity data, stored in databases, are transformed into representations of global biodiversity.

SCALING AS SCIENTIFIC PRACTICE

As social practices generally are, scientific practices of globalizing biodiversity and creating databases are likely to be complex and messy. Studies in the sociology of scientific knowledge have convincingly demonstrated the role of values and interests in the production of scientific knowledge and have argued that science and society are not separate entities but continuously coproduce each other (Shapin and Schaffer 1985, Jasanoff 2004). Recognizing this means (re)conceptualizing science as practice (Pickering 1992). The case of biodiversity databases is no exception to this. As will be demonstrated in more detail, EBONE's project members have to deal with different existing compatible and incompatible biodiversity databases in different countries and regions in Europe. They have to develop standards and criteria that guarantee reliability and harmonization, and are feasible and pragmatic at the same time. To meet their objectives, EBONE's project members have to work with many different actors. This includes large groups of organized and unorganized volunteers who are active in natural history and biodiversity recording. Possibly, there is a tension between the ideals and motives of the volunteers and the scientific ideals that drive the rationalization of biodiversity in the EBONE project (Lawrence and Turnhout 2010). Thus, in the EBONE project we expect to encounter the simultaneous occurrence of scientific ideals related to reliability, objectivity, and complete interoperability, and pragmatic considerations of feasibility and how to make use of existing data. It is relevant to

document this in detail, because in science, these social and practical aspects tend to disappear from view when the end product is published and portrayed as being natural, objective, and true (Latour and Woolgar 1979).

We zoom in on one particular issue involved in biodiversity databases: scale. Scaling in this context refers to the extrapolation of in-situ biodiversity observations to global, full coverage representations of biodiversity and doing so in a statistically and scientifically acceptable way. Scale is a much debated topic in ecology, geography, and various other scientific disciplines (Termeer et al. 2010, Buizer et al. 2011). Ecologists distinguish between the scales of individual organisms, communities, ecosystems, or bioregions. Within policy studies it is common to distinguish between different administrative levels or scales of governance (local, regional, national, or global). Complex environmental issues such as climate change or biodiversity loss often simultaneously involve two or more conceptions of scale (Meadowcroft 2002, Cash et al. 2006). Implicit in these arguments is an essentialist conception of scale in which, "scales are [...] taken for granted as almost `natural' units for social existence" (Swyngedouw 2000:68). In contrast, much of the current literature holds that scales are constructed in representational practices of scaling (Jones 1998). Or, as Delaney and Leitner (1997:93) put it: "geographic scale is [...] socially constructed rather than ontologically pre-given, and [...] the geographic scales constructed are themselves implicated in the constitution of social, economic and political processes&#8221. None of this means that scales are not real, on the contrary, their socially constructed character points to the fact that they are made and thus real. However, their construction in practice does point to their pragmatic, arbitrary and malleable nature.

Within political geography, considerable attention has been paid to the politics of scale, which points to the politics involved in creating scales and the political implications of created scales. Defining a scale involves demarcation and boundary work. Necessarily, items belonging to a certain scale are demarcated from those that do not, and as such scaling involves exclusion. Apart from exclusion, co-option is also likely to take place because scales often involve hierarchies (Bulkeley 2005). Rather than local scales coexisting next to global ones, global scales can be seen to co-opt the local ones when the local is considered to be represented by the global. Politics is involved in scale making as "the continuous reshuffling and reorganizations of spatial scales are an integral part of social strategies and struggles for control and empowerment" (Swyngedouw 2000:70). The outcomes of these processes are important because scales, once produced, have real consequences (Bulkeley 2005). Again quoting Swyngedouw (2000:70): "In a context of heterogeneous social and ecological regulations, organized at the corporeal, local, regional, national, or international level, mobile people, goods, capital, and hypermobile information flows permeate and transgress these scales in ways that can be deeply exclusive and disempowering for those operating at other scale levels".

