Deliverable 1.1 (D1.1)
Gap analysis and priorities for filling identified gaps in data coverage and quality
M22

Project acronym:	EU BON
Project name:	EU BON: Building the European Biodiversity Observatio
Call:	ENV.2012.6.2-2
Grant agreement:	308454
Project duration:	01/12/2012 – 31/05/2017 (54 months)
Co-ordinator:	MfN, Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Germany
Delivery date from AnnexI:	M22
Actual delivery date:	M22
Lead beneficiary:	MfN
Authors:	Dr. Florian Wetzel (Task Lead), Dr. Anke Hoffmann, Alexander Kroupa, Günther Korb, Dr. Christoph Häuser (MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany); Prof. Urmas Köljalg, Dr. Kessy Abarenkov (UTARTU, University of Tartu, Estonia); Tim Robertson, Mélianie Raymond PhD, Dipl. Biol. Andrea Hahn, Dr. Donald Hobern (GBIF, Secretariat of the Global Biodiversity Information Facility); Dr. Isabel Calabuig, Lotte Endsleff (UCPH, University of Copenhagen: Natural History Museum of Denmark, Denmark); Dr. Michael Kuhlmann (NHM, The Natural History Museum, London); Prof. Karol Marhold, Matúš Kempa (IBSAS, Botanicky Ustav Slovenskej Akademie Vied, Slovakia); Dr. Lyubomir Penev, Dr. Pavel Stoev (Pensoft Publishers Ltd., Bulgaria); Dr. Nicolas Bailly, Kathleen Reyes FIN, Fishbase

Information & Research Group, Inc., Philippines);
Dr. Fredrik Ronquist (NRM, Naturhistoriska Riksmuseet, Stockholm, Sweden)
Dr. Sarah Faulwetter, Dr. Christos Arvanitidis, Dr. Eva Chatzinikolaou (HCMR, Hellenic Centre for Marine Research, Greece);
Dr. Corinne Martin (WCMC, World Conservation Monitoring Centre, UK);
Dr. Dirk Schmeller, Dr. Jean-Baptiste Mihoub, Dr. Guy Peer, Prof. Klaus Henle (UFZ, Helmholtz Centre for Environmental Research, Germany);
Dr. Lluis Brotons, Dr. Sergi Herrando (EBCC–CTFC, Centre Tecnologic Forestal de Catalunya, Spain);
Anton Güntsch, Dr. Eckhard von Raab-Straube, Andreas Kohlbecker (FUB-BGBM, Freie Universität Berlin, Germany);
Dr. Stefan Stoll, Prof. Peter Haase, Dr. Jonathan Tonkin (SGN, Senckenberg Gesellschaft für Naturforschung, Germany);
Nils Valland, Wouter Koch (NBIC, Norwegian Biodiversity Information Centre, Norway)
Dr. Aaike de Wever (RBINS, Royal Belgian Institute of Natural Sciences, Belgium)
Dr. Quentin Groom (Botanic Garden Meise; Belgium)
Dr. Donat Agosti, Terry Catapano, Jeremy Miller, Guido Sautter (PLAZI, Plazi Inc., Switzerland);

Furthermore:
Dr. Alexander Sennikov and Dr. Pertti Uotila (Finnish Museum of Natural History, Finland),
Dr. Yde de Jong (Netherlands),
Dr. Christian Schmid-Egger (Germany).

This project is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union **Dissemination Level**		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308454.

Content

1 OVERVIEW DELIVERABLE D1.1

1.1 INTRODUCTION

This report aims to assess the relevant data sources on biodiversity on a European and global scale. The assessment particularly evaluates the gaps of available biodiversity information sources and, after outlining the most important gaps, to identify priorities for improving the data availability and to give recommendations of how the gaps can be closed. The gap analysis has a focus on biodiversity information on a European scale and is based on an assessment of current marine, terrestrial and freshwater biodiversity data sources. The gaps are evaluated against the needs of the different stakeholders; this includes the demands from European policy (like the EU biodiversity strategy to 2020), international relevant processes (like the Convention on Biodiversity and the Aichi biodiversity targets) and the scientific community.

There are various requirements for biodiversity data, and the gap analysis aims to check how far these requirements are met in present available data. On the one hand the underlying information needs to meet specific criteria so that it could be used for scientific analysis and on the other hand the data should fulfill the needs for policy makers on a European scale. Thus, within the process of the gap analysis, a set of categories were developed to define under which aspects the available biodiversity sources should be evaluated to assess the quality and coverage of the datasets. The results of the gap analysis will give some needed background information on the usability of data sources and datasets for other work packages of EU BON, particularly for partners that will analyze the data for patterns, processes and trends.

The results of the gap analysis of data sources will not only focus on outlining main gaps in terms of data quality and coverage, but also on drafting recommendations for improving data availability and data access on a European scale. This will help to develop guidelines for European biodiversity information management, as well as for data mobilization efforts and Citizen Science approaches in the project.

1.2 PROGRESS TOWARDS OBJECTIVES

The work of the deliverable is divided into several sections, outlining different aspects of biodiversity information, evaluating the most important questions on biodiversity, the underlying data sources on a European scale and highlighting the most important gaps in general, combined with a detailed analysis of the gaps of some selected data sets. Specifically, this deliverable contributes in fulfilling the following aims of the work package 1:

- Evaluation of relevant information sources and their data characteristics, such as coverage, accessibility, quality, and format, as a basis for a detailed gap analysis with regard to GEO BON needs.
- Identifying gaps of information sources and setting priorities in filling the identified gaps.

1.3 ACHIEVEMENTS AND CURRENT STATUS

The current activities in the gap analysis can be divided into five main sections:

1. Screening of scope and aims of the gap analysis

The first phase of the gap analysis in Task 1.3 was an initial planning phase to outline how the gap analysis should be conducted and what approaches are needed. Some first discussions took place at a meeting of the Informatics Task Group, which was held in Trondheim, Norway (Initial Informatics Workshop, 29–31 May 2013) and in a conference call some weeks later.

2. Finding the target high level questions for biodiversity

The survey "High Level Questions on Biodiversity" was designed to ask the EU BON partners of the relevant work packages on their view regarding the most important questions that a European Biodiversity Observation Network should answer. The high level questions should help to prioritize and structure the EU BON work on gaps in biodiversity datasets. The questions were based on the needs of European environmental policy and what scientists see as the most relevant questions. To determine the most important questions, the online questionnaire was distributed to EU BON partners of the work packages 1,2,3,4 and to some participants from other projects.

In the survey, there were 29 proposed questions, divided in seven sections (see Annex 1). The partners could rank each of the 'high-level questions' regarding biodiversity on a European scale by using a drop-down menu. The importance of the question could be ranked from 5 (highly relevant) to 1 (less relevant). The participants were also asked to evaluate the availability of data (from 5-1, i.e. from very good to poor data quality) and list high-quality data sources that should be integrated in the EU BON gap analysis and that could be potentially used as valuable datasets on a European level. From the highest ranked questions the most relevant questions were chosen and a set of seven target high level questions were defined at the Work Package 1 meeting in Stockholm in January 2014.

Overall, 25 partners of EU BON or related projects participated in the survey. There were 894 specific votes that ranked the relevance of the questions and data availability of datasets. The partners that participated in the survey were from 15 different countries (Belgium, Denmark, Estonia, Finland, Germany, India, Israel, Italy, Norway, Philippines, Portugal, Slovakia, Spain, Sweden, United Kingdom).

3. Assembling a preliminary list of datasets and gaps for biodiversity information on a European scale

Results from the online survey on high-level questions were assembled to obtain an overview of existing data sources regarding biodiversity and environmental variables on a European scale. In turn, an online and freely-accessible spreadsheet was created, that contains a list of biodiversity data sources. The partners could, in turn, add additional datasources to complete the overview of datasources.

4. Conducting an in-depth gap analysis for specific datasets

The specific gap analysis was planned in more detail after evaluating the general data availability for the high level questions and pointing out some general gaps. To streamline the work of the different groups in the gap analysis of EU BON, a work plan was drafted for all partners in the task group (task 1.3). There were different working groups formed for evaluating gaps in different topics, e.g. the gap analysis for distribution data of certain taxonomic groups. Some of the groups focus specifically on analyzing gaps in databases that contain information on species traits, genes or taxonomic information, e.g. Fauna Europaea for animal or Euro+Med for plant species. Each of the partners was supposed to structure their work along the target high level questions.

Specifically the datasets will be analyzed to:

- outline spatial and temporal gaps.
- determine gaps and biases in terms of the taxonomic information
- evaluate the data accessibility: restricted access to the data or unrestricted access.
- check other aspects of data quality: are duplicates removed in the dataset, are recent observations included and does proper metadata information exist?
- outline trends in the accumulation of occurrence data and the integration of historical data.

Some of the datasets will be evaluated according to the above outlined scheme. For some datasets existing evaluations will be used, e.g. the conducted gap analysis on European monitoring schemes (cf. EuMon).

5. Additional surveys

In addition to the outlined activities of the gap analysis there was a further analysis related to the gap analysis. The United Nations Environment Programme - World Conservation Monitoring Centre (UNEP-WCMC) prepared a report that was supported by EU BON. The report assesses the potentials and limitations of remote sensing applications for monitoring trends and changes in biodiversity, particular with regards to the Aichi Biodiversity Targets. The final version was released at the beginning of 2014 and was also part of the Seventeenth meeting of the Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA) of the Convention on Biological Diversity (CBD) meeting in Montreal, Canada and its official documentation (CBD SBSTTA17 Information document number 16, UNEP/CBD/SBSTTA/17/INF/16).

1.4 FUTURE DEVELOPMENTS

The gap analysis provided some important insights on the gaps, limitations and biases of current biodiversity datasets. The results of the gap analysis will be used for the further work of the project, particularly by filling the gaps of datasets due to enhanced mobilization efforts that will be developed within the project (e.g. in Task 1.4). Further specific evaluation of gaps will be conducted within the project time as needed, e.g. for Work Package 6 (Science-Policy Interface). There will be further work conducted under the Work Package 1 with relevance for the gap analysis, specifically regarding the needs of monitoring schemes in Task 1.1. Further findings can be in turn included in the Deliverable D1.1, which should be seen as a living document. Furthermore, publications are planned for disseminating D1.1 findings.

2 OVERALL OVERVIEW OF GAPS AND LIMITATIONS OF BIODIVERSITY DATASETS

2.1 HIGH LEVEL QUESTIONS ON BIODIVERSITY IN EUROPE AND BIODIVERSITY DATA

2.1.1 Introduction

The gap analysis started with the survey "High Level Questions on Biodiversity" to develop guidelines for determining the gaps of biodiversity data, both on a general level (Chapter 2) as well as on a specific one (Chapter 4). The survey was designed to ask EU BON partners of the relevant work packages their views on the most important questions that a European Biodiversity Observation Network should answer. The high level questions should furthermore help to prioritize and structure the EU BON work on gaps in biodiversity datasets. As a start, 29 high level questions for biodiversity were determined as a baseline to see whether the recent biodiversity datasets are able to cover the information for answering these questions (Annex 1). The questions were based on the needs of European environmental policy and what scientists see as the most relevant questions.

In the survey, there were 29 proposed questions, divided in seven sections (numbers in the parentheses reflects the number of questions):

- Species and habitats (8)
- Ecosystems, biodiversity and their functions (3)
- Ecosystems and their services (5)
- Sustainable land-use and use of freshwater systems and oceans (2)
- Protected areas (3)
- Drivers of change (5)
- Invasive and biodiversity (3)

The partners were asked to rank each of the 29 questions. First, the participants ranked the scientific and political relevance of the question (from a score of 5, highly relevant, to 1, less relevant). In a second step, the availability of the data was ranked (5; very good availability, 3: fair availability, 1: poor availability).

It is recognized that this questionnaire reports on the opinions a fairly narrow group of stakeholders and that other interest group might have different opinions. Nevertheless, we believe that the largely academic background of the respondants will reflect the plurality of views amoung all stakeholders.

2.1.2 Key findings: The most and least important thematic sections, data availability and gaps

Regarding the thematic sections the highest score was given to the section "invasive species and biodiversity" (4.2 points out of 5), the second most important section was the one dealing with "ecosystems and their services". The sections containing questions regarding sustainable agriculture and forestry, status and trends of species and drivers of change got lower rankings (see Fig. 1, Table 1). Lowest relevance was given to the section with questions on ecosystems, biodiversity and their functions and protected areas/biodiversity. Data availability seems generally to be limited, as no section had a data availability of good or very good. Data availability was ranked as "fair" in most cases. However, some variation exists and at least some sections have slightly better datasets available. The thematic section with the highest ranked relevance ("invasive species and biodiversity") seems to have the

best data availability (value of 3.3). Data seems to be limited particularly for questions on "ecosystems, biodiversity and their functions" as they received the lowest rating (2.0 points, see Table 1), probably also because a wider range of thematic data is needed to answer this questions, compared to the invasive species where mainly distribution and some trait is needed to answer the questions.

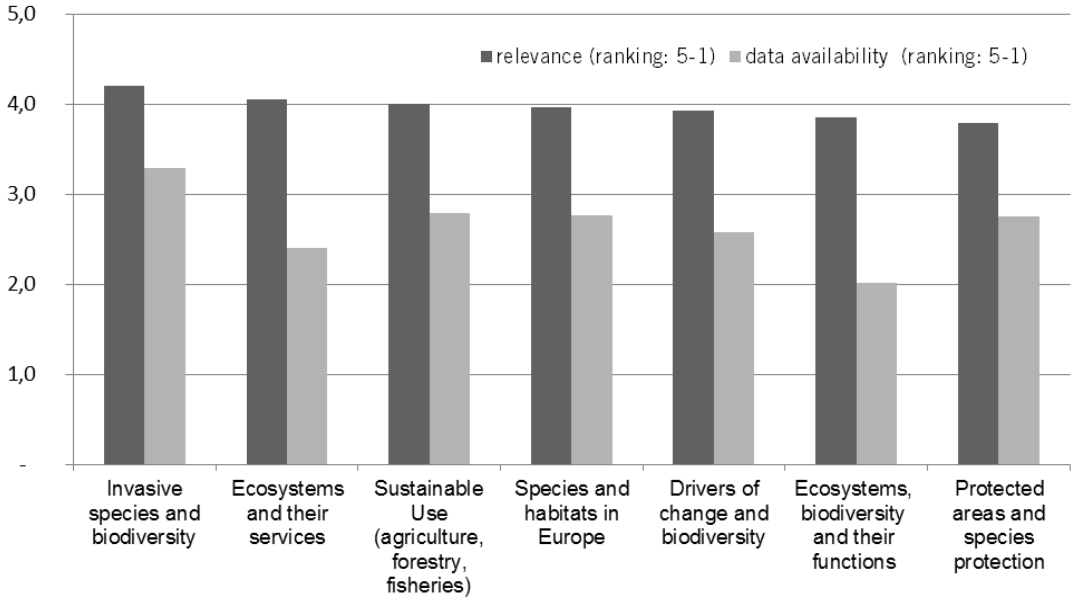


Fig. 1: Thematic Sections and ranking of relevance and data availability, results are based on the input of 25 EU BON partners and associated specialists (listed according to the ranking of relevance, ranking score from 5; high relevance/data availability to 1: low relevance/data availability).

Table 1: Thematic sections and ranking of relevance and data availability.

Thematic section	relevance (ranking: 5-1)	data availability (ranking: 5-1)
Invasive species and biodiversity	4.2	3.3
Ecosystems and their services	4.0	2.4
Sustainable Use (agriculture, forestry, fisheries)	4.0	2.8
Species and habitats in Europe	4.0	2.8
Drivers of change and biodiversity	3.9	2.6
Ecosystems, biodiversity and their functions	3.9	2.0
Protected areas and species protection	3.8	2.8

(listed according to the ranking of relevance, green: high relevance – red: low relevance, ranking score from 5; high relevance/data availability to 1: low relevance/data availability).

Data on ecosystem services in general seems to have quite large gaps; also the section "ecosystems and their services" received a rather low ranking regarding data availability (2.4 points). In the middle field regarding data availability are three sections that were ranked with "fair" data availability: for questions on sustainable land-use, species and habitats in Europe, protected areas and species protection.

However, the ranking according to the different sections reveals a first direction for the thematic focus regarding EU BON and relevant biodiversity topics on a European scale. However, in some of the thematic sections, particularly the ones with a larger number of questions, there is considerable variation in the ranking of relevance and data availability. Thus, the next chapter evaluates the ranking of the questions itself and lists them according to their relevance and data availability.

2.1.3 The most and least important questions and data availability

To gain further insight on the most relevant biodiversity topics, a question-specific analysis was conducted. The 29 questions were ranked according to the average value of the question for relevance, data availability and combined index. Table 2 shows the ranking according to the combined index, which rates the research questions higher if they are highly relevant, but we lack the data to answer them.

The 10 questions that show the highest combined score are mostly related to ecosystems and their services. The highest ranked question deals with the interactions between degradation and the loss of biodiversity and ecosystem services. A further relevant question, where data availability is limited, are future scenarios of ecosystem services and what effects global change processes have, the question of how biodiversity and ecosystems are linked to human health and how biodiversity can improve human health. Another question with rather large gaps is on the relationship between species diversity and ecosystem functions and also questions that involve protected areas and their current state. Other questions with high rankings in the combined index relate to land-use, for example whether there is a measurable improvement in the conservation status of species and habitats due to sustainable land-use and environmental-friendly management plans.

At the same time, it is also interesting to evaluate which questions have a rather low combined index, which means questions that have a lower relevance and/or better data availability. From the group of the 10 questions with the lowest ranking there are some for which sufficient datasets are available, such as on alien invasive species, for example the question on whether priority invasive species and their pathways are sufficiently identified. Other questions with a relatively good data basis are on status and trends of species and on how the preservation of European protected areas positively affects biodiversity. Data with a relatively low relevance are questions like on the genetic diversity of species, protected areas and their carbon storage capability and the question of how a changing European demography and economic activities will affect species (via the human footprint) in a temporal and spatial perspective.

Table 2: A ranking of relevance of the high level questions on biodiversity (ranking according to average rating of the relevance of the question, value range from 5-highest to 1-lowest), ranking of the data availability (5-good data availability, 1-low data availability) and a combined index. For the combined index, highest ranked data shows a high relevance of the question and/or low data availability. i.e. relatively large gaps (max. value of 10). The index is calculated as sum of 'relevance' and the inverse value of 'data availability'.

	Question	relevance	data availability	Combined score r(elevance x inverse data availability)
High relevance x large Gaps	Interaction degradation/biodiversity loss and ecosystem services	4,50	2,23	7,27
	Biodiversity and resilience of ecosystems	4,32	2,07	7,24
	Future scenarios of ecosystem services	3,88	1,75	7,13
	Effects of global change drivers and interactions	4,28	2,38	6,89
	Status of European ecosystems and services	4,53	2,64	6,89
	Biodiversity/ecosystems and human health	3,41	1,73	6,68
	Sustainable land-use and species conservation	4,33	2,70	6,63
	Mapping and modeling of biodiversity	4,09	2,47	6,62
	Relationship species diversity and ecosystem functions	3,82	2,25	6,57
	State of marine and terrestrial protected areas	4,17	2,67	6,50
Medium relevance	Biodiversity loss / extinction	4,43	2,94	6,48
	"Umbrella" protection of species	4,23	2,80	6,43
	Effects of land-use on biodiversity	4,18	2,75	6,43
	Identification of important drivers	4,22	2,86	6,37
	Human footprint and Biodiversity	3,53	2,29	6,24
	Fragmentation of species population	3,86	2,68	6,18
	Ecosystem resilience and restoration/conservation	3,50	2,38	6,13
	Impacts of subsidies	3,59	2,50	6,09
	Invasive Species in Europe	4,21	3,19	6,02
	Taxonomy and biodiversity	4,14	3,13	6,02
Low relevance	Alien species and Biodiversity	4,22	3,29	5,94
	Genetic diversity of species	3,09	2,20	5,89
	Protected areas and carbon storage	3,13	2,25	5,88
	Changing European demography / Economic activities/Footprint and Biodiversity	3,43	2,57	5,86
	Current status and trends of species	4,27	3,42	5,85
	Priorities for ecosystem restoration	3,81	3,00	5,81
	Invasive Species and their pathways	4,18	3,38	5,79
	Mapping of sustainable land-use + threatened species	3,67	2,89	5,78
	Protected areas positively affect biodiversity	4,06	3,33	5,73

2.1.4 Ranking: Data availability and gaps

After evaluating the combined index, we focus on the data availability. The average rating regarding data availability for the data that will be needed to answer the high level questions on biodiversity was 2.6 points (out of a possible 5 points), a value that shows that there are useful datasets on a European scale, but, at the same time, some major limitations in availability of biodiversity information still exist. As there are different datasets needed to answer the specific questions, there is also a quite considerable variation in terms of data availability (see also Table 2), which varied between 1.7 and 3.4 points, i.e. the participants of the online questionnaire ranked the data availability between "low" to "fair".