Generally, scaling up involves making something bigger, more important, and more universally valid. As such, it is not surprising that many scientific practices involve scaling. Often natural science scaling practices involve statistics, maps, and other methods or scaling devices[1] such as computer models. Important issues are whether and how you can scale up, for example, from a single scientific experiment or measurement to statements about large areas, or scale down, for example, from national surveys to site specific information. The spatiality implicit in the notion of scale invites talking about scaling as a form of travel, but one that involves multiple transformations and translations. Biodiversity observations are transcribed into standardized forms. These forms are sent to institutions that process the forms and digitize them. These data then move to sites of quality control. This makes the data fit for yet another movement: storage in the database. And from then on, the journey will involve travel throughout the network of interoperable biodiversity databases as envisioned in the mission of projects like EBONE or the GTI. The data are now fit for linkage with other data and for performing the various analyses, statistical and other, which are required for their globalization. In the journey from observation to database, biodiversity data become disassociated from the context of observation; they are validated, standardized, and objectified and stored in databases (Lawrence and Turnhout 2010). As such, the data become ever more abstract and ever more universal. None of this is automatic though. Scaling is a scientific practice in the sense that people actively work and cooperate to collect, integrate, process, analyze, and scale up biodiversity data in order to achieve full coverage representations of biodiversity.

THE EBONE PROJECT

EBONE is a collaborative project of the Seventh Framework Program within theme 6: Environment (Topic 4.1.1.2. Contribution to a Global Biodiversity Observation System). The project is designed to respond to the widely recognized problem of limitations in the linkages among existing monitoring systems, databases, and monitoring sites. These limitations refer to the different and uncoordinated approaches in measuring and reporting reliable trends and changes in biodiversity across geographical and temporal scales, as can be seen in the online database DaEuMon, a result from the EuMon project, in which a detailed picture is given of monitoring biodiversity practices in Europe (Halada et al. 2009). As such, the strength of EBONE is that it builds on available knowledge and existing networks and makes use of different existing monitoring systems, such as those developed in earlier European Framework projects like ALTERNET, BioHab and EuMon, in order to end up with efficient indicators and a well balanced sampling program (Halada et al. 2009). Policy relevance is an important objective for EBONE. The project was set up to achieve a coherent system of data collection that can be used for internationally comparative assessments on an international level. By selecting indicators based on existing policy frameworks, EBONE will be able to provide relevant information for evaluating the implementation of policies such as Natura 2000 (Halada et al. 2009). EBONE works with different types of data. The first source is in-situ data about species and habitats. Most of the species data come from country-based recording organizations and monitoring schemes. The habitat data come from different existing habitat monitoring schemes such as the Countryside Survey (UK), NILS (Sweden), SISPARES (Spain), and SINUS (Austria) and will be integrated based on the habitat categories of the BioHab project. It is EBONE's intention to build as much as possible on existing data and networks and collaborate with the organizations involved, such as research institutes and recording organizations. In this way, EBONE can also achieve its aims in a cost-effective way. A crucial aspect of EBONE is the integration of these in-situ data with earth observation data. Combining these two data sources in a way that is scientifically acceptable is a major challenge for EBONE, especially since the project will be working with existing in-situ data that is often fragmented, patchy, incompatible, or collected in ways that do not meet commonly agreed scientific standards. As such, the material about the EBONE project presented here offers a unique opportunity to look behind the scenes at databasing and scaling in action and the difficulties, dilemmas and techniques involved.

The material we present is from two main sources. The first is publications and documentation about the project. The second is transcripts from one of the project's meetings, specifically the second project meeting. Note that the unreferenced quotations included in the next section are from participants at this meeting. The main aim of the meeting was to measure progress and discuss the state of the art of the different work packages; note that EBONE has 10 different work packages (see http://www.ebone.wur.nl/UK for more information). It was the kick-off meeting of Workpackage 6, which focuses on "Field validation of the methodological framework". Many of the discussions focused on Workpackage 5, which is called "Intercalibration of EO data with in-situ observation". These discussions proved very valuable for the insights they provided into practices of databasing and scaling. The meeting was held from 15-17 April 2009, one year after the start of the project in April 2008, in Aix-en-Provence, in the south of France. The participants in this meeting included the partners of the EBONE project. These consist of 18 European nature research institutes and universities, which already have a history of collaboration over many years, as well as partners from Israel and South Africa. These were added to test the EBONE framework on a global level. In general, the participants all had an ecological background. The presentations and discussions that took place were all taped by the second author who participated in the meeting. Subsequently, all recordings were played in full (a total of 15 hours). Based on that, 13 hours of recording were selected for full transcription. The transcripts were analyzed for key topics of interest related to the issues addressed in the theoretical sections on databases and scales. In particular, the material was analyzed for references to scaling methods and devices, the use of standards, and the practical dilemmas and challenges involved in the creation of databases. Quotes from the transcripts are anonymized in a traceable way by numbering the different participants (P1, P2 ...P9) and listing the transcript file number (T1, T2, ...T7). The material is organized into five subheadings, which emerged from the initial analysis of the transcripts and represent key topics discussed during the meeting.