The table below (Table 3) shows a list of questions with a relative high (> 4 points) or low data availability. As the analysis shows, there are questions for which the data availability was ranked quite low. For example, data that could determine how different drivers of global change interact seems to be quite limited. Also datasets seems to be sparse that are needed to answer questions on the human footprint on biodiversity, i.e. the question of how human consumption and changes in the consumption patterns influence species and their habitats.

Table 3: High level questions with a rather good (average data availability rating > 4 points) or low data availability (ranking ranged from a very good to low availability).

High Ranked Data Availability	**Low Ranked Data Availability**
Current status and trends of species	Effects of global change drivers and interactions
Invasive Species and their pathways	Ecosystem resilience and restoration/conservation
Protected areas positively affect biodiversity	Human footprint and Biodiversity
Alien species and Biodiversity	Relationship species diversity and ecosystem functions
Invasive Species in Europe	Protected areas and carbon storage
Taxonomy and biodiversity	Interaction degradation/biodiversity loss and ecosystem services
Priorities for ecosystem restoration	Genetic diversity of species
Biodiversity loss / extinction	Biodiversity and resilience of ecosystems
Mapping of sustainable land-use + threatened species	Future scenarios of ecosystem services
Identification of important drivers	Biodiversity/ecosystems and human health

In addition to the above-mentioned examples, a lot of issues connected to ecosystem services have currently no sufficient datasets on a continental scale. For example, data seems to be scarce for the interaction between ecosystem services, biodiversity loss and land degradation or regarding biodiversity and resilience of ecosystems. Also, according to the specific voting of the participants of the online survey, there is little knowledge on likely future ecosystem service change and related scenarios. Furthermore, limited available data exists on genetic diversity of species or on how biodiversity and intact ecosystems are linked to human health and how biodiversity can possibly improve human health.

However, a detailed gap analysis is also needed for those questions where it seems that good datasets exist. An example is the question "current status and trends of species" and the suggested datasets for answering the question. As an evaluation of available European data shows, there are datasets available based on the EU Article 17 reporting of the habitats directive (EC 1992), where member states are requested to undertake surveillance of habitats and species considered to be of community interest. But a closer evaluation of the available

data shows that there is an insufficient temporal coverage, as the reporting period is only every 6 years and is restricted to species of the habitats/birds directive.

The low ranking of some questions for data availability in part reflects the difficulty of answering some types of question with the available data. Some questions such as those on genetic diversity and ecosystem services can only be addressed using comparatively time consuming and technical methods. In contrast, many questions on the abundance and distribution of species can be answered using observation data that relatively easy to collect.

Large data gaps may force the development of new techniques that answer the same questions using different methods. For example, some of the questions on ecosystem services that have traditionally answered through terrestrial surveys can now be addressed at a continental extent using remote sensing and only limited groundtruthing. In the field of molecular genetics, so called environmental DNA and next generation sequencing could be used to address the gaps in genetic data.

2.2 EUROPEAN DATASETS, ESSENTIAL BIODIVERSITY VARIABLES AND GAPS

After determining the high-level questions, we generated a list of relevant biodiversity data sources on a European scale and evaluated how these datasets can possibly be used for generating Essential Biodiversity Variables (EBVs). The EBVs cover the thematically most relevant aspects of biodiversity data and can be used in turn to answer high level questions on biodiversity. Fig. 2 shows, for a list of sample datasets, for which EBV classes the data sources supply data. For the sample datasets, only data sources with unrestricted access were used or data with partly unrestricted access (i.e. data owners need to be asked in written form before data can be used).

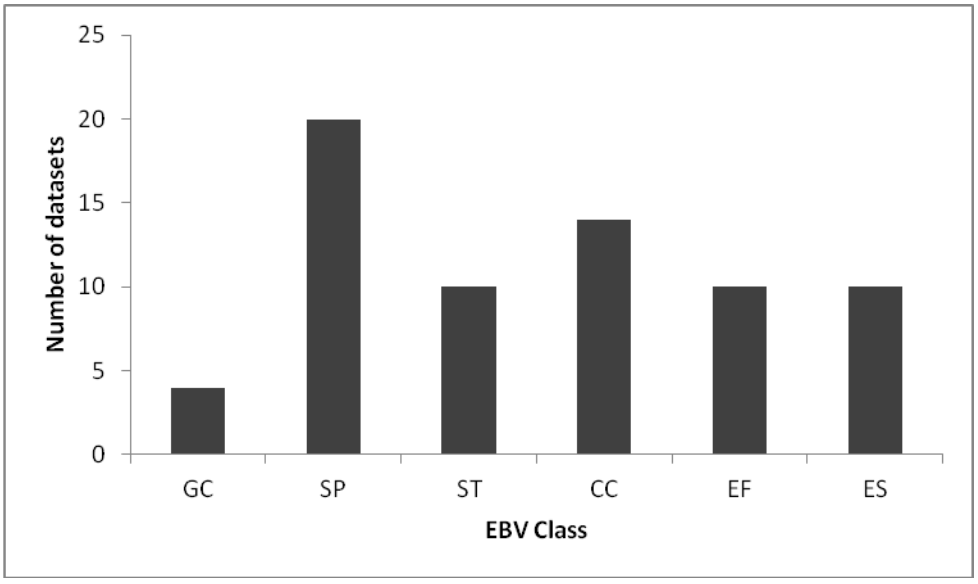


Fig. 2: Essential Biodiversity Variable classes and number of available biodiversity datasets that could be used to contribute data for a specific EBV class. The figure is based on sample set of selected European data sources. (EBV Classes Abbreviations: GC: Genetic composition, Species populations: SP, Species traits: ST, Community composition: CC, Ecosystem function: EF, Ecosystem structure: ES).

As the analysis shows, most of the large European datasets that we have included in the analysis contain data that can be used for analysis of questions related to the EBV class "species populations". This means the datasets include data for example on species distributions, population abundances or -structure. The most obvious gaps regarding biodiversity data on a European scale exist for the EBV class "genetic composition". Some genetic databases do exist (see below in the next chapter), however, as our evaluation in

section 3.9 shows, the spatial and temporal coverage is mostly not sufficient for continent-wide analyses.

The data for the other EBV classes are ranked on a medium level, however, as the more in-depth analysis in the following chapter will show, there are also some significant gaps in community composition and ecosystem function datasets.

2.3 A FIRST OVERVIEW OF GBIF, PESI, INSD AND DATAONE DATA

As a first approach, we present here an overview, comparison and outline of gaps of three important data providers: GBIF for species specimen and observation data, PESI for taxonomic and INSD for genetic data.

Specimen and observation based taxon occurrences were analyzed using GBIF (The Global Biodiversity Information Facility)[1] datasets. The International Nucleotide Sequence Database Collaboration (INSDC: GenBank, ENA, DDBJ)[2] datasets cover DNA sequence data, The Pan-European Species directories Infrastructure (PESI)[3] is used for the analyses of taxonomic backbone data, and the Data Observation Network for Earth (DataONE)[4] covers ecological data including LTER and Dryad datasets. In addition third party annotated INSDC datasets of UNITE[5] were used in the analyses.

The summary of the analyses on the number of species in Europe is shown in Table 4 and Figs. 3-5. Data are shown on Kingdom level and separately also for the birds (Aves), insects (Insecta) and mammals (Mammalia). PESI dataset provides a checklist of European species, which is compiled by the taxonomy experts. The data on the presence of species is not accompanied with specimen or other types of taxon occurrences. On the contrary all European species in GBIF and INSDC datasets are based on specimen/observation or DNA data respectively. Therefore the plain expectation would be that the PESI checklist is much more complete and outnumber the two other datasets.

Table 4: Number of species in Europe based on taxon occurrences and current names in databases.

Taxon name	GBIF Specimen	GBIF Observation	GBIF Total	PESI checklist	INSD DNA
Animalia	76932	39536	85225	157142	28623
Aves	1323	1469	1932	833	1488
Insecta	42650	24451	46361	95981	12367
Mammalia	771	488	894	298	1144
Plantae	54129	16454	56088	27306	21156
Fungi	31610	11950	32867	21751	10262
Bacteria			2488	137	3512
Archaea					83

[1] http://www.gbif.org
[2] http://www.insdc.org
[3] http://www.eu-nomen.eu
[4] http://www.dataone.org
[5] http://unite.ut.ee

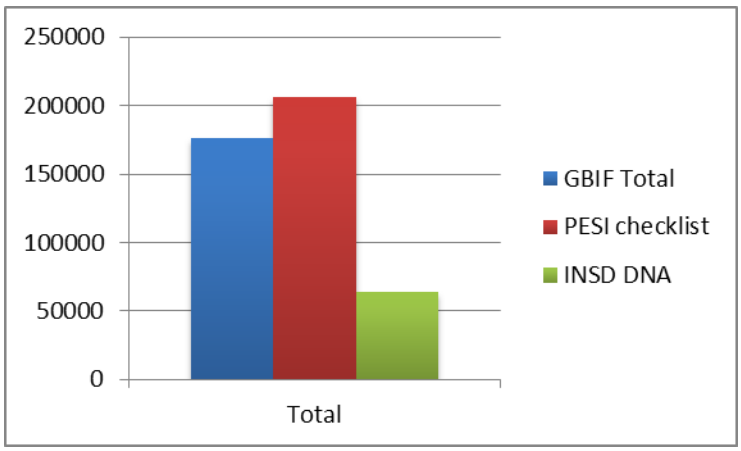


Fig. 3: Number of species in Europe based on GBIF taxon occurrences (specimens and observations), PESI checklist of European species names and INSD species with DNA sequences

For all taxa this is true (see Fig. 3) also for the animals (Fig.4) but not for plants, fungi and bacteria. This can be at least partly explained by the synonymic or defective names used for the specimen and observational data in GBIF datasets. However this should be analysed further to see how to improve our current knowledge on species present in Europe. It is also important to find out why GBIF data on animals do not outnumber PESI checklist but plants, fungi and bacteria do. Plants are certainly easier to study and they are maybe better represented in specimen datasets, however, this is probably not the case for the fungi and bacteria that are presumably much less studied than plants and animals. The number of species in Europe where DNA data are available is much lower compared to GBIF and PESI datasets (Fig. 3 and 4). But this is mostly because of large differences in the number of insect species. There are also major exceptions. Mammals and bacteria have higher species number based on DNA than PESI and GBIF datasets. These discrepancies should be investigated in future studies. INSD datasets probably include also synonymic species names.

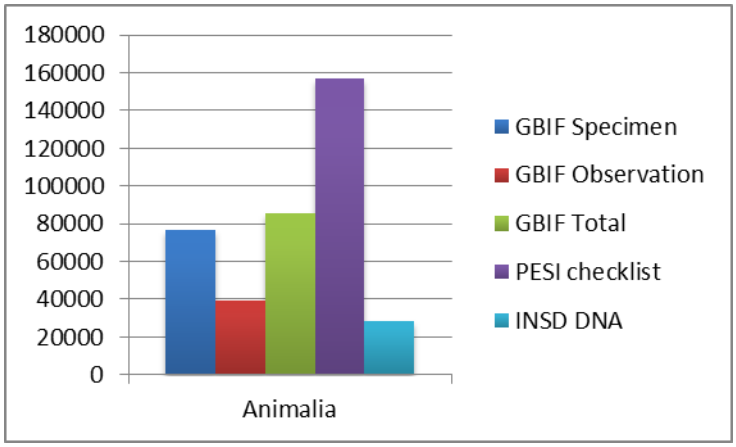


Fig. 4: Number of animal (Animalia) species in Europe based on GBIF taxon occurrences (specimens and observations), PESI checklist of European species and INSD species with DNA sequences

Fig. 5 shows the number of European species occurrence data in GBIF, split between kingdoms and major classes of Animalia. It is obvious that most data in GBIF are for animals and plants. And most animal data are actually bird (Aves) observations. The same basic structure of the data is visible for the global occurrence data. Based on these data it is clear that many taxa like insects, fungi and bacteria are much less studied compared to other animals and plants.

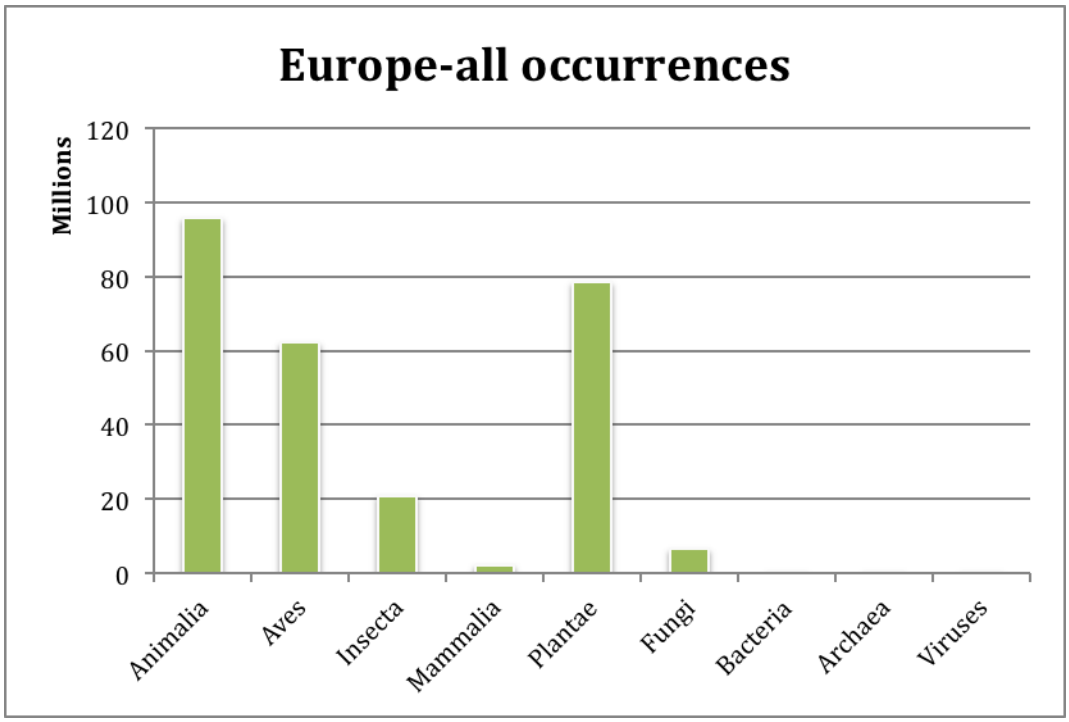


Fig. 5: Number of all taxon occurrences in Europe based on GBIF datasets (specimens and observations).

We also analysed DataONE data. However, only the metadata of the datasets were analysed, as the datasets are not directly available. The datasets are submitted in different file formats and before analyses of the data can take place, the data has to be imported into a common format. EU BON has neither the aim nor the resources to analyse the underlying datasets itself, therefore this first overview analysis has limited results. Metadata of the DataONE datasets include only modest information on the taxa involved. We analysed 144,198 datasets and found that metadata of nearly 14,000 datasets has information on taxa. This is approximately 10% of the analysed datasets. The number of datasets where metadata include species level information is much lower (Fig. 6).

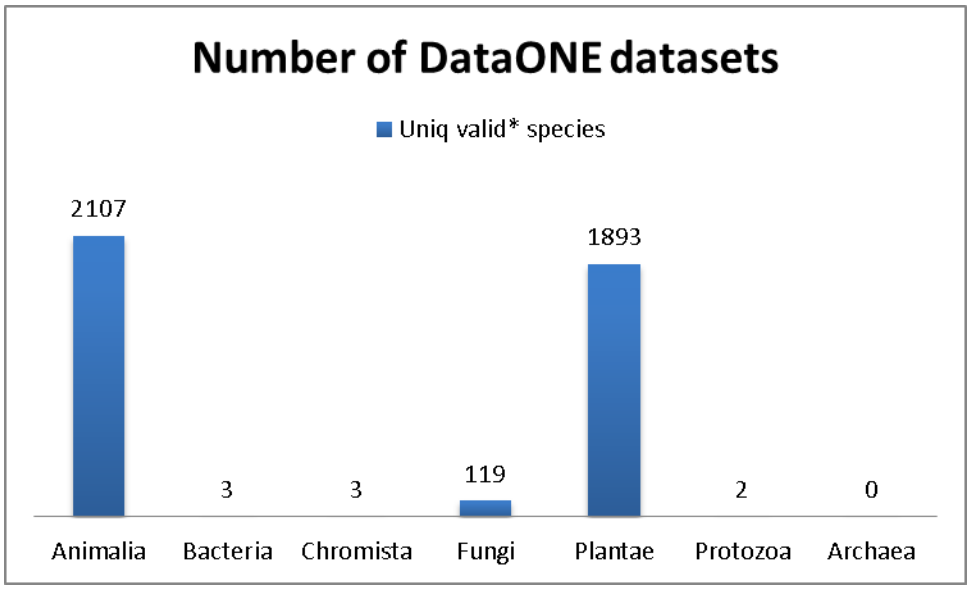


Fig. 6: The number of DataONE datasets with unique valid species name in the metadata file

We also analysed the content of genetic databases (INSD, UNITE). Fig. 7 demonstrates the potential of DNA data for biodiversity analyses on a species level. There are less than 30,000 fungal species names in INSD accompanied with DNA data. However, most DNA data have no full species names because they do not come from voucher specimens but from samples like soil, air, water, skin, etc. These sequences can be identified on species level only if there are fully identified sequences already available. This is usually not the case for most taxa, including bacteria, protists as well as for many animals and plants. This problem can be resolved in part by DNA analyses whereby sequences are clustered into species. This is possible for specific genes where a species threshold value (e.g. similarity) is accepted. In Fig. 7, the number of sequence based fungal species based on INSD/UNITE dataset is much higher than the number of species names behind DNA data. UNITE made these sequence based species hypotheses available for the environmental analyses by giving them stable identifiers. It means these species can be communicated even when full species name is not available. So in DNA data, there is an obvious gap in sequenced based species and UNITE species names.

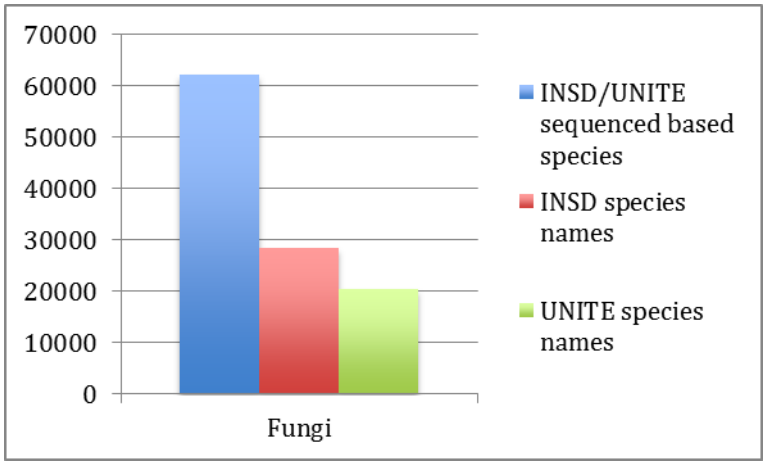


Fig. 7: Number of fungal species on the Earth based on: 1) sequence similarity analyses; 2) unique species names in INSD datasets; 3) unique species names in UNITE datasets.

3 KEY FINDINGS - OVERVIEW OF MAIN GAPS OF THE SPECIFIC ANALYSIS OF GLOBAL AND EUROPEAN DATASETS AND RECOMMENDATIONS

After outlining some general gaps of biodiversity data in Europe, we also conducted a specific gap analysis for a selection of several main global and European datasets. The datasets represent some main sources for biodiversity data, either for specific realms (terrestrial, marine, freshwater), taxonomic groups, thematic fields (taxonomy, genetic databases) or networks of European test sites (LTER).

Here we present a short overview of the main gaps of datasets, compiled from Chapter 4. Specific gap analyses were conducted for a whole set of different databases (see Box 1 below, also for the Acronyms of the datasets that will be mainly used from now on). For the authors and EU BON partner institutions that contributed to the analyses, please see Annex 2.

Box1: Overview of databases where an in-depth gap analysis was conducted, acronyms and URL (see the overview below the box and a more detailed analysis in Chapter 4)

General Biodiversity Data

- Data provided by the Global Biodiversity Information Facility (GBIF) http://www.gbif.org/
- European Monitoring schemes (EuMon) http://eumon.ckff.si/index1.php

Data for specific taxonomic groups

- Atlas Florae Europaeae (AFE) http://www.luomus.fi/en/atlas-florae-europaeae-afe-distribution-vascular-plants-europe
- European Vegetation Archive of European Vegetation Survey (EVS) http://euroveg.org/eva-database
- FishBase www.fishbase.org
- Polytraits for annelids (polytraits.lifewatchgreece.eu),
- Ocean Biogeographic Information System (OBIS) http://www.iobis.org/
- Marine and coastal data holdings of the UNEP World Conservation Monitoring Centre (UNEP-WCMC) http://data.unep-wcmc.org/
- Checklist of Western Palaearctic Bees http://westpalbees.myspecies.info/
- European Union data on Habitats Directive Article 17 http://www.eea.europa.eu/data-and-maps
- Atlas of the European Bees http://www.atlashymenoptera.net/
- Atlas of European Breeding Birds http://www.ebcc.info/
- Pan-European Common Bird Monitoring Scheme (PECBMS) http://www.ebcc.info/pecbm.html

Genetic datasets

- International Nucleotide Sequence Databases (INSD) http://www.insdc.org and other DNA databases: DNA Data Bank of Japan (DDBJ) http://www.ddbj.nig.ac.jp, European Nucleotide Archive (ENA) http://www.ebi.ac.uk/ena/, The National Center for Biotechnology Information (GenBank) http://www.ncbi.nlm.nih.gov/genbank/)
- UNITE community (Database on Genetic Data) http://unite.ut.ee/

Taxonomic Data:

- Fauna Europaea (FaEu) http://www.faunaeur.org/
- Euro+Med PlantBase (E+M) http://www.emplantbase.org/home.html
- PESI http://www.eu-nomen.eu/portal/

Environmental Data:

- Long Term Ecological Research Network (LTER) http://data.lter-europe.net/deims/

Further important datasets mentioned/discussed:

WORMS, Catalog of Fishes (CofF), Trawlbase, Catalogue of Life, Sealifebase (www.sealifebase.org), TRAITBANK-EOL, Biofresh-Datasets, Datras, Delivering Alien Invasive Species Inventories for Europe (DAISIE), Global Invasive Species Information Network (GISIN) others.

Here we summarise some of the main findings of the more detailed gap analyses in Chapter 4. This overview of gaps is divided thematically in five sections, namely in the subsections (a) spatial gaps, (b) temporal gaps, (c) taxonomic gaps, (d) other types of gaps and (e) data availability.

3.1 SPATIAL GAPS

As the gap analysis shows, there are some specific gaps in biodiversity data regarding the spatial coverage. Here we outline some of the most important spatial gaps. In the specific analysis, the data were evaluated regarding the resolution of the data and the spatial coverage. Further questions on the spatial aspects of gaps were to determine whether presence data exists or if also absence data is available. Another important aspect was to analyse for all records of the biodiversity datasets whether spatial information exists and to determine the accuracy and precision of the georeferenced information.

General Biodiversity Data: GBIF-mediated data

- For data served by the Global Biodiversity Information Facility (GBIF), most data is available for the GBIF participating countries (e.g. 37 voting countries, 15 associate countries) whereas some non-participating countries have less data available. Non-participating countries can be found particularly in Africa, in the Middle East and Asia and some in South America.
- 3% of GBIF records lack any location, however, even when point locations are given, this information sometimes contains inaccuracies. For example, attention is needed to ensure the accuracy of values and to review or improve accompanying metadata, e.g. to clarify where coordinates indicate centroids or corner points of grid cells in a monitoring scheme rather than actual geo-location of the locality of occurrence.
- For GBIF records it is also relevant not only to determine that records for a species exist, but also to determine the geographic coverage for the recorded species.
- For many species there are only a small number of records available with geographic coordinates. More data mobilisation efforts are needed to close the gaps. The more specific gap analysis shows mobilization efforts in GBIF data for recent years and outlines for which regions/countries more data is needed.

Specific Taxonomic Groups: Data on plant species

- Spatial gaps in data coverage could be detected in all of the plant datasets analysed, particularly for Eastern Europe, i.e. the Russian Federation and other Eastern European countries such as Belarus, Bulgaria and Hungary but also for some Western European countries, e.g. Italy.
- There is a large bias in the recording effort across Europe. While the counrties of Scandinavia, Western Europe, and Greece are well covered, many Central European, and most of East and South-East European countries are covered only poorly or not at all.
- The European Vegetation Archive of European Vegetation Survey (EVS) provides considerable amount of data for Austria, Germany, Netherlands, Czech Republic and Slovakia and some data for other Central European countries. Gaps are in other countries are obvious, for example Belarus, Moldova and the Caucasian countries.
- Overall, Atlas Florae Europaeae (AFE) distribution maps show the presence of the taxa over the whole European continent with only minor spatial gaps. The resolution is limited with point locations for 50 x 50 km squares and the data does not contain time-series.

Specific Taxonomic Groups: Data on marine species

- The Ocean Biogeographic Information System (OBIS) is an important source for data on marine species. However, only half of the estimated 115,000 valid species that have occurrence data in OBIS have more than three (occurrence) points.
- Marine Data: Some databases contain spatial information (polygons) that overlap – such overlaps have to be excluded. For example, overlaps have to be excluded in species distribution datasets of the International Union for Conservation of Nature (IUCN). Also, in most cases datasets contain only information on presence data and not absence data. This will generate bias as survey effort is uneven and usually unknown. This can lead to both under and overestimates of the likelihood of a species' presence.

Specific Taxonomic Groups: Example of pollinator species – wild bees in Europe

- There are various sources for data on wild bee species in Europe. As the comparison on distribution data shows (see chapter on specific gap analysis, assessment of pollinator data), GBIF contains quite a large amount of data for countries of Scandinavia, Western- and Central Europe. Less covered regions regarding bee species occurrence records are Eastern European countries, the Caucasus region, the Balkans and Turkey. Additional expert datasets like the Checklist on Western Palearctic bees (http://westpalbees.myspecies.info/) or the Atlas of European bees (http://www.zoologie.umh.ac.be/hymenoptera/default.asp) can contribute additional important information on the distribution of bee species.
- The analysis also shows some significant data gaps regarding bee species occurrence in European countries, as particularly exemplified in the study for Denmark. Well covered countries are the United Kingdom, Ireland, Sweden and Germany whereas poorly covered countries are for example Latvia, Albania, Montenegro or Moldova.

Metadata on European Monitoring schemes – the EuMon database

- EuMon, a metadatabase on European animal monitoring schemes, also shows considerable gaps regarding the spatial coverage of monitoring schemes. Less mean taxonomic coverage was achieved for several countries, e.g. Greece, Croatia, Portugal or Romania. The highest mean taxonomic coverage was achieved in Poland, Germany, Estonia or the Netherlands. Please note thet these results are on the basis of the number of monitoring schemes, which doesn't have to correlate with the amount of records generate ot the effort put in to surveying.

Example for Monitoring data: Bird monitoring data from the European Bird Census Council

- In general, bird data, compared to other taxonomic groups, are among the best available datasets on a European scale. However, in the Atlas of European Breeding Birds, based on well-established monitoring schemes, Turkey, Cyprus and the Canary Islands are not covered. Spatial gaps with no data exist particularly for parts of Russia, parts with incomplete coverage for Belarus, Ukraine and Romania. Some rather small areas with incomplete coverage can be also found in Spain and Italy.
- The resolution of the Atlas of European breeding birds offers data in a 50×50km square resolution, which limits its applicability for some scientific analyses.

European Taxonomic data

- Taxonomic information is available for most European countries. However, there are some countries and regions which lack sufficient information. For example the faunistic taxonomic database Fauna Europaea covers European countries except for the Caucasus region. Also in the floristic taxonomic database Euro+Med the Caucasus and Near East are not fully covered.

European research sites – the LTER site network as an example

- The Long Term Ecological Research (LTER) network, a network of field sites which collect environmental data, consists mainly of terrestrial sites. In its current form, marine environments are the most under-represented domain. Sites are most dense in Central Europe and the United Kingdom. Continental, Atlantic, Mediterranean, and alpine regions are the best represented regions, while nemoral, boreal and northern alpine regions, are underrepresented compared to their area.
- The United Kingdom and Italy have the largest number of LTER sites compared to the rest of Europe. In the Netherlands and Belgium, national LTER networks have been started only recently and consist of fewest sites. Another bias is the spatial coverage of the number of project objectives / research topics at each LTER site, where some sites focus on a few research objectives whereas others incorporate a huge variety of research topics. Urban areas represent a clear gap in the LTER network and thus should be a focal area in the future.

Genetic Datasets: Example of fungal species

- The International Nucleotide Sequence Database (INSD) contains nucleotide sequence data. An analysis (see Chapter 4) of a quality filtered fungal test dataset of 276,898 sequences shows that still quite many records do not contain georeferenced information. Despite the fact that 65.7% of the sequences have a country of origin, only 15.9% have geographic coordinates specified.
- INSD shows a sampling bias towards North-America and Europe, with South-America, Australasia and Africa being clearly under-represented.

3.2 TEMPORAL GAPS

For determining temporal gaps, datasets were evaluated according to their temporal coverage (year or time periods). Long-term datasets are needed to determine trends in species ranges and their populations, so the temporal gap analysis also included an assessment of long-term datasets, in case data was continuously recorded over the years. Additionally, other temporal aspects of the datasets were evaluated, for example the number and intervals of collection events or whether detailed information on the year, month and day of the collection event were documented. In the following chapter, some of the most obvious temporal gaps and limitations of the datasets are outlined:

General Biodiversity Data: GBIF-mediated data

- There are still quite some significant temporal gaps in the datasets existing. However, due to mobilisation efforts, gaps in the data can be closed. For example, the relative proportion of records with a missing or incomplete occurrence date has decreased from about 30% in 2008 to 20% in 2014.

Specific Taxonomic Groups: Data on plant species

- A particularly interesting question is whether long-term datasets for plant data exist. However, as our analysis shows, there are only few long-term datasets available for a quite limited number of species, particularly when looking for studies which span a large area. Atlas Florae Europaeae (AFE) data shows no time series regarding the distribution of plants, only some time series are available for the European.Vegetation Survey (EVS) data.

Specific Taxonomic Groups: Data on marine species

- In general, data for evaluating long-term trends in marine species are scarce and only for some species datasets are available that show trends in distribution over decades and that allow long-term studies (or even studies of phenology). This applies also for marine data of UNEP-WCMC where most datasets show data at a given point in time and only few datasets cover a longer time period (e.g. the dataset on the mean sea surface productivity in December 2003-2007).

European research sites – the LTER site network as an example

- Most LTER sites offer datasets and records with a temporal coverage of one to several years. However, there are some LTER sites that have operated for longer time periods. In fact, 48 sites have been operating for 50 or more years and 12 have been operating for over 100 years.

Other datasets analysed

- Also for the other datasets that were analysed, long-term datasets are rare and, if available, limited to some taxonomic groups (birds, butterflies).

3.3 TAXONOMIC GAPS

To determine the completeness and coverage of available biodiversity information, a further step was to evaluate for which taxa information and datasets exists. For example, it was evaluated whether data was collected for a certain limited set of species or for a broad set of taxonomic groups. If the metadata of the datasets allowed a more thorough investigation, the taxonomic coverage in comparison to the complete number of species in the taxonomic group was determined.

General Biodiversity Data: GBIF-mediated data

- There are quite significant differences for which taxonomic group's data are available, not only on a global but also on a country level. There are quite large differences of the taxonomic coverage among countries globally and thus a different kind of taxonomic bias in a country. Sweden for example mostly contributes animal records (around 80%) whereas Japan contributed proportionatly more records on plants.
- On a global level, occurrence records are accessible for around 40% of animal species through GBIF. GBIF occurrence data have, for example a strong bias towards bird species due to the large amateur community and the rapid web publication of such data in recent years. Analysis of species richness is impacted by uneven taxonomic resources. Significant gaps exist for example in catalogues of molluscs, beetles, algae, and some groups of higher plants, as well as for fossil species.

Specific Taxonomic Groups: Data on plant species

- There are still taxonomic gaps in plant datasets existing. For example 20-25% of the European vascular plants are covered in the AFE database, which is a quite considerable effort, given the large number of species in the different families. However, there are still many species to be included in the assessment.
- The taxonomic coverage of the data largely varies according to plot datasets in the plant database EVS. Furthermore, there is no overall information on taxonomic coverage available for the datasets available.

Specific Taxonomic Groups: Data on marine species

- The marine data shows also some large gaps with regards to the species that are covered. For example OBIS covers only 230,000 marine species out of 221,000 species currently listed in WoRMS. However, some taxonomic group are better covered than others. For example fish species are a quite well covered, at least there are point data for 16,100 species out of ca. 17,100 fish species available. However, only 84% of the species with occurrence information are well documented at a country scale. Fore some groups data is sparse, for example on sea snake species and for many marine mammals.

Specific Taxonomic Groups: Data on freshwater species

- A lot of data mobilisation effort is also needed for freshwater species, as the Freshwater Animal Diversity Assessment (http://fada.biodiversity.be/) database contains species names for roughly one third of the estimated 150,000 freshwater species.

Metadata on European Monitoring schemes – the EuMon database

- There is also a taxonomic bias with regards to the groups covered in EuMon, which contains metadata for European montitoring schemes. Relative to their natural abundance and diversity, birds are the best covered species group, followed by butterflies and bats. Least covered species groups are reptile, mammal and fish species. However, it is important to keep in mind that the results in EuMon are on the basis of the number of monitoring schemes, which is not correlated with the amount of records generated or the effort that was put in to surveying.

European research sites – the LTER site network as an example

- For the LTER field sites, there are also differences in coverage of taxonomic groups. Biodiversity data of LTER sites cover mostly plant species. Birds are less well represented in terms of the number of sites that include them as a research topic.

European Taxonomic data

- For the taxonomic databases, some taxonomic groups are overlooked and gaps exist in some families. The overview in Chapter 4 will show some examples for existing gaps such as in plant or animal families.
- Regarding the overall taxonomic completeness the analysis shows that Euro+Med covers 92% of the European flora of vascular plants. Fauna Europaea was calculated to include 99.3% of the known European fauna (actual number of databased species 128,692; estimated number of described species 129,647). The faunistic coverage is less complete, but nevertheless including 90-95% of the total fauna. Both Euro+Med and Fauna Europaea have some families that need to be edited and updated as some larger gaps exist.

Genetic Datasets: Example of fungal species

- For sequence-based data, obvious gap are the lacking names for many species. For example in the INSD database, approximately 35% of species known from DNA as fungi cannot be assigned to any known species for which a full species name in Linnaean classification is available. Many species still lack their representation in INSD, as only approximately 20% of the formally described fungal species are represented with DNA barcode sequence in INSD.

3.4 OTHER GAPS

Besides the gaps mentioned in the subsections above, the specific analysis of the databases showed that there are several other types of limitation and biases in current databases. In the following short overview we outline some of the most obvious gaps.

- Gaps on functional traits: There are only some databases that focus on species traits (e.g. TRY, TraitBank/EOL). However, databases and information on species traits are still scarce.
- Major gaps exist with regards to data on species populations, although there are at least some examples of well-documented population databases (e.g. for marine fishes and for plant species).
- Genetic data: There is still some work needed regarding determining genetic barcode data. For example among marine species, large gaps in DNA Barcode Data exists as data is available only for 10,185 fish species (at least one barcode in the BOLD

system) out of 33,065 currently validly described species (Catalog of Fishes Feb 2014).

- Standards of the data: Some of the datasets have no sufficient documentation of their metadata. A proper metadata description, compliant with existing international standards such as ISO 19115 or INSPIRE is strictly needed. Moreover, metadata schemes (e.g. GBIF, EuMon, DataOne) need to include additional fields to facilitate thematic and/or qualitative selection of datasets. For example, more data on the type of collection (random observation/monitoring) or data quality is needed.
- Closing the gaps: The results of our gap analysis also show how mobilization efforts of recent years have already helped to close existing gaps. For example, in GBIF, the data volume more than tripled between the years 2008 to 2014.
- Dark taxa – unknown species: There are still many unknown species. Even for finfish species the exact number is not known – and the number is increasing each year (200-400 newly described species/year). In other taxonomic groups the number of undescribed species is even higher, for example in bacteria. This leads in turn also to an underestimation of species richness. For example in the case of Fauna Europaea, there is still a high number of newly described species each year for Arthropoda, Molluscs and Nematodes.
- Gaps from published literature to databases: There is also a gap regarding the delay of data - from data being published in papers until data becomes available in databases, e.g. for newly described species. As the example of some online databases show, there is a substantial time delay.
- Main gaps genetic datasets: The main gaps in nucleotide sequence data deposited in INSD are related to sequence quality, missing- and misidentifications and lack of or misuse of metadata standards.

3.5 DATA AVAILABILITY

Generally, sharing of data and thus both accessibility (e.g. by online access) and availability (e.g. if the data has unrestricted access or not) are essential for the further usage of biodiversity information. To secure the free usage of biodiversity data for all kinds of analyses or modelling efforts, free access to European or global datasets is one of the most important prerequisites. So, a fundamental question regarding the evaluated data sources was whether there is unrestricted access (for example, online access under a Creative Commons Zero licence) to the data. In case the access is restricted, the type of restriction was further analyzed, for example if (a) raw data can be downloaded under an open license or waiver or (b) raw data can be downloaded under a restrictive license (such as non-commercial or research only) (c) raw data can be downloaded, but without a license (re-use must be requested) or finally (d) raw data cannot be downloaded but the data can be browsed online. (e) raw data cannot be accessed).

Up to now, only a small percentage of data is shared openly. For example, in a survey, where 1,389 researchers participated, only 25% of the researchers make their data openly available for everyone (Kuipers and van der Hoeven 2009).

For the datasets analysed in this survey, we found huge differences in terms of data availability. As Fig. 8 shows (cf. Table 5), there are still quite many datasets that have, in some respect, restricted access to their data (around 67%). Only one third of the data that was surveyed in this report has completely unrestricted access, There are many reasons why access to biodiversity data is restricted. In several cases, one has to be either part of a

consortium or member of the research network to obtain data. This “tit for tat” approach ensures that in order to obtain data, one has to give access to your own datasets that are related to this area. Such examples are some datasets of the European Vegetation Survey. Other data providers, particularly of specific expert databases with validated data, give access to their data only after a (formal or informal) request, like the data of the Atlas of European Breeding Birds or the Atlas of European bees.

At the same time, there are many data providers that do not have a strict policy regarding the sharing of datasets. This means that some of the datasets have unrestricted access and can be freely downloaded, for others access is only granted after a formal request. For example in the LTER metadatabase, some datasets can be freely accessed via links to webpages where data can be downloaded, whereas for other data sources access is restricted. Some providers show also a different data policy for biodiversity data, for example the aggregated data for the Article 17 reporting of the EU members states is freely available (e.g. as species distribution maps) whereas underlying raw data is not available from most of the member countries.

However, some data providers are strictly committed to open data access and open source software. For example many datasets can be accessed via GBIF , which has mobilized 440 million data records (June 2014), and data can be freely downloaded and used (however, for some of the datasets there is a defined use, i.e. some data providers claim restrictions regarding the use of the data).