NEGOTIATING AND SCALING BIODIVERSITY DATA

Bootstrapping and the use of existing data

EBONE was faced with the persistent bootstrapping difficulties in creating biodiversity databases: the interdependence between the structure and design of the database, the objectives of the project, and the data that is available. What EBONE can achieve depends on what is there to build on and vice versa. It was EBONE's explicit intention to link up with existing systems and use existing data: "[EBONE will] develop a conceptual framework for monitoring, utilizing the existing institutional context of European monitoring, databases, observation points and observing organizations, agencies, and NGOs" (Halada et al. 2009:180). As one of the participants put it, "we do not want to invent something new but we want to base ourselves on existing data models and existing solutions".

This was also key to achieve cost-effectiveness because if EBONE linked up with what biodiversity recording organizations are already doing and if EBONE provided them with something they can work with, they would do part of the work themselves as part of their jobs:

> If you design something which is not done in a large part of the monitoring schemes already. They will not be willing to implement [it] and change their habits;
> It is very important that we have a flexible protocol in the end that can be used by a wide range of users.

The importance of using existing data was never questioned. It was however considered to be a big problem. The first challenge was to know which data and which monitoring systems exist out there and how they can be used by EBONE. As the following quotes from the meeting participants illustrate, this was considered necessary before EBONE could start its work and decide on what it wants to achieve:

> Nobody in their right minds wants to design a [...] system without being absolutely sure what the goals are. And yet, we really do not actually have those;
> [we need to] know what we are really monitoring [...]. Because only by then the database system itself can be designed. Before that there are only technical thoughts about what this could look like.

Statistics

There was a general realization that although there were a lot of data available, using them would be an entirely different matter. According to the participants, much of the available species data did not meet scientific standards: "monitoring needs to improve. [...] [in order] to present an unbiased and realistic picture of the state of Europe's biodiversity" (Halada et al. 2009:179). According to one of them, the information collected did not even deserve the qualification "data": "most [...] monitoring has no data behind it at all, it's a walk of a person through a site with a tick list".

Also for the habitats, existing systems served as the basis. However, these were not compatible in terms of their sampling methods and several of them did not meet the standards of the participants:

> Different countries are doing slightly different things, and it is far from a perfectly random sample of sums. Most of these times they are selective because that is where people have traditionally done this.

The preferred sampling procedure is stratified random sampling. This means that certain non-overlapping strata are indentified within which samples are selected randomly. Using strata is a way of ensuring that all relevant categories are sampled, and in the case of biodiversity, meaningful strata are often based on habitat, soil type, biotype, or land use classifications. Based on a comparative assessment of some of these schemes, one of the participants concludes: "sometimes it is not clear what is meant with random sampling".

This makes clear that there are different ways of doing random sampling. Some forms of random sampling were criticized because they made it impossible to calculate the sampling error and the probability with which the sample is selected:

> The [...] samples within the strata are not selected independently. [...and] this seriously complicates the estimation of the sampling variance.
> if you want to quantify the uncertainty then I have problems with [some of the existing schemes].

The question was how to include these imperfect existing data in a way that was scientifically acceptable for the participants, as one participant commented, "we have to consider how we are going

to incorporate the data, which may be softer and may not be stratified random".

It was important that statistics would not get in the way of feasibility. As one participant cautioned:

I have the feeling that we are trying too much to be statistically correct, while in the end [...] each country is going to go with its own method [...]. Are we not trying too much?

Thus, the EBONE project would have to apply statistics in a pragmatic way and make compromises; as one participant put it, " the [...] bigger [a] picture we want to create of biodiversity [...], the more we have to make these compromises between perfect design and making the best use of information".

Although there was some strong opposition to this based on a fear that all scientific validity would be lost, according to the participants, a compromise was reached in the end to go for minimal standards and work with the imperfections of existing data.

Biodiversity indicators

The transcripts show that discussions of monitoring and sampling procedures and statistics dominated the meeting. There was very little mention of biodiversity itself. It entered into the discussions only when indicators and the kind of data that would be included in EBONE were discussed:

These are the priority indicators: [...] habitat assessment, abundance, and distribution of selected species, fragmentation of natural and semi-natural areas; the total area of habitat types, land cover types [...], the temporal trend [...], the spatial mean on the total of biodiversity indices, the quality of the habitat type. And so on....

The use of indicators shows that biodiversity is not assessed directly. It is calculated by means of indicators that serve as proxies. As pointed out before, the necessary data for these indicators would have to come from existing monitoring schemes, especially birds, plants, and butterflies because, according to participants, for those species groups, most data were available. However, the ecological significance of birds and butterflies was doubted because they migrate:

I have a problem with birds and butterflies. Not that I do not like them but [they are not] good indicators. [...] I think we have to be careful not just to go with what people are doing because [...] the data [...] are very convenient. We have to go for good indicators.

This points again to the difficult dilemmas involved between scientific validity and using existing data, which kept emerging throughout the meeting.

The network

Most of the species and habitat monitoring that serve as a basis for EBONE is done by a wide variety of national organizations and it is essential that EBONE establish a "sound institutional framework to ensure continuity and long-term collaboration between partners" (Halada et al. 2009:178). However, this adds greatly to the complexity of the project. As commented by one participant:

EBONE [...] has an amazingly complex and difficult institutional goal of actually establishing an operational network. That is going to be very, very tricky.

EBONE would have to find a way to coordinate all the different existing monitoring efforts. As stated by one participant, "[t]The key activity is to determine how existing monitoring data can be coordinated into a standardized framework"

Getting all these people to do their monitoring in the same way was difficult because some of the categories they would have to monitor were quite ambiguous. The next quotation lists two examples that illustrate this ambiguity:

The question is how many [wetland] species do you need to have before you go from pure grassland with some broadleaved plants to a mixed swamp. [...] That is an absolutely critical decision. [...] It means that it changes from being outside the Annex 1 to the inside of Annex 1[2]. So it is an absolutely crucial change. [...] There is another example, also very important. There is a definition in the Annex 1 habitats called blanket bog. A blanket bog is a habitat complex, it is not a simple habitat. It is a complex of grassland and shrubs and wetlands and bog pools. The critical thing is that if you are mapping,[...] you can map

an area of grassland in that bog and call it grassland. Yet the whole landscape feature of blanket bog covers the whole landscape.

The different possible interpretations would have to be aligned. One way to achieve that was to develop a standard protocol. However, this protocol would have to have a certain flexibility and be open to the requirements of stakeholders:

We should design a system [that is] applicable and worldwide acceptable, that means that we have to, not only to develop the technical systems, we have to go to users and [...] tell them how they can do it and we have to involve their views.

The protocol would require coordination. For example, according to one participant, if somebody suggested a change or an addition, this would have to be agreed centrally. Another way to coordinate and harmonize monitoring was to develop a usable handbook that describes "the methodology appropriate for coordinating information on habitats in order to obtain statistically robust estimates" (Bunce et al. 2005:11). This handbook would have to contain detailed rules because "otherwise real changes cannot be separated reliably from background noise" (Bunce et al. 2005:11). It explains for example how to map a road margin:

If you have a road that is three and a half (3.5) meters wide, with a margin along the side of one meter [...], you will be mapping that. And it will be mapped as an area. [...] If there is one meter on either side, then you [...] map those two areas as lines. [...] They may be the same. They may be just grass in which case you record it as grass. But one may be grass and the other may be shrubs. [...] How do I know? The rule is that if the vegetation is the same on the side of the track as it is in the surrounding area, then you do not map it. [...] that is all in the handbook.

However, the handbook is very detailed and it was recognized at the meeting that to really understand and be able to apply the handbook, training is required:

Getting to know the rules is also very important. [...] There is a lot of work there, [...] if you pick up the handbook and [...] try to apply it;

[there should be] one participant [that] [...] organize[s] [...] training in his own institution and supervise[s] the work.

Finally, to really make sure that the data are reliable and collected in the right way, control was considered necessary. According to Bunce et al. (2008:20), "quality control (i.e., supervision of surveyors) and assurance (i.e., independent checks of recording) are all essential to produce robust data. [...] so that policy makers and scientists would have confidence in the results". By means of the standard protocol, the handbook, the training sessions, and the quality control and verification procedures, EBONE would be able to standardize and coordinate the monitoring, manage the network, and achieve reliable results.

Scaling up

As we have seen, the starting point for EBONE is very local and fragmented. It involves clusters of data about different aspects of biodiversity, collected in different sites, with different methods, and for different purposes. EBONE was faced with the challenge of scaling up these patchy data to become usable at a higher level:

The problem is how can you go from those sorts of observation [...] to a larger area? [...] We [...] need to build up from sample-based inferences for local areas to [...] inferences at a larger scale.