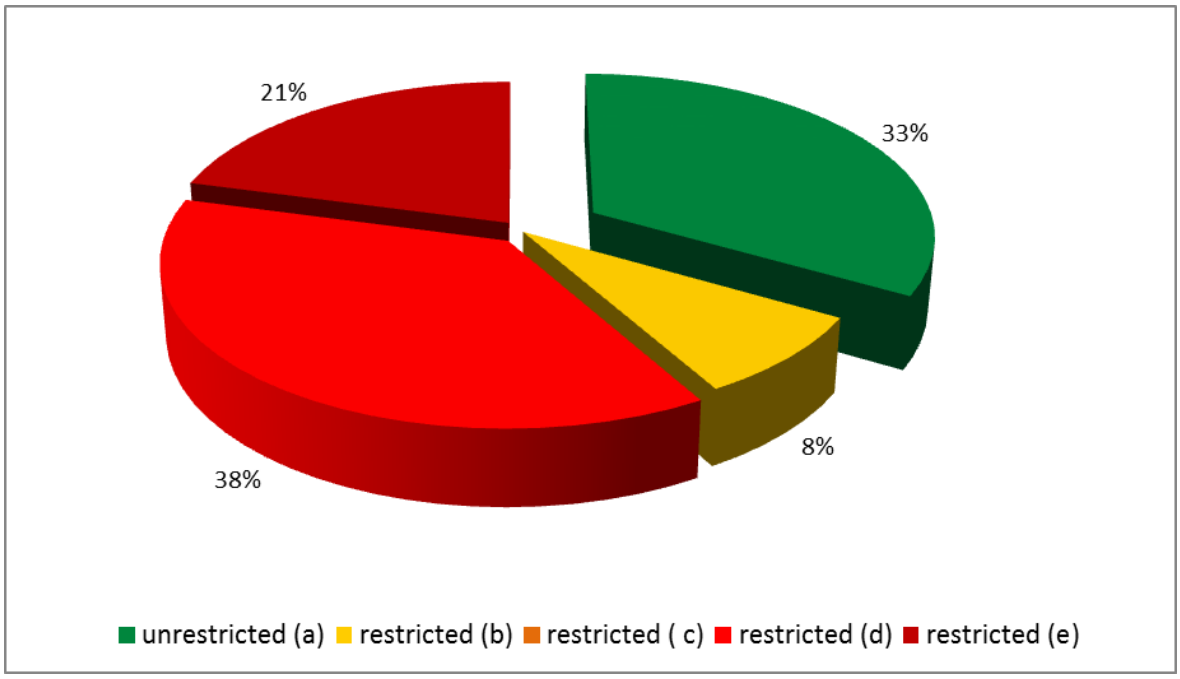


Fig. 8: Accessability of datasets surveyed in the gap analysis (a) raw data can be downloaded under an open license or waiver or (b) raw data can be downloaded under a restrictive license (e.g. non-commercial, research only) (c) raw data can be downloaded, but without a license (re-use must be requested) or (d) raw data cannot be downloaded but the data can be browsed online or (e) raw data cannot be accessed).

Table 5: Overview of datasets analysed and data accessability

Dataset	**Accessability**
Data provided by the Global Biodiversity Information Facility (GBIF)	unrestricted (a) , restricted (b)
Atlas florae europaeae (AFE)	restricted (e)
European Vegetation Archive of European Vegetation Survey (EVS)	restricted (e)
FishBase	unrestricted (a)
Ocean Biogeographic Information System (OBIS)	unrestricted (a)
World Register of Marine Species (WoRMS)	restricted (d)
European Union data on Habitats Directive Article 17	aggregated data: Unrestricted (a), raw data: mostly restricted (e)
PolyTraits	unrestricted (a)
Marine and coastal data holdings of the UNEP World Conservation Monitoring Centre	restricted (b) or (d)
Checklist of Western Palaearctic Bees	restricted (d)
Atlas of the European Bees	restricted (d)
Atlas of European Breeding Birds	restricted (d)
Pan-European Common Bird Monitoring Scheme	restricted (d)
International Nucleotide Sequence Databases (INSD)	unrestricted (a)
The National Center for Biotechnology Information (GenBank)	unrestricted (a)
Fauna Europaea	restricted (d)
Euro+Med PlantBase	restricted (d)
LTER Data	partly unrestricted (a), some data restricted (e)
Trawlbase	restricted (e)
ERMS	restricted (d)

* counted twice for the Figure as data accessability differs withing the datassets of the database

3.6 GENERAL RECOMMENDATIONS FOR CLOSING EXISTING BIODIVERSITY DATA GAPS

For improving the current situation for biodiversity data in Europe, particularly regarding the quality and comprehensiveness of available datasets, we highlight in this report a whole set of recommendations. The following more general recommendations target the European and national policy levels, but also scientific communities, institutions, and individual researchers. Several of the following action points should also be further addressed in the developing work plan of EU BON (Hoffmann et. al. 2014). Furthermore, we give specific advice in the relevant sections in the text (see Chapter 4) for individual data sectors and information realms. These specific recommendations should help to close the gaps of existing databases and to mobilize additional datasets to complete their spatial, temporal and taxonomic coverage.

Here, we offer some general recommendations for improving the data quality and quantity at European scale and how to improve the data availability of biodiversity data in general.

3.6.1 Recommendations to the EU and national authorities

1. Provide financial and other support at European level to relevant regional and national databases to close existing gaps in data coverage and availability.
2. Make all primary data(-sets) under the auspices of the European Union/European Commission and national authorities fully and openly available.
3. Provide incentives for individual researchers and projects to openly share data online, and provide guidance and best practice examples for data management procedures (e.g., recommended embargo periods for releasing research data, standardized citation and acknowledgement practises for online data sources, recommendations for intellectual property right issues for compiled datasets, etc).
4. Strengthen analytical services that use available biodiversity information to provide further incentives and guidance for data providers and information infrastructures.
5. Support relevant international data aggregators and information infrastructures, such as the International Union for Conservation of Nature (IUCN), the Group on Earth Observations-Biodiversity Observation Network (GEO-BON) and the Global Biodiversity Information Facility GBIF) through funding and policy specifically to mobilize datasets only available at local or regional scales.
6. Support digitization of biodiversity data for European protected areas and their sharing, preferably through an international network such as GBIF. Extend the networks of protected areas that undertake monitoring activities and increase monitoring efforts in these sites (e.g. Biosphere reserves).
7. Encourage and support EU member states and other European countries to participate in and become members of GBIF, and to share data via the GBIF portal.
8. Increase funding and support at European level for the collection and provision of key biodiversity and ecosystem data currently not fully available (e.g., genetic information, traits, ecological interactions, etc).
9. Develop and sustain standardized biodiversity monitoring schemes in Europe for generating long-term data sets over larger areas.
10. Undertake and promote regular assessments of available data sources and mechanisms for identifying and closing gaps in data coverage and quality.

3.6.2 Recommendations for European and National Research Networks

1. Support data collection for ecosystem services on a local, regional and European scale. These data are in short supply particular for certain areas such as non-fisheries marine services. Encourage collaboration between EU research networks and disciplines. Particularly strengthen the links to the Earth Observation community, e.g. for an increased use of remote sensing and other spatially-explicit products.
2. Ensure the digitization of historical and legacy data, particularly where those data have been generated by European projects. Continue developing the mechanisms to extract, and provide through it, observation records from the published records either using specialised workflows to extract this data (cf. Agosti and Egloff 2009) or by promotion of advanced publishing in journals. Increase the efforts to make existing data from monitoring schemes available, also from museum collections, citizen science projects and small databases.
3. Implement the Global Name Architecture (GNA) procedures between several components. Main bodies (like GBIF, Catalogue of Life) should be responsible and increase their work and participation in the GNA.
4. Organize and update the information systems on alien species to support much faster data exchange – choose a reference system (such as the DAISIE - Delivering Alien Invasive Species Inventories for Europe or the Global Invasive Species Information Network, GISIN) and align their activities with the work of the European Alien Species Information Network (EASIN).
5. Enhance the collection of socio-economic data for analyses and the establishment of interdisciplinary projects that share their datasets and knowledge or at least the essential datasets.
6. Contribute to and support species assessments (e.g. the Red List) and inventories, as these have been shown to be of particularly high value for conservation policies.
7. Specific recommendations for closing GBIF gaps: (a) support GBIF participant nodes to enable them to publish the most accurate, complete and diverse content possible (b) review the GBIF endorsement model to allow engaging a wider variety of data publishers than currently possible (c) encourage data providers in GBIF to implement specific quality checks of their data, to secure the high quality standards.
8. Speed the detection and description of new taxa by supporting DNA barcoding initiatives and the rapid integrated publication tools.
9. Increase the cooperation between different projects and initiatives with similar databases or taxonomic and geographic overlap.
10. Prioritise data collection goals so that funding is targeted at answering priority questions.

3.6.3 Recommendations for the Scientific Community and individual researchers

1. Actively promote free access to biodiversity data in the larger scientific community. Develop and provide incentives for data guardians, and promote fair and best practice rules for open data sharing.
2. Apply rigidly recognized international data standards for biodiversity information and promote the use of those standards in universities, research institutions and agencies.
3. Coordinate and standardise also the sampling efforts at various scales for sampling different organism groups and environmental variables simultaneously.
4. Adequately cite the origin of data, wherever possible, using the Digital Object Identifier of those data. Centralising data or other measures to provide easier access to all data would benefit users seeking to use data.
5. Encourage the enlargement of research site networks (e.g. LTER and others) with specific focus on underrepresented areas and topics.
6. Increase curation of taxonomic databases (e.g. FaEu, Euro+Med, ERMS and subsequently PESI) to reveal information on species names and work on detecting new species / cryptic species / dark taxa.
7. Support additional gap analysis particular at a fine scale to provide additional support for prioritization and policy support. Develop a strategy for faunistic and floristic checklists at various levels (local/continental/global).

3.6.4 Recommendations to the biodiversity informatics community, to data managers and operators of information infrastructures

1. Improve integration efforts for available datasets (observation records, remote sensing) through the adoption of a limited number of standards and interoperable exchange formats and make metadata for datasets accessible. Adopting such standards could help to overcome the current lack of integration that causes a severe bias in status- and trend analyses.
2. At the same time, extend the current metadata standards with additional fields to give more of the needed additional information to the users (e.g. on presence, absence and abundance data). Support the development of standards, e.g. on Global Unique Identifiers for point data records and specimens.
3. Revive exisiting databases like the BioCASE metadatabase about European collections and find strong incentives for institutions and curators to respond to questionnaires or to generate automatic metadata.
4. Create a powerful data conversion tool that facilitates data exchange between all important formats that are used in other communities.

3.7 LITERATURE

Agosti, D., W. Egloff. 2009. Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2009, 2:53, http://www.biomedcentral.com/1756-0500/2/53/abstract

European Council (1992): European Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora ((OJ L 206, 22.7.1992), p. 7

Kuipers T., Van der Hoeven J. (2009): Insight into digital preservation of research output in Europe. PARSE. Insight Project deliverable D3.4 Survey Report. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

Hoffmann A, Penner J, Vohland K, Cramer W, Doubleday R, Henle K, Kõljalg U, Kühn I, Kunin WE, Negro JJ, Penev L, Rodríguez C, Saarenmaa H, Schmeller DS, Stoev P, Sutherland WJ, Tuama EO, Wetzel FT, Häuser CL (2014): Improved access to integrated biodiversity data for science, practice, and policy - the European Biodiversity Observation Network (EU BON). Nature Conservation 8: 49-65. doi: 10.3897/natureconservation.6.6498

4 SPECIFIC GAP ANALYSIS OF EUROPEAN AND GLOBAL DATABASES

4.1 CRITERIA FOR ASSESSING THE GAPS

In the following chapters, the results of an in-depth and specific gap analysis of several important global and European datasets are presented. The datasets represent some main sources for biodiversity data, either for specific realms (terrestrial, marine), some taxonomic groups or thematic fields (taxonomy, genetic databases) or networks of European field sites (LTER).

Each of the specific gap analysis chapters are structured along a common outline. First, the dataset is shortly described with its main features and the specific focus of biodiversity data. Also included is an overview of the spatial coverage of the dataset, for example, if the data includes global datasets or focuses on European data or specific countries.

The general description is followed by an analysis of gaps and biases of the datasets that were determined by applying specific criteria. These specific criteria were defined for each of the type of gaps (see below) and taking into account the requirements for taxonomic data to be spatially modelled in EU BON (internal EU BON document by Ingolf Kühn et. al.). In most of the cases, two or several databases were compared, to detect the specific gaps, limitations or biases of a dataset. Gaps in biodiversity information can be detected on various scales and are strongly dependent on the needs of data users with respect to the key questions on biodiversity (cf. the high level questions in the first chapter).

In the specific analysis (see the different sections in this chapter), there was a focus on evaluating gaps taking into account the seven target high level questions (see Box 2).

Box 2: Target high level questions on biodiversity

From the highest ranked questions the most relevant were chosen and a set of seven target high level questions were defined at the Work Package 1 meeting in Stockholm in January 2014 (see Annex 1 for the complete list). These seven target high level questions were used for the gap analysis on specific datasets:

1. Can we identify status and trends of [European] species? Can we identify status and trends of biodiversity taking interspecific phylogenetic or intraspecific genetic diversity into account? Can we assess the risk of extinction?
2. Can we assess the status and trends of [European] ecosystems and ecosystem services?
3. Are we closing the biodiversity knowledge gap (poorly known organisms, ecosystem services, areas)?
4. Are we filling the gaps in historical knowledge (in relation to available historical data in collections, literature and non-mobilized digital datasets) so we can evaluate long-term trends?
5. Can we identify trends in the spread and effects of alien and invasive species [in Europe]?
6. Can we identify drivers behind [European] changes in biodiversity over time?
7. Can we assess the effect of [European] marine and terrestrial protected areas on the conservation of biological diversity?

For answering the seven target questions on biodiversity the data has to fulfil certain requirements, e.g. with regards to geographic and temporal coverage. Biodiversity datasets have to fulfil some specific minimum criteria in order to serve as a source for biodiversity assessments, e.g for evaluating trends in species or to answer other target high level questions like for assessing the effects of protected areas.We grouped thus the gap analysis according to four different types of gaps in the data and this structure was also used for the thorough gap analysis of the data sources. These different types of gaps were (a) spatial, (b) temporal and (c) taxonomic gaps of the data and on (d) data availability.

For each type of gaps, the datasources were evaluated taking into account several criteria:

a. For **spatial gaps**, the data was evaluated with regards to the resolution of the data and if there exist gaps in the spatial coverage. Further questions on the spatial aspects of gaps are whether only presence data exists or whether there is also abundance or even data on species absences available.

b. For the **temporal gaps** of data, the data was tested whether a temporal reference is available (date, year, period), for which years or time periods data exists, e.g. if particularly long-term data exists (> 10 years, since 1980) and if there is a homogenous distribution of records over time.

c. For **taxonomic gaps** it was evaluated for which taxonomic groups/species data is available or whether data gaps for some groups exist. Other questions on the data were if the datasets cover terrestrial as well as freshwater and marine species or if biases in the data are apparent.

d. Another important aspect to test the data was to evaluate the data **availability and accessibility**. To secure free usage of biodiversity data for all kind of analyses or modelling efforts, a free access to European or global datasets is one of the most important prerequisites. So, a fundamental question regarding the evaluated data sources was if there is and restricted or unrestricted access (for example a free online access) to the data. In case the access is restricted, the kind of restriction was further analyzed, for example if (a) raw data can be downloaded under an open license or waiver or (b) raw data can be downloaded under a restrictive license (e.g. non-commercial, research only) (c) raw data can be downloaded, but without a license (re-use must be requested) or finally (d) raw data cannot be downloaded but the data can be browsed online or (e) raw data cannot be accessed.

4.2 GENERAL REVIEW OF GAPS IN BIODIVERSITY DATA: MONITORING TRENDS IN GBIF MOBILIZED CONTENT TO HELP ADDRESS GAPS

To help enable assessment in the trends of data biasing and quality of species occurrences data accessible through the network of the Global Biodiversity Information Facility (GBIF, http://www.gbif.org/), the GBIF Secretariat has developed reporting tools to visualise key characteristics of the data published and integrated through the GBIF network and to highlight changes in these characteristics over time. This work is based on comparative analysis of 25 snapshots of the complete global GBIF data index covering a period from late 2007 to the present. Here we introduce the results of this early work and provide some recommendations for how this can be used to help address some of the gaps present in the content mobilized through the GBIF network.

The early results of this work are currently available on http://analytics.gbif-uat.org and will be integrated into the GBIF website during the course of 2014 following feedback from GBIF stakeholders and progressively enhanced.

4.2.1 Introduction

GBIF aims to bring together all available evidence for the recorded occurrence of any species at any time and place, along with associated observations, measurements and links to further information (e.g. deposited specimens, images, sequences, collecting event). As of June 2014, the GBIF network has mobilised more than 440 million data records. These records standardly include the accepted identification for the organism, the associated locality (normally with coordinates) and date, and classification of the record according to type of supporting evidence (e.g. specimen, sequence, multimedia object, human observation). Additional data elements may be included according to the nature of the specimen or observation underlying the record.
The GBIF data index integrates these data elements and accordingly offers a summary of the taxonomic, spatial and temporal distribution of available data. The GBIF data index is served through the GBIF website (http://www.gbif.org). Earlier versions of this index were presented through the GBIF data portal http://data.gbif.org (due to be decomissioned in 2014). Over time, additional data sets are included and changes and corrections are made to existing data. It is possible to use changes in the GBIF data index to analyse how on-going efforts to mobilise and curate biodiversity data result in changes in data availability.

The current study is based on 25 snapshots of the GBIF data index from between December 2007 and May 2014. These have all been processed using the same interpretative rules and organised according to a common taxonomy to allow for direct comparison. GBIF will continue regular analysis using future snapshots to measure changes in coverage, completeness and fitness-for-use for different purposes over time.

By applying consistent quality control and identical taxonomic organisation to both current and historical data, it becomes possible to compare views over time, to determine if gaps are being closed, and if more data are available and fit for use. Since GBIF is an international

initiative with many activities funded at the national level, separate reports have also been generated for each country to review data shared by institutions within the country and to report on levels of data relating to biodiversity within the country. In many cases, national GBIF Nodes are well placed to take action on the data quality issues detected, and also to target specific data holdings to address current biases. This work is in an early stage, and will be enhanced in collaboration with national participants and with users wishing to assess content available through GBIF. In some cases (e.g. Sweden), the prepared data may be used for more detailed analysis at a national level. The EU BON funds allocated to GBIF through task 1.3 have thus been used to seed the development of consistent reporting that will be enhanced and run at regular intervals (monthly or quarterly; yet to be decided) to help with future gap analysis, assessment and monitoring of how the gaps are closing.

4.2.2 Data and code accessibility

GBIF is committed to open data access and open source software.

- The most recent view of the GBIF index is open to the internet and available through the GBIF.org website (http://www.gbif.org/occurrence/search) and API (http://gbif.org/developer).
- The compiled snapshots for this study are available on request. The number of records represented add up to a total 7.4 billion records, and compressed in RCFile format are 0.7 terabytes (Approx 7 TB uncompressed). Other formats can be generated as required.
- The digested views for each page report on http://analytics.gbif-uat.org are available online. For any given report, using a "/csv" directory suffix on the URL will list the files. e.g.
    - The report: http://analytics.gbif-uat.org/country/SE/publishedBy/index.html
    - The CSV files: http://analytics.gbif-uat.org/country/SE/publishedBy/csv/
- The scripts used in this iteration of work are available at http://github.com/gbif/analytics.
- All data interpretation code is available in http://github.com/gbif/occurrence

4.2.3 Methodology

The methods to reproduce this work are documented within the github project on https://github.com/gbif/analytics and will be continually enhanced as the work progresses. The high level process is summarized as:

- Verbatim content from 24 historical views of the GBIF index was restored from archives and processed to the latest quality control (https://github.com/gbif/occurrence) and taxonomic backbone using the GBIF high performance processing environment (Hadoop)
- A single snapshot of the latest index was taken and processed using the same routines as the historical data.
- A series of SQL scripts (https://github.com/gbif/analytics/tree/master/hive/process) was run on the Hadoop cluster (Hive) to process the large number of records (7.4

billion) into smaller summary views suitable for export and subsequent processing using R on a laptop.
- A series of R scripts (https://github.com/gbif/analytics/tree/master/R) were run to generate charts and a static site generator was used to produce the site now deployed at http://analytics.gbif-uat.org.

4.2.4 Coverage of the dataset

The content represented in the GBIF index is global in geographic and taxonomic scope and covers content dating back to before 1700. Content includes records documenting *in situ* species observations as well as specimen collection events, the latter primarily coming from specimen labels, with some field notes and literature sources.

Data are added to the GBIF network by GBIF Participants endorsing publishing organizations. The content in this study represents trends in data mobilized through GBIF since 2007, during which time participation in GBIF has changed as follows:

December 2007:
- 80 Participants (29 voting countries, 17 associate countries, 34 associate organizations)
- 230 publishing organizations sharing 1,655 datasets (123 million records)

June 2014:
- 89 Participants (37 voting countries, 15 associate countries, 37 associate organizations)
- 611 publishing organizations sharing 15,171 datasets (441 million records)

4.2.5 Outline of gaps and biases

Global geographic coverage

GBIF is an intergovernmental initiative, a fact that influences the nature of content available through GBIF. This is prominently visible in the view of all georeferenced data available through GBIF (June 2014, Fig. 9a), which illustrates that data are most abundant in those countries participating in GBIF (http://www.gbif.org/participation/summary, Fig. 9 b). At least part of the reason for this lies in the current endorsement model of GBIF (under revision, more details discussed below), which requires an institution to be formally endorsed by a GBIF Participant before published data are included in the GBIF index.