To achieve this, the use of earth observation data was indispensible. In fact, it was seen as one of EBONE's main innovative aspects: as "EBONE is about developing new methods for integrating data. [...] combining in-situ [and] observation data effectively". The integration of species and habitat data with earth observation data was the key to achieving full coverage scaled-up biodiversity information (Figure 1).

The main scaling device was the statistical procedure of intercalibration through the use of correspondence matrices. As one of the meeting participants explained:

[We have] in-situ observations [...] and we also have a land cover map which provides full coverage. So what [...you can do...] in terms of intercalibration is [...] produce

Fig. 1. The relation between species, habitats and earth observation data (http://www.EBONE.wur.nl/ UK/Project+Work+Packages/WP5+Inter-calibration/, accessed March 16, 2010)

correspondence matrices. [...] And these correspondence matrices actually help us to identify how good the link is between the two.

Subsequently, the correspondence matrices could be used to correct the in-situ data in such a way that they could be connected to the earth observation data, thereby creating a new, integrated, scaled-up, full cover result (Figure 2).

Importantly, this procedure also increased the robustness of the outcomes. The problems that were identified with the species and habitat data, could, at least partly, be ameliorated by combining them with observation data:

> *[You can use the] correspondence matrices by overlapping in-situ with the land cover map, to then produce a statistically better result;*
> *[One...] parcel [...] will be incorporated in the correspondence matrix [...] but actually that parcel extends right [...] out into the surrounding squares. So you have a benefit from using the extra surrounding squares. This procedure will increase that sample number [...]. So you will actually be greatly reducing the standard error.*

As such, the statistical procedure of intercalibration through correspondence matrices proved to be of critical importance in integrating data from different sources, in up-scaling them and in improving their overall reliability. EBONE's scaled-up and integrated outcomes will be able to function as globalized representations of biodiversity.

PRACTICES OF DATABASING AND SCALING

The findings illustrate several of the points raised earlier. First of all, the "maddening difficulties" (Bisby 2000:2309) involved in databasing were obvious throughout the results. The first difficulty involved bootstrapping: the interdependent relation between the database and what is to be stored in it. EBONE had to be based on existing data and monitoring schemes, so before the design of the database could start, it was important to find out what data were out there and how usable they were. However, to assess their usability, it was important to know more about the database and its requirements. This discussion went back and forth throughout the meeting. A second difficulty was related to the reliability and compatibility of the existing data. It was clear that the data out there were all collected in different ways, some of which were more reliable and up to standards than others. Although ecological criteria were mentioned, the standards used for the most part were statistical and referred to the sampling method that was used and how that affected the possibilities for statistics and analysis. A third main difficulty referred to the complex institutional network; the different local organizations and institutions involved in the monitoring work. EBONE's challenge was to coordinate all the different monitoring efforts and try to achieve a certain degree of harmonization and standardization. In dealing with these difficulties, there was a general consensus about the need to be pragmatic. Designing the database would have to start before everything was clear about existing data. Statistics were indispensible but should not stand in the way of feasibility and should be employed in a pragmatic way. Harmonization and standardization

Fig. 2. A visualization of the intercalibration procedure (adapted from http://www.EBONE.wur.nl/UK/ Project+Work+Packages/WP5+Inter-calibration/, accessed March 16, 2010). The question mark indicates EBONE's results.
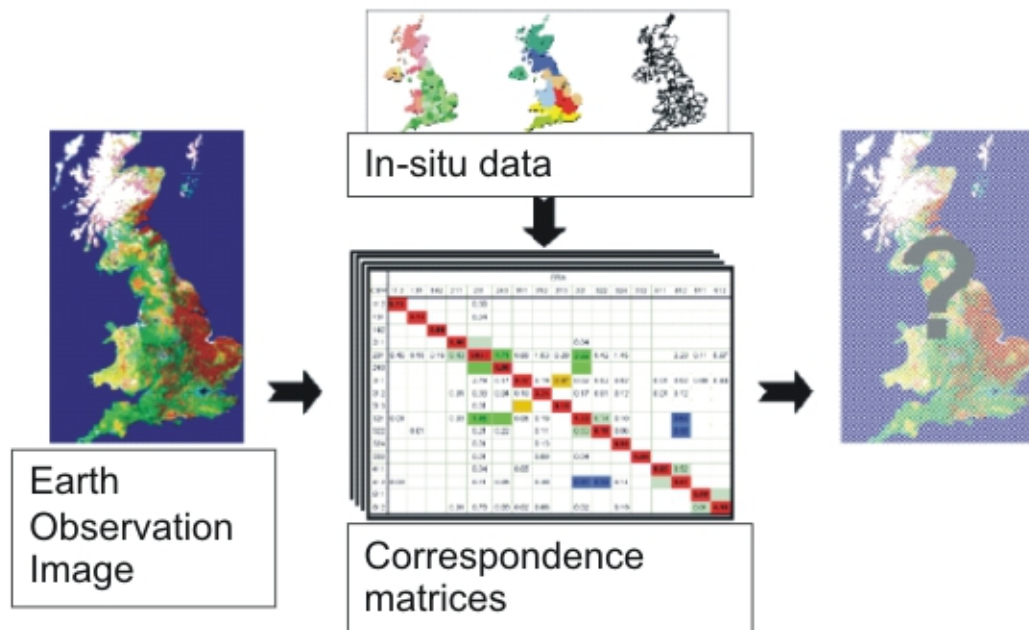
were important, but to keep everybody on board, EBONE's requirements would have to be minimal and fit with what everybody was doing already. The EBONE project is thus a typical example of a scientific practice in which the actions of the actors resulted from the continuous interactions between scientific ideals and practical considerations.

EBONE's centralizing features are striking. EBONE aims to serve as a crucial mediator between the European Commission and its nation states by producing information about Europe's biodiversity that is linked to European environmental policies. EBONE also attempts to integrate all biodiversity information in Europe, leading to one monitoring system for everybody to use. To achieve this, the project wishes to coordinate, harmonize and standardize all of the ongoing monitoring. Although EBONE realizes that this will probably not be feasible in the short term, effort is put into ensuring this for the long term by means of protocols, handbooks, and training. As such, EBONE envisionsbecoming an obligatory passage point for biodiversity information in Europe (Callon 1986).

Interestingly, no mention was made of how to achieve this centralization and standardization while at the same time ensuring the cooperation of the current recording and monitoring organizations and the volunteers who do most of the recording and monitoring. This may pose some real challenges for EBONE in the future (Bell et al. 2008, Lawrence and Turnhout 2010). While the terms gateway or obligatory passage point may suggest a neutral function for EBONE, the term "center of calculation (Latour 1987)" does more justice to the work that goes in before EBONE can play that role. Statistics were important to standardize the data, calculate the variance and sampling errors involved. This was very important for the reasons of authority and credibility. The potential political implications of the data were significant; if EBONE identified certain areas as containing rare and important habitats and species, EU member states were required to take action to protect them. Moreover, sometimes the boundaries between habitats of different conservation status were quite blurry and unclear. EBONE recognized that statistical validity was essential to protect the data from contestation

and deconstruction (Porter 1995). However, given the nature of the project and its dependence on existing, imperfect, patchy, and incompatible data, statistical validity could only be achieved in a pragmatic way. The key solution that was chosen in the meeting was to integrate the in-situ and earth observation data in such a way that it increased the reliability of the end result. Integration with earth observation data was also the main factor in scaling up the in-situ data to achieve full coverage representations of Europe's biodiversity. This integration was achieved using correspondence matrices as scaling devices. Again, this required statistics as the matrices were produced using statistical intercalibration techniques and procedures.

GLOBALIZING BIODIVERSITY

We have used empirical material from the EBONE project to analyze how biodiversity is globalized in practices of databases and scaling. Our findings demonstrate not only the importance of databases and scaling but also the dilemmas and negotiations involved, in particular: (1) the bootstrapping issues involved in designing databases, (2) the issue of how to achieve statistical validity while using existing, imperfect data, and (3) the issue of how to scale up the data. Our findings also point to an important role of statistics in these practices and the use of correspondence matrices to intercalibrate in-situ and earth observation data as scaling devices.

What is striking is the relative absence of the term biodiversity in the meeting. All the technical talk of data and statistics seemed to have replaced discussing biodiversity itself. The EBONE project addresses biodiversity in an abstract way, relatively detached from the actual species and habitats that make up biodiversity and from the individuals who collect the data upon which EBONE relies. The EBONE project illustrates what was suggested earlier: creating a biodiversity database is as much about collecting knowledge as it is about forgetting (Bowker 2005). The databasing and scaling practices involved in the globalization of biodiversity have an impact on how biodiversity is represented. EBONE's end results take the shape of full coverage, scaled-up, decontextualized maps of Europe's biodiversity, which enable an objective and detached understanding of stock and trends of biodiversity. From a scientific perspective this is perfectly understandable. Detachment and decontextualization are inevitably involved in the globalization of biodiversity. They are absolutely

vital if EBONE's results are to be seen as scientifically valid and if they are to assume a wider usability beyond the direct context in which they were created.