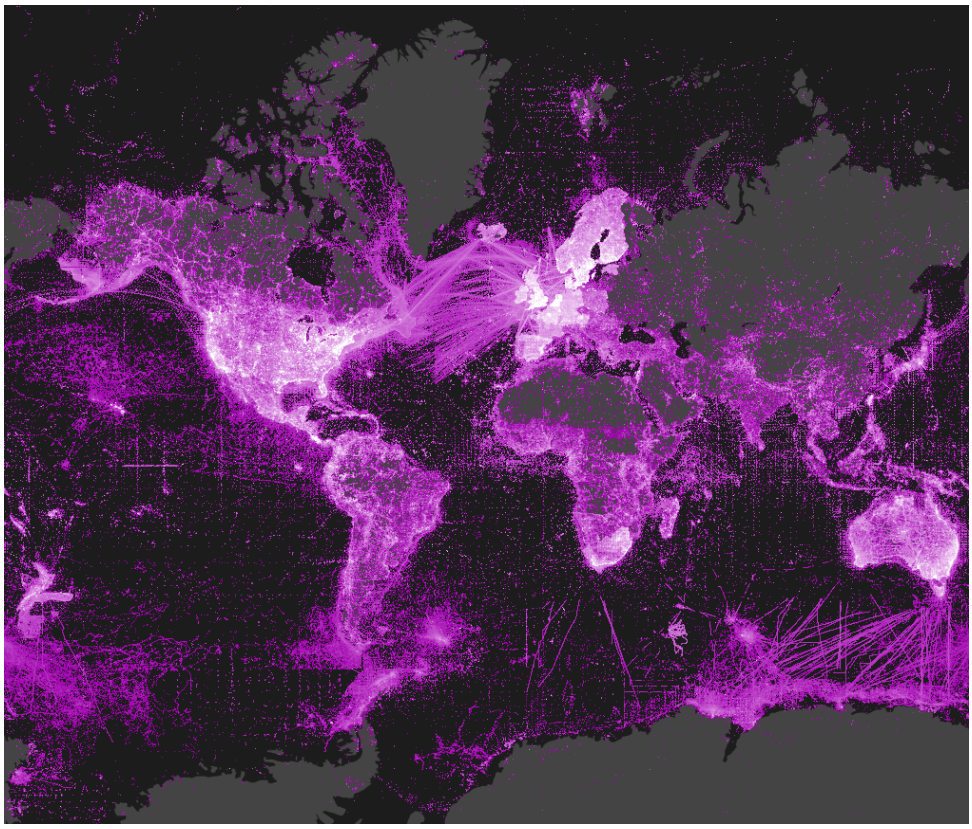


Fig. 9a: Georeferenced records available through GBIF, June 2014. Heat map representation with lighter colours indicating more available data (interactive version at http://www.gbif.org/occurrence).

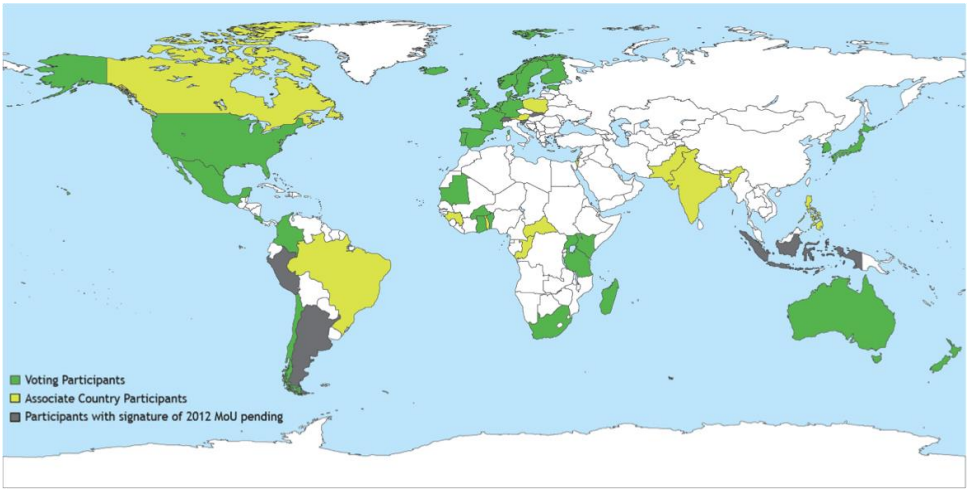
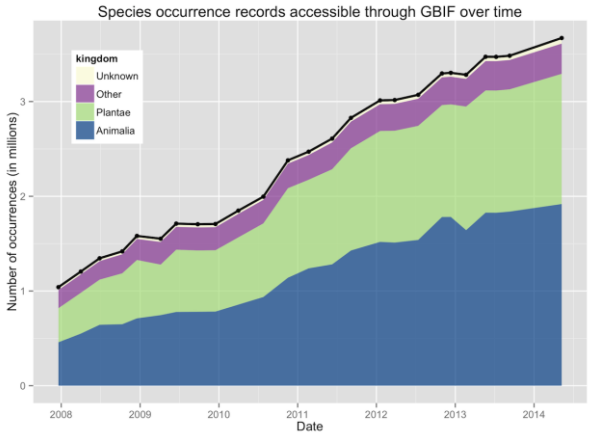


Fig. 9b: GBIF Country Participants, June 2014

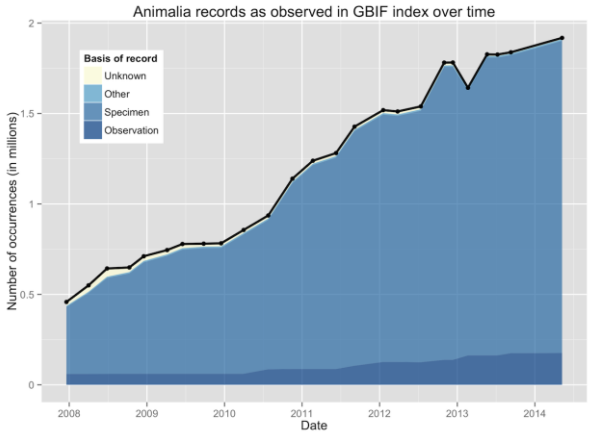
Comparison between countries

In analysing the trends in accumulation of occurrence data (next section, below), a global view of data is applied. It has to be noted here that this may obscure differences at country or regional levels. To demonstrate this, the following figures, Fig. 10 and Fig. 11, illustrate species occurrence counts for Japan and Sweden:

a)



b)



c)

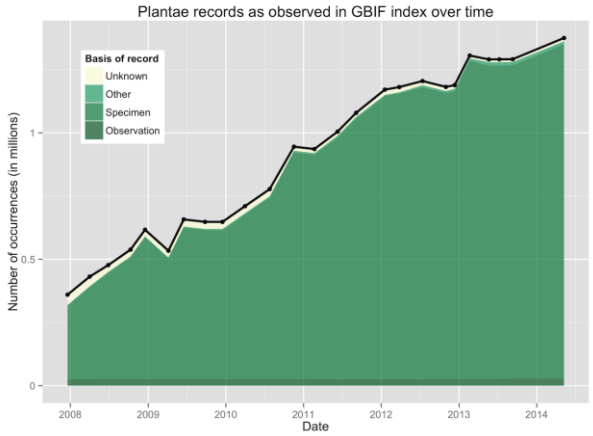
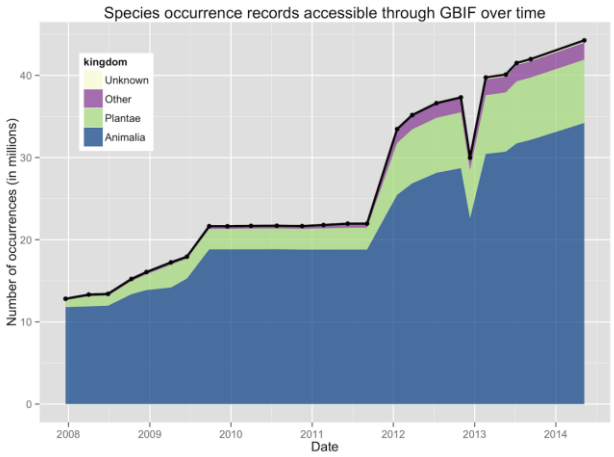
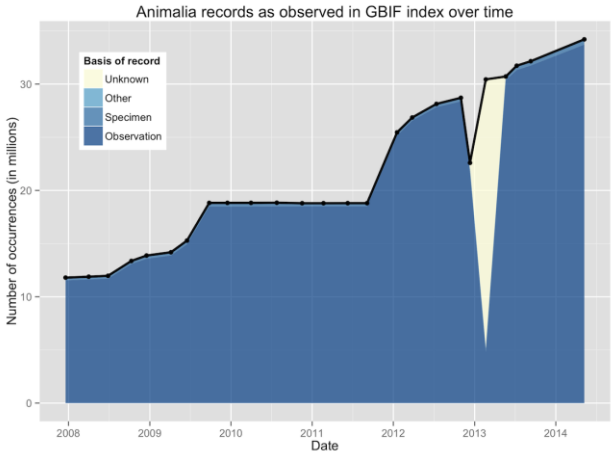


Fig. 10: Japan, species occurrence records available over time. a) total, b) animal records only, c) plant records only. The colours in b and c indicate the basis of record (observation, specimen, other or unknown)

a)



b)



c)

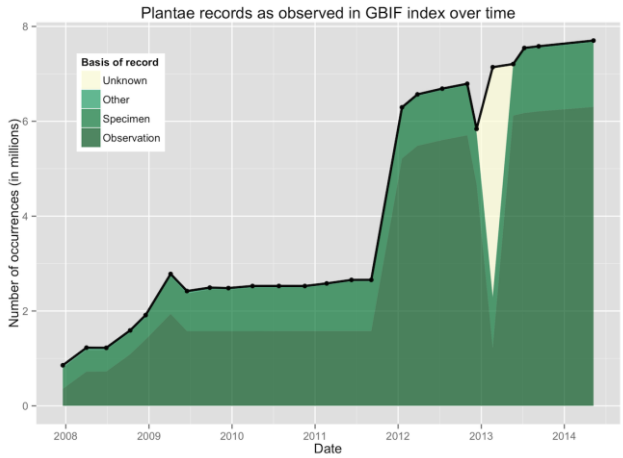


Fig. 11: Sweden, species occurrence records available over time. a) total, b) animal records only, c) plant records only. The colours in b and c indicate the basis of record (observation, specimen, other or unknown). The drop and subsequent gain in numbers and the high percentage of records with an unknown base of record in early 2013 is an artifact, most likely caused by the reconfiguration of a major dataset.

In both cases there is a general continued increase in record numbers, although the absolute numbers differ by a factor of about 10. The basis of record (specimen, observation, or one of several less frequently reported types such as fossils and living materials) is known for the

vast majority of records from both countries. There are however significant differences between the summaries for these countries.
In Swedish datasets, about 80% of all records concern animals, while the Japanese data show many more plant records and around 20% of records are for other groups, including fungi and microorganisms. These other groups are only visible in Swedish datasets from 2012 onwards. Aside from the relative proportions of plant and animal records, the main difference lies in the basis of record: while Japanese datasets contain predominantly specimen records, i.e. from the digitization of physical collection objects, most records in Swedish datasets are based on field observations. Differences like these may document different strategic approaches to data digitization, differing engagement with communities such as citizen scientists, or varying approaches to collection management and open access to data.

Neither country represents a better or worse model than the other, but differences like these need to be taken into account in any analysis of data coverage. At a global level, these differences may be hidden, so any action plans for addressing gaps and targeting specific needs must be based on more fine-grained review, particularly at the national scale.

4.2.6 Trends in accumulation of occurrence data / integration of historical data

Global number of occurrence records

Fig. 12 shows the number of available occurrence records categorized by kingdom for each of the 25 index archives, normalized against the most recent GBIF backbone taxonomy. The "Unknown" category includes records with taxonomic information that cannot be linked to available taxonomic checklists.

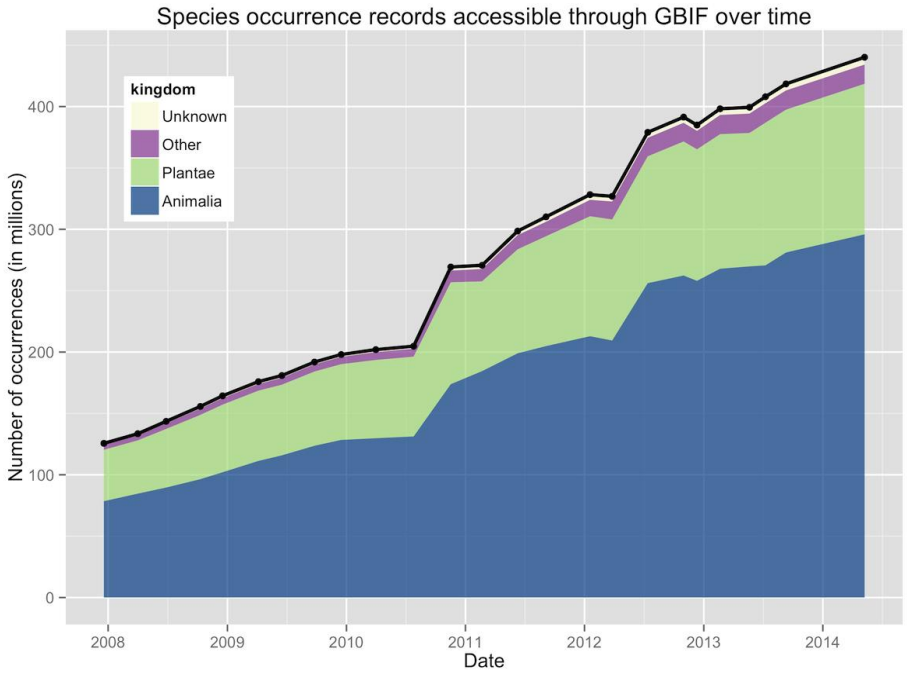


Fig. 12: Data volume for major taxonomic groups. Number of occurrence records per kingdom over Index snapshot version, 2008-2014. The category of "Other" summarizes fungi as well as several groups of microorganisms, bacteria and viruses, but typically predominantly consists of fungi.

Fig. 13 gives the number of species with available occurrence records, categorized by kingdom. In this case, species counts are based on the number of binomial scientific names for which GBIF has received data records, organized as far as possible using information on known synonyms, as recorded in key databases such as the Catalogue of Life. Since many names are not yet included in these databases, some proportion of names will not be recognised as synonyms and incorrectly treated as separate species. Therefore these counts can be used as an indication of richness only, and do not represent true species counts. All data have been processed using the same, most recent, version of the common GBIF backbone taxonomy, and comparisons over time are therefore realistic.

For all major taxonomic groups, occurrence record numbers in the index have been increasing more or less steadily since 2007. Animal records constitute the largest component throughout, with up to two thirds of the total data volume, while plant records show more rapid growth. While data mobilization is clearly dominated by these two major groups, numbers for other taxonomic groups have also been steadily increasing. The most significant group included under "Other" in Fig. 13 is Fungi.

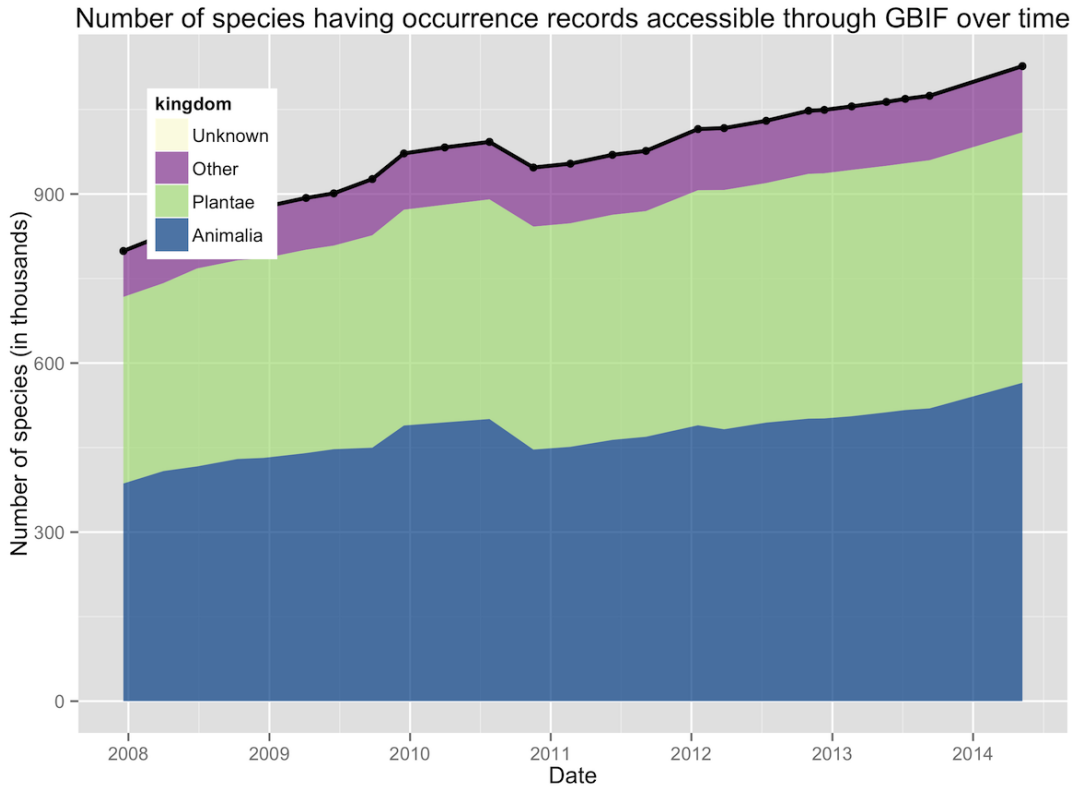


Fig. 13: Taxonomic richness. Number of species names documented with occurrence records over Index snapshot version, 2008-2014. Note the caveat (text above) concerning the interpretation of a binomial as a species.

Fig. 13 shows a complementary view to Fig. 12: the number of species-level taxon names for which occurrence records are documented in the index for each snapshot version. Here again, numbers for all major taxonomic groups have risen more or less steadily over time. However, while the proportion of occurrence records for groups other than plants and animals (Fig. 12) is relatively low, the number of species-level names for voucher or

observation data are available represents a larger proportion of the total. In addition, the number of plant species with associated occurrence records is almost as great as for animal species, although there are far fewer plant species described (about 321,000 vs 1.36 million, according to the IUCN Red List version 2010[6]).

As with the country bias discussed in the previous section, taxonomic bias becomes more apparent when reviewing taxonomic groups below the kingdom rank. Occurrence coverage, for example, is known to have a strong bias towards bird observation data, caused by the high data accumulation rate for these data from the large citizen science community and the rapid web publication of such data in recent years. Analysis of species richness is impacted by uneven taxonomic resources. Significant gaps exist for example in catalogues of molluscs, beetles, algae, and some groups of higher plants, as well as for fossil species. For any targeted study, therefore, these gaps would need to be analysed in more detail.

Geographic completeness of available records

The chart in Fig. 14 illustrates changes in the number of available records which include coordinates, focusing on those for which no known issues have been detected in the automated checking routines. For records without accepted valid coordinates, this chart also shows the number of records for which the country of occurrence is known.

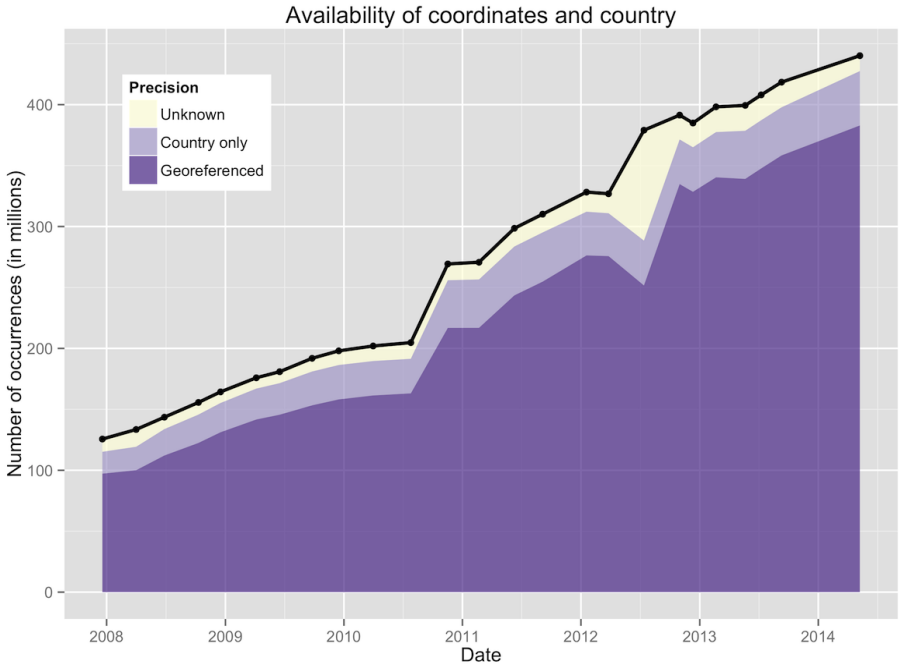


Fig. 14: Availability of georeferencing information. Number of occurrence records per available information type over Index snapshot version, 2008-2014. Note that the availability of a country name is only counted if coordinate data are not available for the same record.

The graph shows that, over time, the percentage of records with coordinates has gone up from about 80 to about 90%. Together with the textual information of interpretable country names, about 97% of all records include some indication of geo-location, though with varying precision and accuracy. The remaining 3% of records lack any georeferencing information either because of technical errors (invalid numbers, partially missing values), or

[6] http://www.iucnredlist.org/documents/summarystatistics/2010_1RL_Stats_Table_1.pdf

because of content errors (e.g. coordinate values not matching a given country) or as a result of intentional or unintentional omissions. While the percentage appears comparatively high, more detailed analysis shows that more attention is needed to ensure the accuracy of values and to review or improve accompanying metadata, e.g. to clarify where coordinates indicate centroids or corner points of grid cells in a monitoring scheme rather than actual geo-location of the locality of occurrence.

Geographic coverage for recorded species

Fig. 15 illustrates changes in the number of species for which records are available from a range of localities. The earth's surface is divided into a series of one-degree grid cells (for finer grids, see http://analytics.gbif-uat.org/global/index.html). All species are then categorized according to the number of such cells for which GBIF has any available data for the species from 1970 onwards. The chart shows the proportion of species recorded in each snapshot from only one such grid cell, or from between two and twenty such grid cells, etc. Various reasons may limit the number of cells from which any species has been recorded (rarity, obscure taxonomy, few observers, detectability, etc.). However, greater numbers of records typically indicate a likelihood that the data will be more suitable for various modelling activities.

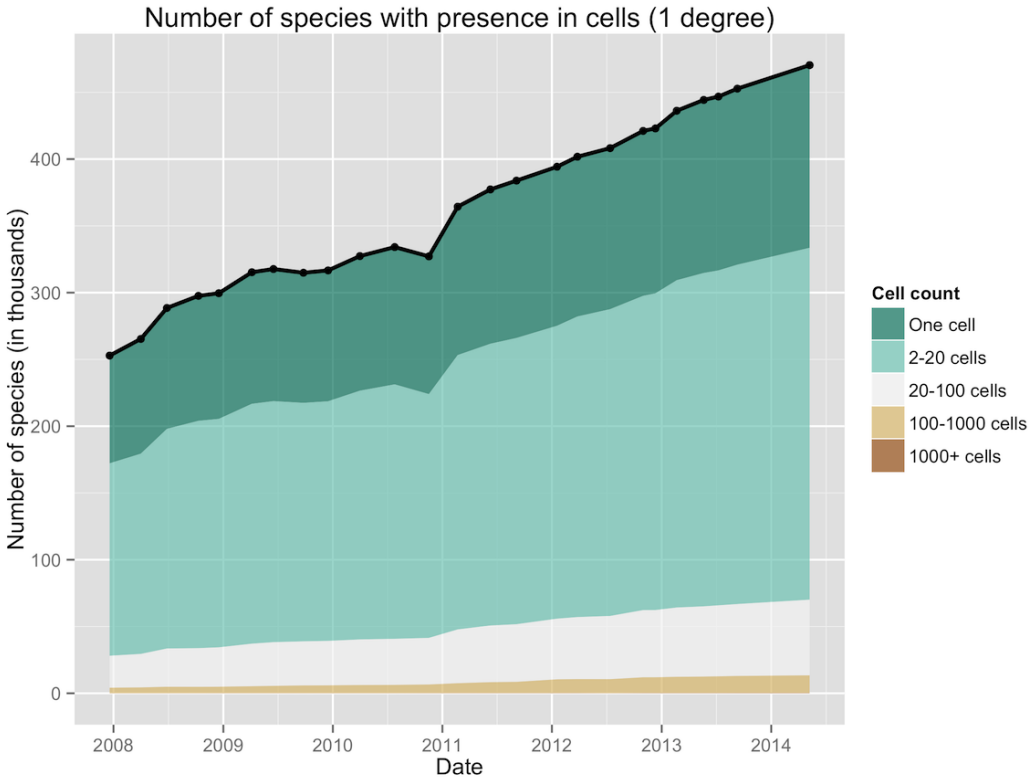


Fig. 15: Geographic spread of species occurrences. Number of species per spread classification group over Index snapshot version, 2008-2014.

In the 2012 index version, for example, 50,000 of ca. 400,000 species were documented to occur in at least 20 different 1-degree grid cells. In 2014, the same applied to about 70,000 of 470,000 species.

Taxonomic precision

Fig. 16 illustrates changes in the number of available records which include an identification at least to the rank of a species. The numbers of records identified to an infraspecific rank, to a genus or to a higher taxonomic rank are also shown.

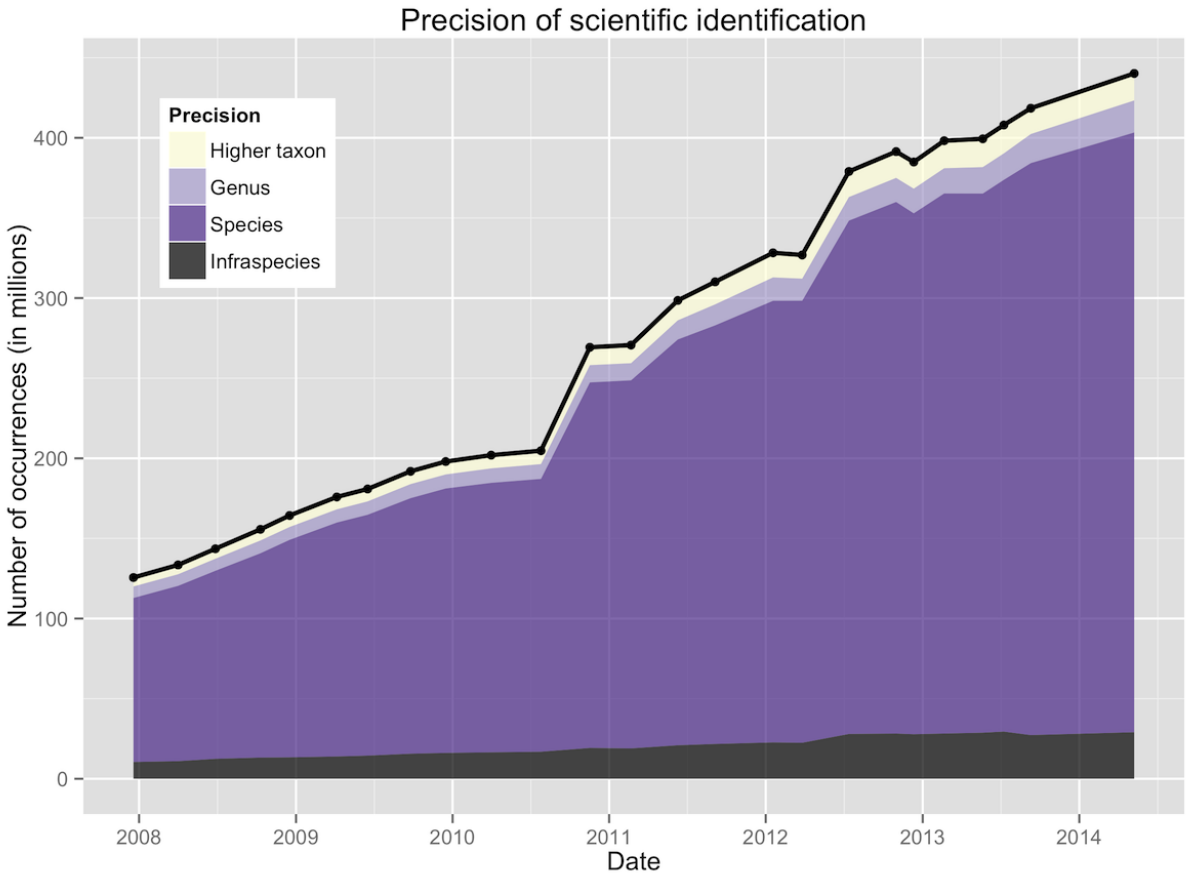


Fig. 16: Precision of scientific identification. Number of occurrence records per precision of scientific identification over Index snapshot version, 2008-2014.

Throughout all index versions from 2008 onwards, the vast majority of occurrences were identified to species level or below. Only about 10% of records were identified only to a higher taxonomic rank. These records arise e.g. when specimens collected and archived during projects and provisionally filed for later examination, when taxonomy is obscure or diagnosis is difficult, or when monitoring projects involve observers with different degrees of taxonomic expertise. This category also includes a small proportion of records supplied with binomial names but where these names could not be matched against the backbone taxonomy and the records were therefore processed at higher ranks during indexing.

Temporal precision

Fig. 17 illustrates changes in the number of available records which include a complete date, consisting of year, month and day. The numbers of records including only the month and year or only the year are also shown.

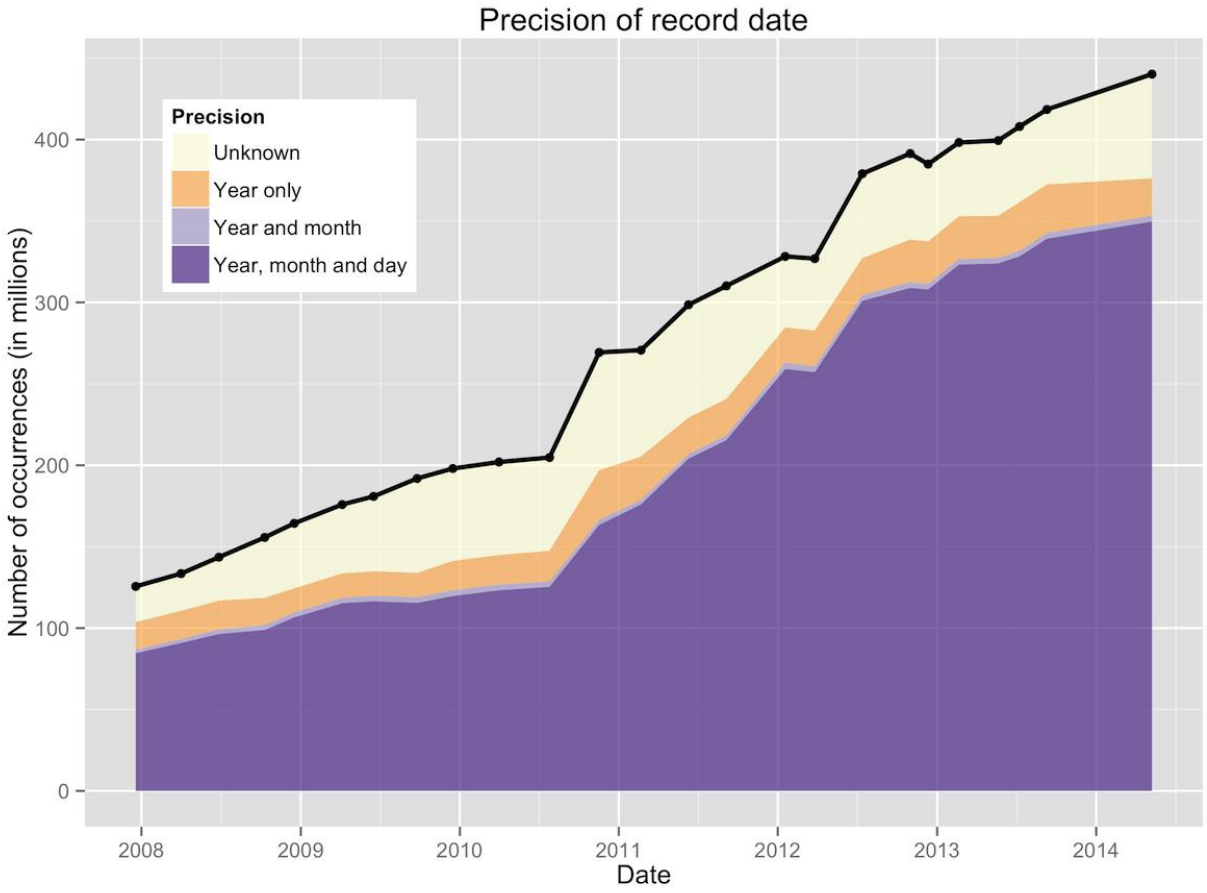


Fig. 17: Precision of occurrence dates. Number of occurrence records per completeness status over Index snapshot version, 2008-2014.

There has only been relatively little change in relative proportions of records in these categories between index versions from 2008 to 2014. While the relative proportion of records with a missing or incomplete occurrence date has decreased from about 30% to about 20%, the absolute number has increased during that period. A certain percentage of these records is due to dates not being supplied in valid or interpretable date formats. In the remainder of cases, no dates or only partial dates have been supplied. With the exception of historical data, where e.g. a specimen label did not in fact record the collection date, most of these gaps could potentially be filled through intensified collaboration with data publishers, to review and improved data capture processes, date conversions, issues around mapping of values in publishing software used, and generally highlighting the importance of this item for a range of data uses.

4.2.7 Recommendations for closing the gaps

The key recommendations to address the gaps and biases present in GBIF mobilized content are to:

- Support GBIF participant nodes to enable them to publish the most accurate, complete and diverse content possible
- Review the GBIF endorsement model to allow engaging a wider variety of data publishers than currently possible

These are outlined in more detail in the following sections.

Supporting the work of GBIF Participant Nodes

Participant Nodes play a central role in facilitating data mobilization in the GBIF network. When joining GBIF, each Participant country or organization commits to establishing a Participant Node as "a mechanism by which a Participant coordinates and supports its GBIF networked data-sharing activities" (2012 GBIF MOU). The GBIF Nodes are a very diverse network of organizations that share a common objective "of promoting, coordinating and facilitating the mobilisation and use of biodiversity data among all the relevant stakeholders within the Participant's domain, primarily to help address the Participant's biodiversity information needs and priorities" (2012 GBIF MOU). This common objective means that trends showing efforts to mobilize and improve the fitness-for-use of data at the country level are an important tool for national Nodes to help set objectives, track general-level progress and report to their stakeholders, for example.

The Participant Nodes have requested this type of analysis on previous occasions. During the development of the GBIF website[7], a consultation was run with Nodes to scope requirements for country pages. The country pages serve several purposes including promoting the role of country Participants in GBIF, redirecting users towards the Nodes and providing an overview of the data available about and published by each country. Among the requested features were trends on the national data publishing efforts and on the data availability about a country[8]. Many Nodes include some data mobilization trends in their annual reports[9] and websites[10], and some have even carried out in depth analyses at the national level[11].

There are several ways in which Nodes can use this type of analysis for self-assessment in support of data mobilization activities. Firstly, it will help to highlight issues with the data that have already been mobilized. In their coordination role, many Nodes will be well placed to use these charts to immediately identify actions for follow up with individual data publishers to increase the fitness-for-use of the data. Secondly, the charts can help to reveal biases that could be used to set mobilization targets, for example through targeted data mobilization projects. Over time, this information can also help to monitor national-level

[7] www.gbif.org
[8] http://community.gbif.org/pg/groups/27781/designing-country-pages-for-the-new-gbif-portal/
[9] E.g. Ireland http://www.biodiversityireland.ie/downloads/annual-reports/
[10] E.g. Atlas of Living Australia http://dashboard.ala.org.au
[11] Otegui J, Ariño AH, Encinas MA, Pando F (2013). Assessing the Primary Data Hosted by the Spanish Node of the Global Biodiversity Information Facility (GBIF). PLoS ONE 8(1): e55144. doi:10.1371/journal.pone.0055144

progress and results of specific investments. This can feed into relevant reporting processes that Nodes engage with, for example reporting under the Convention on Biological Diversity.

This work has been shared with Nodes at an early stage to ask for feedback and to invite those interested to contribute to further development. All feedback is being captured in an online issue tracking system. Following an early consultation phase lasting until 30 June 2014, the feedback received will be sorted into issues to be addressed before this work is integrated into the GBIF website by September 2014 and further enhancements are identified for the 2015 GBIF work programme.

Reviewing the GBIF endorsement model in support of wider collaboration

This year, GBIF is undertaking a review to gather input on proposed changes to the processes for adding data publishers to the GBIF network and for assessing the fitness for use of datasets. This fulfils a commitment in the 2014 Work Programme to expand the model for endorsement of datasets, and to engage with expert communities in order to provide a richer assessment of the value of each dataset to potential users. The proposed changes seek to "*retain the strengths of the current model and improve scientific oversight but accelerate integration of relevant data where possible*[12]". A full consultation document is available[13] that outlines the strengths and weaknesses of the current situation and invites feedback from GBIF stakeholders on suggested changes to the model.

Currently new data publishers are added to the GBIF network only after a GBIF Participant provides an endorsement (by email) for the publisher. New publishers are expected to receive such endorsement from the relevant national Participant or alternatively from a relevant thematic organisational Participant. Once endorsed, a data publisher can publish new datasets without further approval.

This model has been in place since GBIF was first established. Where it works well, it offers the following strengths:

- The endorsement process could detect and reject inappropriate data publishers and data sets without these ever becoming visible through the network.
- The endorsing Participant performs a review of the data being published and may work with the publisher to resolve data quality issues before the data are integrated in the network.
- National Participants can be closely involved in the relationship between their data publishing institutions and GBIF. This may be important for at least two reasons
    - The relationship may help to create a strong national GBIF community and engage relevant stakeholders around GBIF activity.
    - The role as endorsing body may assist the national Node in demonstrating its importance and relevance to relevant national agencies.

[12] GBIF Work Programme 2014-2016
[13] Consultation document in English

There are however also several weaknesses to this model:

- Potential data publishers which are located in a non-Participant country (or which do not have access to a relevant Node) are not readily able to publish data to GBIF. This has three consequences:
    - GBIF does not benefit from potentially valuable datasets, particularly from regions with poor representation.
    - GBIF has reduced opportunities to begin building a profile within the biodiversity research community in the country.
    - GBIF is unable to maximise its role as an infrastructure and network supporting free and open access to biodiversity data.

- Some Participants lack resources to carry out any review of new datasets.
- The current process does not require endorsement for any data sets subsequent to the first dataset offered by a publisher (which limits the direct relevance of the process in improving data quality). The alternative approach (requiring separate endorsement for every dataset) might be unmanageable in cases where publishers produce very large numbers of small data sets.

In brief, the proposed changes are:

- Any potential data publisher should be able to register with GBIF and submit datasets without the need for prior endorsement - such datasets would initially be 'unevaluated'.
- GBIF Secretariat staff would monitor new datasets to detect attempts to abuse the infrastructure.
- As datasets are indexed, they would undergo automated checks for standards compliance and data consistency, and the results would be included on dataset pages in the GBIF website.
- Relevant GBIF Participants, based on location and indications by the publisher, would be invited to review the dataset and provide an endorsement, if appropriate.
- GBIF country pages would include a list of all datasets endorsed by a national Participant, as at present.
- Additional pages would be developed to list all datasets endorsed by organisational Participants and affiliates, encouraging curation of data within areas of expertise.
- Datasets may receive multiple endorsements from any number of Participants, and these would all be displayed on a dataset page, helping to evaluate fitness for use.
- Further elements to assist such evaluation will be explored, including user comments and annotations on individual records, and use of datasets or records in research activities.
- Filters would be developed to help users select or exclude datasets carrying different levels of endorsement, including unevaluated data.

The consultation period runs until 16 June 2014, after which a final recommendation will be prepared and the feedback will be summarised in a report.