However, this may come with risks. Through the decontextualization, abstraction, standardization, quantification and scaling up it involves, the globalization of biodiversity runs the risk of losing touch with the actual habitats and species that make up biodiversity and the people involved in its recording (Lawrence and Turnhout 2010). Although at this point it is too early to tell, this may also be the case for EBONE in the future. To keep the volunteer recorders and the recording organizations on board, it is important that projects such as EBONE recognize the different motivations of those involved and ensure transparency about what happens to the data. Finally, and perhaps most importantly, to achieve mutual trust and exchange, it will be crucial that these projects are perceived as part of the recording community, rather than as external institutions that are only interested in the data. The challenge is to organize the creation of biodiversity databases and other initiatives to globalize biodiversity in a socially robust way that is open to multiple kinds of data, perspectives, users, and participants.

Responses to this article can be read online at:
http://www.ecologyandsociety.org/vol16/iss1/art35/
responses/

LITERATURE CITED

Bell, S., M. Marzano, J. Cent, H. Kobierska, D. Podjed, D. Vandzinskaite, H. Reinert, A. Armaitiene, M. Grodzinska-Jurczak, and R. Muršic. 2008. What counts? Volunteers and their organisations in the recording and monitoring of biodiversity. *Biodiversity and Conservation* 17:3443-3454.

Bisby, F. A. 2000. The quiet revolution: biodiversity informatics and the internet. *Science* 289:2309-2312.

Bowker, G. C. 2000. Biodiversity datadiversity. *Social Studies of Science* 30:643-683.

Bowker, G. C. 2005. *Memory practices in the sciences.* The MIT Press, Cambridge, Massachusetts, USA.

Bowker, G. C., and S. L. Star. 2000. *Sorting things out, classification and its consequences.* The MIT Press, Cambridge, Massachusetts, USA.

Buizer, M., B. Arts, and K. Kok. 2011. Governance, scale, and the environment: the importance of recognizing knowledge claims intransdisciplinary arenas. *Ecology and Society* 16(1):21. [online] URL: www.ecologyandsociety.org/vol16/iss1/art21/

Bulkeley, H. 2005. Reconfiguring environmental governance: towards a politics of scales and networks. *Political Geography* 24:875-902.

Bunce, R., G. B. Groom, R. H. G. Jongman, and E. Padoa-Schioppa. 2005. *Handbook for surveillance and monitoring of European habitats.* ALTERRA - Research Institute for the Green World, Wageningen, The Netherlands. [online] URL: http://www2.alterra.wur.nl/Webdocs/PDFFiles/Alterrarapporten/AlterraRapport1219.pdf

Bunce, R., M. Metzger, R. Jongman, J. Brandt, G. De Blust, R. Elena-Rossello, G. Groom, L. Halada, G. Hofer, D. Howard, P. Kovar, C. Mücher, E. Padoa-Schioppa, D. Paelinx, A. Palo, M. Perez-Soba, I. Ramos, P. Roche, H. Skånes, and T. Wrbka. 2008. A standardized procedure for surveillance and monitoring European habitats and provision of spatial data. *Landscape Ecology* 23:11-25.

Callon, M. 1986. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. Pages 196-223 *in* J. Law, editor. *Power, action and belief: a new sociology of knowledge?* Routledge, London, UK.

Cash, D. W., N. Adger, F. Berker, P. Garden, L. Lebel, P. Olsson, L. Pritchard, and O. Young. 2006. Scale and cross-scale dynamics: governance and information in a multilevel world. *Ecology and Society* 11:8. [online] URL: http://www.ecologyandsociety.org/vol11/iss2/art8/.

Delaney, D., and H. Leitner. 1997. The political construction of scale. *Political Geography* 16:93-97.

Ellis, R., and C. Waterton. 2004. Environmental citizenship in the making: the participation of volunteer naturalists in UK biological recording and biodiversity policy. *Science and Public Policy* 31:95-105.

Ellis, R., C. Waterton, and B. Wynne. 2010. Taxonomy, biodiversity and their publics in twenty-first century DNA barcoding. *Public Understanding of Science* 19:497-512.

Escobar, A. 1996. Construction nature: elements for a post-structuralist political ecology. *Futures* 28:325-343.

Hajer, M. A. 1995. *The politics of environmental discourse, ecological modernization and the policy process.* Oxford University Press, Oxford, UK.

Halada, L., R. H. G. Jongman, F. Gerard, L. Whittaker, R. G. H. Bunce, B. Bauch, and D. S. Schmeller. 2009. The European Biodiversity Observation Network - EBONE. Pages 177-188 *in* J. Hrebicek, J. Hradec, E. Pelikán, W. Pilmmann, I. Holoubek, and T. Legat, editors. *Proceedings of the European conference of the Czech presidency of the Council of the EU: towards e-environment.* Masaryk University, Prague, Czech Republic. [online] URL: http://e-envi2009.org/proceedings.pdf .

Haraway, D. 1988. Situated knowledges: the science question in feminism and the privilege of partial perspective. *Feminist Studies* 14:575-599.

Hine, C. 2008. *Systematics as cyberscience: computers, change, and continuity in science.* The MIT Press, Cambridge, Massachusetts, USA.

Jasanoff, S., editor. 2004. *States of knowledge: the co-production of science and social order.* Routledge, London, UK.

Jones, K. T. 1998. Scale as epistemology. *Political Geography* 17:25-28.

Jones, R. 2009. Categories, borders and boundaries. *Progress in Human Geography* 33:174-189.

Latour, B. 1987. *Science in action: how to follow engineers and scientists through society.* Harvard University Press, Cambridge, Massachusetts, USA.

Latour, B., and S. Woolgar. 1979. *Laboratory life, the construction of scientific facts.* Princeton University Press, Princeton, New Jersey, USA.

Lawrence, A., and E. Turnhout. 2010. Personal meaning in the public sphere: The standardisation and rationalisation of biodiversity data in the UK and the Netherlands. *Journal of Rural Studies* 26 (4):353-360.

McGoey, L. 2007. On the will to ignorance in bureaucracy. *Economy and Society* 36:212-235.

Meadowcroft, J. 2002. Politics and scale: some implications for environmental governance. *Landscape and Urban Planning* 61:169-179.

Pickering, A. 1992. From science as knowledge to science as practice. Pages 1-28 *in* A. Pickering, editor. *Science as practice and culture.* The University of Chicago Press, Chicago, Illinois, USA.

Porter, T. M. 1995. *Trust in numbers: the pursuit of objectivity in science and public life.* Princeton University Press, Princeton, New Jersey, USA.

Scott, J. 1998. *Seeing like a state: how certain schemes to improve the human condition have failed.* Yale University Press, New Haven, Connecticut, USA.

Secretariat of Convention on Biological Diversity. 2006. *Article 2. Use of Terms.* Accessed September 27, 2010. [online] URL: http://www.cbd.int/convention/articles.shtml?a=cbd-02.

Secretariat of Convention on Biological Diversity. 2007. *What is the problem: the taxonomic impediment.* Accessed September 27, 2010. [online] URL: http://www.cbd.int/gti/problem.shtml .

Shapin, S., and S. Schaffer. 1985. *Leviathan and the air-pump: Hobbes, Boyle and the experimental life.* Princeton Unversity Press, Princeton, New Jersey, USA.

Stone, D. A. 1988. *Policy paradox and political reason.* Scott Foresman and Company, Illinios, USA.

Swyngedouw, E. 2000. Authoritarian governance, power, and the politics of rescaling. *Environment and Planning D* 18:63-76.

Takacs, D. 1996. *The idea of biodiversity, philosophies of paradise.* The Johns Hopkins University Press, Baltimore, Maryland, USA.

Termeer, C. J. A. M., A. Dewulf, and M. van Lieshout. 2010. Disentangling scale approaches in governance research: comparing monocentric, multilevel, and adaptive governance. *Ecology and Society* 15(4):29. [online] URL: http://www.ecologyandsociety.org/vol15/iss4/art29/.

Turnhout, E. 2009. The effectiveness of boundary objects: the case of ecological indicators. *Science and Public Policy* 36:403-412.

Turnhout, E. 2010. Heads in the clouds: knowledge democracy as a utopian dream. *in* R. 't Veld, editor. *Knowledge democracy, implications for science, politics and media.* Springer, Berlin and Heidelberg, Germany.

West, P., and D. Brockington. 2006. An anthropological perspective on some unexpected consequences of protected areas. *Conservation Biology* 20:609-616.

Wilson, E. O. 2000. A global biodiversity map. *Science* 289:2279.

[1] The authors are indebted to Steve Woolgar who introduced the term "scaling device" at the "Scaleography" workshop, Oxford, July 8, 2009.

[2] This refers to the EU Habitats Directive. Annex 1 contains a list of habitat types for which EU member states are obliged to take protective measures.