4.3 FOCUSED-REVIEW OF GAPS IN SPECIFIC DATABASES: ANALYSIS OF DISTRIBUTION DATA OF VASCULAR PLANTS IN EUROPE

4.3.1 Introduction

There are three main datasets that comprise distribution data on vascular plants throughout the European continent, namely datasets provided by the *Global Biodiversity Information Facility* (GBIF, www.gbif.org), data from *Atlas florae europaeae* (AFE, http://www.luomus.fi/en/atlas-florae-europaeae-afe-distribution-vascular-plants-europe) and the *European Vegetation Archive of European Vegetation Survey* (EVA of EVS, http://www.euroveg.org/).

4.3.2 Data accessibility

Only the GBIF database is available in unrestricted electronic form and data can be freely downloaded; electronic access to the other two datasets requires permission. Data from AFE are available in a printed form in most of the large botanical libraries in Europe (see references below); electronic access is based on negotiations with the AFE management and might involve some financial contributions. EVS is a kind of federation of national vegetation databases. It comprises vegetation relevés (data on vegetation plots, where presence of all taxa is recorded, together with estimation of their abundance and dominance), from which distribution data can be extracted. In some cases distribution data from individual national datasets are publicly available, as e.g., for the Czech Republic (see http://www.florabase.cz/, http://florabase.cz/databanka/index.php?lang=en), in other cases national databases are currently opened only for cooperating parties. The European Vegetation Archive as a whole is also currently opened only for cooperating parties (subject to formal project proposal, for conditions of use see http://euroveg.org/download/eva-rules.pdf).

4.3.3 Data Sources

European Vegetation Archive of EVS

The European Vegetation Archive (http://euroveg.org/eva-database) was recently established and is continuously updated. It will contain vegetation relevés (and thus also vascular plant distribution data) from most of the European countries. Its aim is "to create a centralized database of European vegetation plots by storing copies of national and regional databases on a single software platform using a unified taxonomic reference database. EVA does not affect the ongoing independent developments of source databases and it guarantees that data property rights of the original contributors are respected" (Chytrý et al. 2014).

A list of the national datasets comprising data on vegetation relevés is available at http://www.givd.info/list_databases.html?&no_cache=1. GBIF and EVS datasets are to the considerable extent complementary, providing together reasonable information on the distribution data on vascular plants particularly for W and Central Europe. The amount of data provided by EVA continuously grows and it is likely that in foreseeable future it will provide much more detailed picture on Central and Eastern European distribution of vascular plants. We have received permission to use currently available georeferenced EVA data for Caryophyllaceae and Cruciferae families for the purpose of this gap analysis.

Atlas florae europaeae

The most complete dataset for distribution data on European vascular plants for particular plant families is that of AFE, which was built in cooperation of all European countries. The already published 16 volumes of the serial include distribution maps of species and

subspecies cover almost 20-25% of the European vascular plants (altogether 4878 taxa). Data (based on herbaria and literature) and the distributions are presented as grid maps (50 x 50 km squares; http://www.luomus.fi/en/grids-mapping-atlas-florae-europaeae) covering the whole continent. Albeit incomplete in terms of plant families and rough in scale, the AFE data set is an excellent reference, based on which completeness of other datasets (including those publicly available) can be tested. Two large families (Cruciferae, Caryophyllaceae, i.e. content of four volumes of printed maps) are available for the gap analysis as part of EU BON project.

4.3.4 Comparison of AFE, GBIF and EVA data

In this chapter, we present here a comparison of GBIF, AFE and EVA data for both above-mentioned families and for a few selected example species.

For the gap analysis of the GBIF database we downloaded a complete set of the distribution data on all species of the families Cruciferae (Brassicaceae) and Caryophyllaceae.

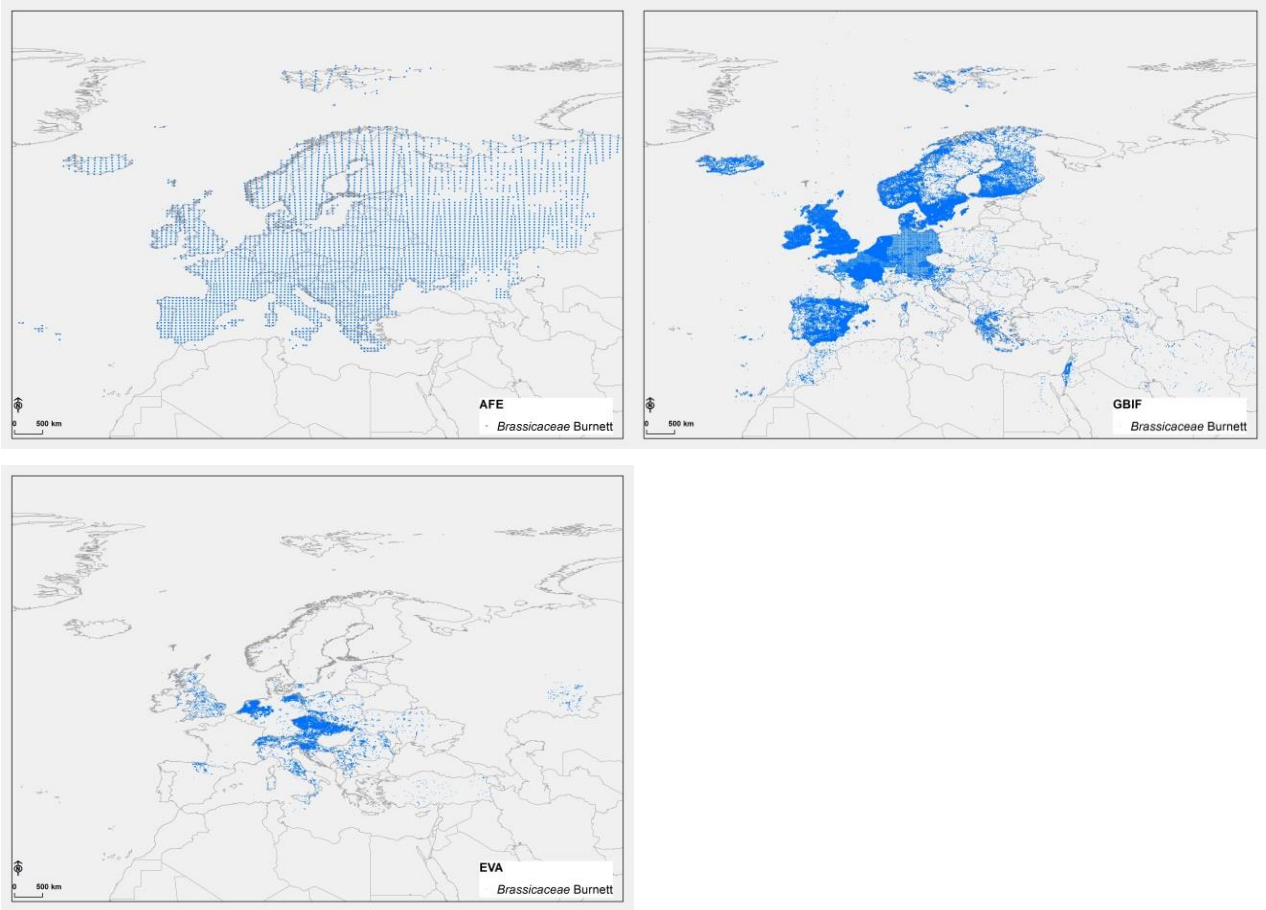


Fig. 18: Synthetic presentation of all Cruciferae (Brassicaceae) distribution data in AFE (occurences are represented by blue dots, map on the left), GBIF (map on the right), and EVA (map in the second row).

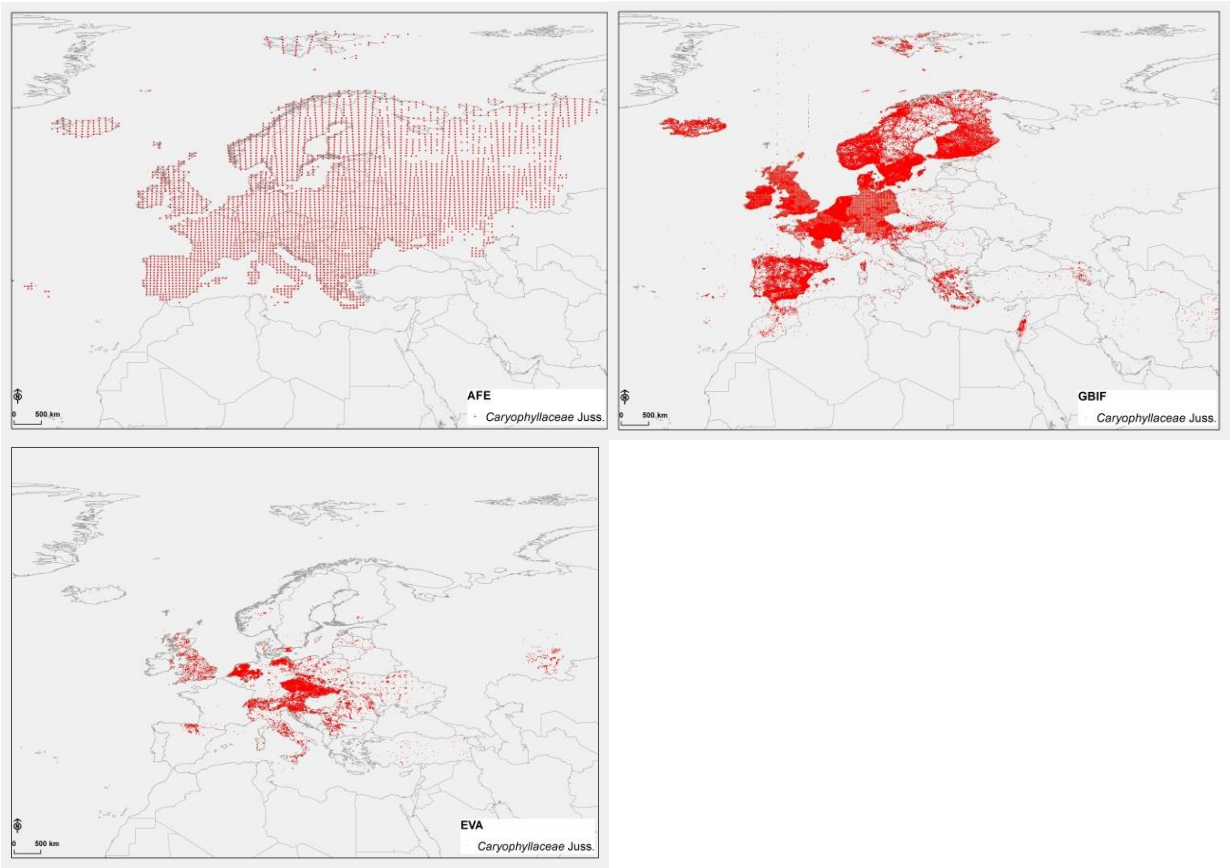


Fig. 19: Synthetic presentation of all Caryophyllaceae distribution data in AFE (occurences are represented by red dots, map on the left), GBIF (map on the right), and EVA (map in the second row).

Overall, AFE synthetic family distribution maps show the presence of the taxa of these two families over the whole European continent. Certain gaps in the area of Russian Federation reflect missing data rather than the real absence of any taxa of these plant families (see Fig. 18 and 19, upper maps on the left). GBIF data, on the other hand, are strongly biased. While the areas of Scandinavia, Island, United Kingdom, Ireland, Denmark, Benelux, Iberian Peninsula, parts of France (but not the whole country), Austria, and Greece are well covered, many Central European, and most of East and South-East European countries are covered only poorly or not at all (see also Fig. 18 and 19, upper maps on the right, Table 6a). This coverage reflects membership of particular countries in GBIF and presence of herbarium specimens from some non-GBIF countries (e.g. Greece) in herbaria of GBIF member states. While it is clear that GBIF vascular plant dataset might be useful for modelling distribution of taxa with the centre of distribution in Scandinavia and Atlantic Europe, it is of little use for those taxa that have their centre of distribution in Central, Eastern and South-Eastern Europe.

The EVA database, on the other hand, provides considerable amount of data for Austria, Germany, the Netherlands, Czech Republic and Slovakia, but some data also for other Central European countries not so well covered by GBIF (see Fig. 18 and 19, lower maps on the left and Table 6b). It is somewhat complementary to the GBIF data and for the taxa with the centre of distribution in Central and Western Europe GBIF and EVA provide reasonable amount of data. Nevertheless, for the taxa distributed in the Eastern part of Europe, neither GBIF nor EVA serves well.

Table 6a. Overview of the amount of distribution records for Cruciferae and Caryophyllaceae families in GBIF database by countries of Europe. (Crucif_all – total numer of records of the family Cruciferae, Crucif_map – georeferenced data used for the presented map; Caryophyl_all – total numer of records of the family Caryophyllaceae, Caryophyl _map – georeferenced data used for the presented map).

Country	Crucif_all	Crucif_map	Caryophyl_all	Caryophyl_map
Albania	94	**16**	90	**4**
Andorra	1700	**1408**	1833	**1626**
Austria	9366	**7945**	7592	**6356**
Belgium	160998	**160620**	184730	**184528**
Bosnia and Herzegovina	56	**4**	64	**9**
Bulgaria	597	**157**	706	**231**
Byelarus	7	**1**	2	**0**
Croatia	286	**94**	364	**96**
Cyprus	323	**20**	211	**13**
Czech Republic	524	**34**	319	**38**
Denmark	26993	**24481**	82215	**81552**
Estonia	56	**28**	94	**55**
Faroe Islands	375	**2**	998	**6**
Finland	70685	**69710**	111540	**110452**
France	204600	**175053**	240792	**215166**
Germany	209726	**199606**	223805	**217401**
Greece	14415	**7862**	15394	**10837**
Hungary	870	**199**	454	**78**
Iceland	13593	**13052**	28074	**27004**
Ireland	71569	**71491**	68280	**68255**
Italy	4019	**1199**	3071	**1247**
Jan Mayen	113	**5**	76	**9**
Latvia	31	**9**	7	**3**
Liechtenstein	13	**11**	12	**10**
Lithuania	37	**6**		**0**
Luxembourg	7471	**7466**	16550	**16544**
Macedonia	149	**68**	77	**11**
Malta	11	**5**	5	**2**
Moldova	28	**8**	1	**0**
Monaco	2	**0**	3	**1**
Montenegro	46	**28**	80	**40**
Netherlands	258659	**255573**	354576	**353610**
Norway	118652	**106765**	152521	**135587**
Poland	3833	**1823**	3218	**2042**
Portugal	4849	**1010**	3335	**1070**
Romania	502	**183**	410	**94**
Russia	5074	**866**	4601	**726**
San Marino	3	**3**		**0**
Serbia	50	**19**	25	**2**
Slovakia	4736	**80**	22781	**1660**
Slovenia	1790	**1476**	1890	**1470**
Spain	225541	**172700**	241927	**193388**

Country	Crucif_all	Crucif_map	Caryophyl_all	Caryophyl_map
Sweden	166808	**135852**	196095	**171227**
Switzerland	1357	**247**	994	**157**
Ukraine	790	**191**	154	**31**
United Kingdom	585804	**582687**	596461	**594908**

Table 6b. Overview of the amount of georeferenced distribution records for Cruciferae and Caryophyllaceae families in EVA database by countries of Europe.

Country	Brassicaceae	Caryophyllaceae
Austria	11498	16797
Bulgaria	202	314
Croatia	2506	2885
Czech Republic	46521	54146
Denmark	166	728
Estonia	394	405
Faroe Islands	0	5
Finland	126	332
France	30	32
Germany	26916	49527
Greece	166	293
Hungary	53	33
Iceland	35	45
Italy	6966	9422
Latvia	91	522
Lithuania	20	129
Macedonia	408	2
Montenegro	8	668
Netherlands	35891	56645
Norway	340	1185
Poland	7224	11818
Portugal	0	1
Romania	3436	4184
Russian Federation	756	2332
San Marino	10	9
Serbia	2098	5090
Slovak Republic	16911	18414
Slovenia	7657	5896
Spain	540	974
Sweden	264	704
Switzerland	2809	2583
Ukraine	1158	2833
United Kingdom	3675	7886

Few examples of particular Cruciferae taxa (four species: *Cardamine bulbifera*, *Sisymbrium strictissimum*, *Cardamine matthioli*, *Sisymbrium polymorphum*) illustrating the above-mentioned patterns are presented here. GBIF and EVA data were translated into the AFE grids in order to enable exact comparison. Depending on the type of the distribution area of particular species, the percentage of intersection of AFE and GBIF data, expressing the

approximate accuracy of the GBIF data set, ranged from 40.42 to 1.2 %, equivalent data for EVA range from 23.34 to 3.03% (Table 7, see also Fig. 20, Annex A)

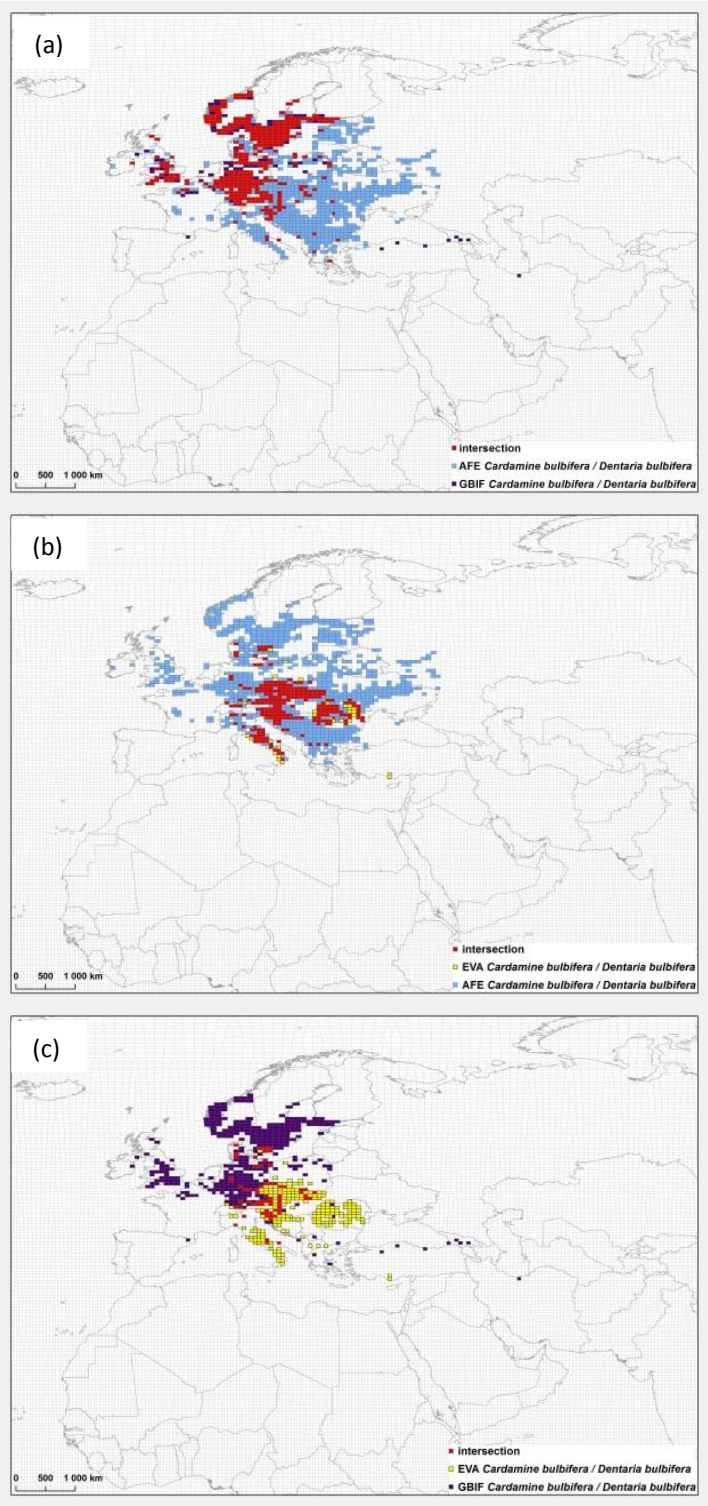


Fig. 20: Graphical comparison of AFE, GBIF and EVA datasets for an example taxa of the family Cruciferae (Cardamine bulbifera/Dentaria bulbifera for (a) AFE and GBIF data, (b) EVA and AFE data and (c) EVA and GBIF data). For the taxa with the centre of distribution in Central and Western Europe GBIF and EVA complement each other reasonable well, but for the Eastern European taxa neither of these databases provides sufficient amount of data.

Table 7: Comparison of AFE vs GBIF and AFE vs EVA datasets for four example taxa of the family Cruciferae expressed as the number of AFE quadrants for which distribution data are provided by GBIF and EVA for these species.

Species	No. of quadrants	AFE quadrants	GBIF quadrants	Intersect. quadrants	
Cardamine bulbifera	1076	1016	496	436	
Sisymbrium strictissimum	450	431	82	63	
Cardamine matthioli	161	158	8	5	
Sisymbrium polymorphum	249	249	3	3	
	AFE quadrants only	**GBIF quadrants only**	**AFE quadrants only in %**	**GBIF quadrants only in %**	**Intersect. quadrants In %**
Cardamine bulbifera	580	60	53.90	5.58	40.52
Sisymbrium strictissimum	368	19	81.78	4.22	14.00
Cardamine matthioli	153	3	95.03	1.86	3.11
Sisymbrium polymorphum	246	0	98.80	0.00	1.20

Species	No. of quadrants	AFE quadrants	EVA quadrants	Intersect. quadrants	
Cardamine bulbifera	1037	1016	263	242	
Sisymbrium strictissimum	436	431	31	26	
Cardamine matthioli	168	158	19	9	
Sisymbrium polymorphum	264	249	23	8	
	AFE quadrants only	**EVA quadrants only**	**AFE quadrants only in %**	**EVA quadrants only in %**	**Intersect. quadrants In %**
Cardamine bulbifera	774	21	74.64	2.02	23.34
Sisymbrium strictissimum	405	5	92.89	1.15	5.96
Cardamine matthioli	149	10	88.69	5.95	5.36
Sisymbrium polymorphum	241	15	91.29	5.68	3.03

4.3.5 Recommendations

As the analysis showed, there exist several gaps in European plant distribution datasets. To overcome these gaps, here are some recommendations that show how data availability and quality could be improved:

- Encourage European countries to become GBIF members and/or to provide data via GBIF portal. This particularly concerns countries of Central, Eastern and South-Eastern Europe and Italy.
- Provide financial support from EU sources to the national vegetation databases, members of the European Vegetation Survey, in providing unrestricted access to the distribution data from their datasets (as is currently done by the Czech Republic). National vegetation databases contain considerable amount of distribution data on European vascular plants, including the areas currently not covered by the GBIF. There is a good will on the side of providers to make these data public, but the main impediment in the countries of Central, Eastern and South-Eastern Europe is the lack of financial support on the national level.
- Provide financial support from EU sources to the Atlas florae europaeae at the University of Helsinki, Finland. Currently this serial publication, which is of European and perhaps also global importance, is supported only from the University sources and its continuation cannot be fully granted.

4.3.6 Literature

List of published volumes of the Atlas florae europaeae (volumes available for the gap analysis in electronic form are marked by asterisk)

Jalas, J. & Suominen, J. (eds.) 1972: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 1. Pteridophyta (Psilotaceae to Azollaceae). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 121 pp. [maps 1–150 + folded base map]

Jalas, J. & Suominen, J. (eds.) 1973: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 2. Gymnospermae (Pinaceae to Ephedraceae). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo,* Helsinki. 40 pp. [maps 151–200]

Jalas, J. & Suominen, J. (eds.) 1976: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 3. Salicaceae to Balanophoraceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 128 pp. [maps 201–383]

Jalas, J. & Suominen, J. (eds.) 1979: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 4. Polygonaceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 71 pp. [maps 384–478]

Jalas, J. & Suominen, J. (eds.) 1980: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 5. Chenopodiaceae to Basellaceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 119 pp. [maps 479–668]

*Jalas, J. & Suominen, J. (eds.) 1983: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 6. Caryophyllaceae (Alsinoideae and Paronychioideae). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 176 pp. [maps 669–1011]

*Jalas, J. & Suominen, J. (eds.) 1986: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 7. Caryophyllaceae (Silenoideae). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 229 pp. [maps 1012–1508]

Jalas, J. & Suominen, J. (eds.) 1989: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 8. Nymphaeaceae to Ranunculaceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 261 pp. [maps 1509–1953]

Jalas, J. & Suominen, J. (eds.) 1991: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 9. Paeoniaceae to Capparaceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 110 pp. [maps 1954–2109]

*Jalas, J. & Suominen, J. (eds.) 1994: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 10. Cruciferae (Sisymbrium to Aubrieta). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 224 pp. [maps 2110–2433]

*Jalas, J., Suominen, J. & Lampinen, R. (eds.) 1996: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 11. Cruciferae (Ricotia to Raphanus). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 310 pp. [maps 2434–2927]

Jalas, J., Suominen, J., Lampinen, R. & Kurtto, A. (eds.) 1999: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 12. Resedeaceae to Platanaceae. — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 250 pp. [maps 2928–3270]

Kurtto, A., Lampinen, R. & Junikka, L. (eds.) 2004: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 13. Rosaceae (Spiraea to Fragaria, excl. Rubus). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 320 pp. [maps 3271–3556]

Kurtto, A., Fröhner, S. E. & Lampinen, R. (eds.) 2007: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 14. Rosaceae (Alchemilla and Aphanes). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 200 pp. [maps 3557–3912]

Kurtto, A., Weber, H. E., Lampinen, R. & Sennikov, A. N. (eds.) 2010: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 15. Rosaceae (Rubus). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 362 pp. [maps 3913–4708]

Kurtto, A., Sennikov, A. N. & Lampinen, R. (eds.) 2013: Atlas Florae Europaeae. Distribution of Vascular Plants in Europe. 16. Rosaceae (Cydonia to Prunus, excl. Sorbus). — *The Committee for Mapping the Flora of Europe & Societas Biologica Fennica Vanamo*, Helsinki. 168 pp. [maps 4709–4878]

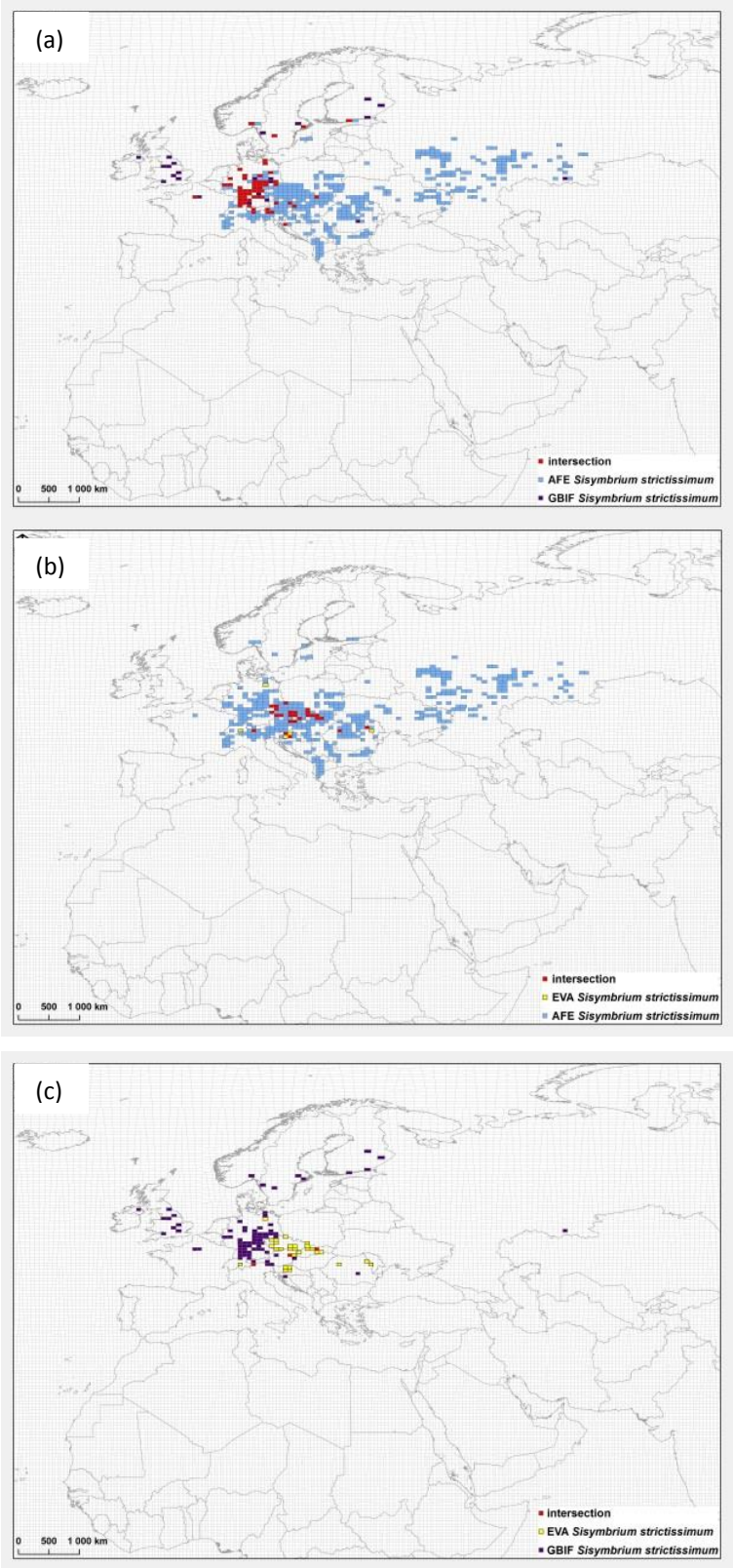
EVA references

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Dengler, J., Agrillo, E., Angelini, P., Apostolova, I., Becker, T., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Carlón, L., Casella, L., Csiky, J., Danihelka, J., Dimopoulos, P., Ewald, J., Fernández-Gonzáles, F., Fitz Patrick, Ú., Font, X., Gárcia-Mijangos, I., Golub, V., Guarino, R., Indreica, A., Jandt, U., Jansen, F., Kącki, Z., Kleikamp, M., Knollová, I., Krstonošić, D., Kuzemko, A., Landucci, F., Lenoir, J., Lysenko, T., Marcenó, C., Michalcová, D., Rodwell, J., Rusina, S., Seidler, G., Schaminée, J., Šibík, J., Šilc, U., Sopotlieva, D., Sorokin, A., Spada, F., Stančić, Z., Swacha, G., Škvorc, Ž., Tsiripidis, I., Turtureanu, P. D., Valachovič, M., Vassilev, K., Venanzoni, R., Weekes, L., Willner, W., Wohlgemuth, T. & Nordic Vegetation Database Consortium. 2014. ***European Vegetation Archive: now EVA really starts!*** In: Čarni A., Juvan N. & Ribeiro D. (eds), 23st International Workshop of the European Vegetation Survey ZRC Publishing House, Ljubljana, Slovenia, pp. 31-32 (available at http://evs.zrc-sazu.si/BookofAbstracts.aspx)

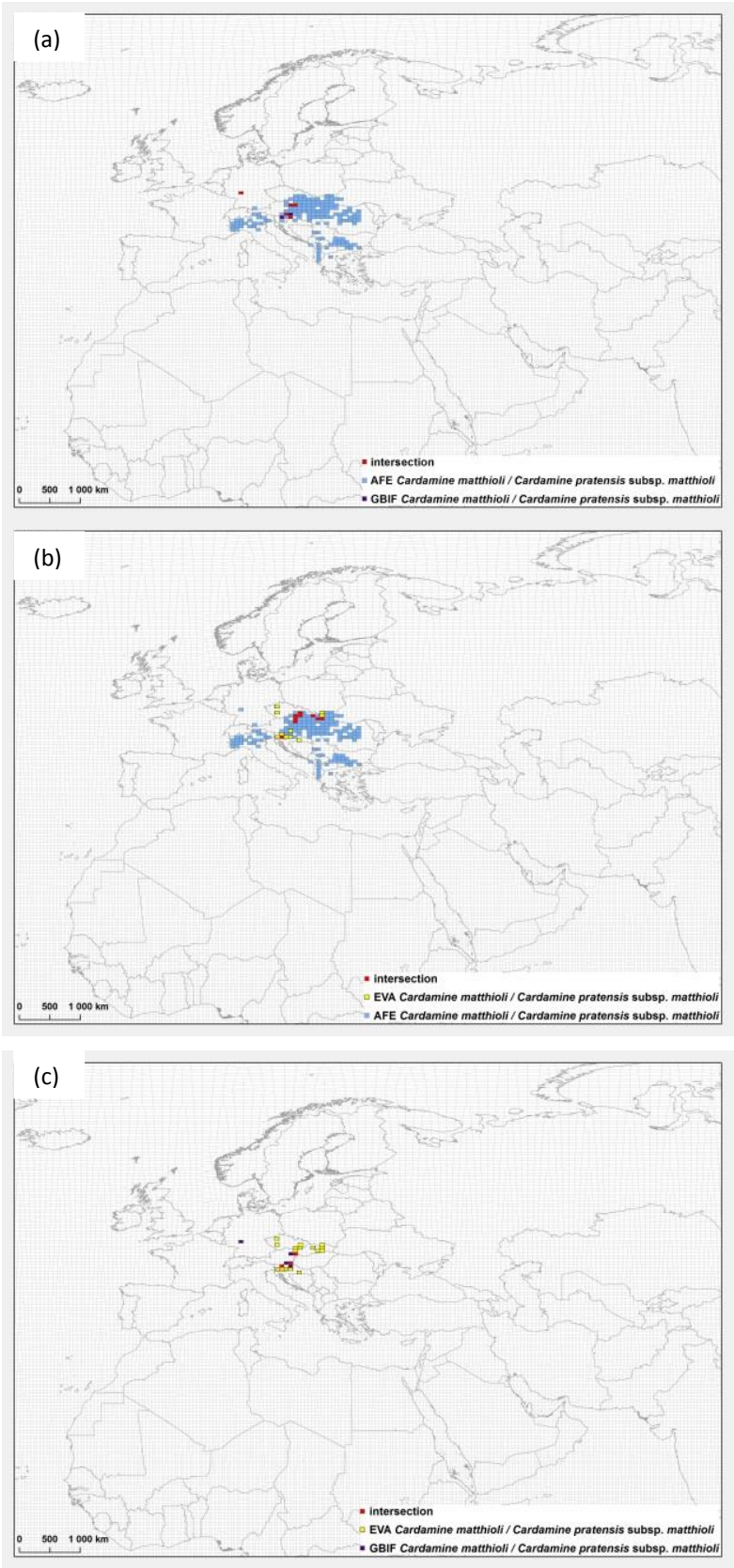
Chytrý, M., Berg, C., Dengler, J., Ewald, J., Hennekens, S., Jansen, F., Kleikamp, M., Landucci, F., May, R., Rodwell, J. S., Schaminée, J. H. J., Šibík, J., Valachovič, M., Venanzoni, R. & Willner, W. 2012. **European Vegetation Archive (EVA): A New initiative to strengthen the European Vegetation Survey**. 21st Workshop European Vegetation Survey. Vegetation databases and large-scale classification. Biogeographical patterns in vegetation. Vegetation and global change, University of Vienna, Austria, p. 12.

Annex A: Graphical comparison of AFE, GBIF and EVA datasets for three example taxa of the family Cruciferae (for (a) AFE and GBIF data, (b) EVA and AFE data and (c) EVA and GBIF data). For the taxa with the centre of distribution in Central and Western Europe GBIF and EVA complement each other reasonable well, but for the Eastern European taxa neither of these databases provides sufficient amount of data.

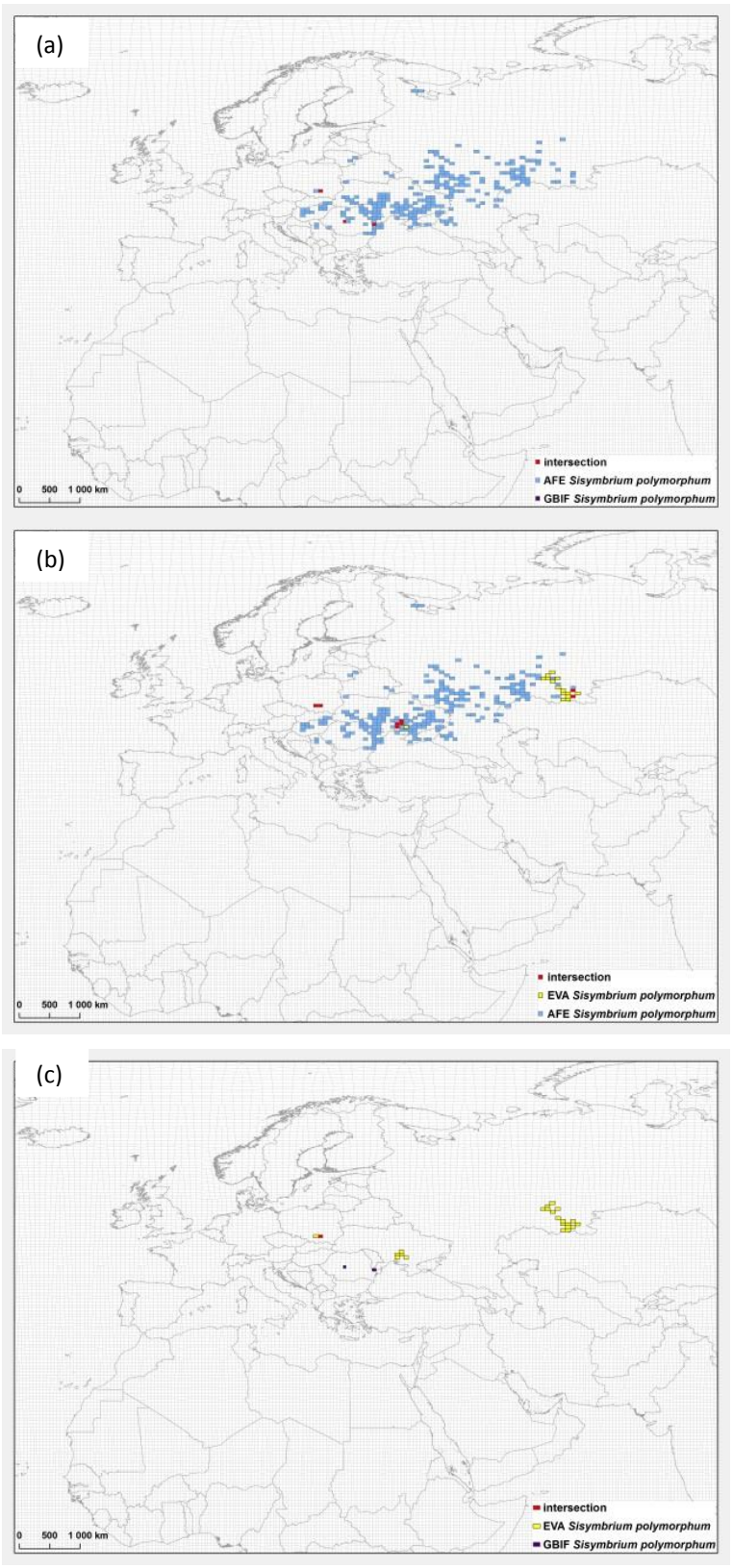
Sisymbrium strictissimum



Cardamine mattioli / Cardamine pratensis subsp. mathioli



Sisymbrium polymorphum



4.4 FOCUSED-REVIEW OF GAPS IN SPECIFIC DATABASES: GAP ANALYSIS ABOUT MARINE SPECIES DISTRIBUTION AND TRAITS

This gap analysis is science-oriented. It lists a number of gaps that scientists deem to be addressed to answer policy makers' questions. However, policy makers will ask questions at higher level of knowledge integration (see the BioFresh report on the Gap analysis for Policy makers), but this is not the goal here to link scientists' data and information gaps with policy makers' knowledge gaps. For the purpose of clarity, we include a short glossary at the footnote.[14]

4.4.1 Introduction and Data Sources

The main target for the marine species are all marine vertebrates (FIN) for the global analysis, and all marine species for Europe (HCMR).

The development of the so-called "Niche modeling" has triggered the creation of the Global Biodiversity Information Facility (GBIF) for all species, and the Ocean Biogeographic Information System (OBIS) for the marine life zone. Both organizations aim at gathering the maximum of occurrence data, if possible precisely geo-referenced point data.

After almost 15 years, two main general conclusions are well established:

- Data are scarce. Very few species have many data allowing long-term studies (or even seasonality studies), most of species do not have electronically recorded data or a few (either data do not exist, or are not computerized at all, or computerized data are not delivered to these aggregators for IPR, technical, or ignorance issues).
- Data are not of good quality. In addition to sampling bias, data are, in general, poorly curated by providers, and feed-back mechanisms from aggregators to providers worked in a handful of cases only.

For marine species, the situation is somewhat better thanks to the effort of the Census of Marine Life for OBIS. Some quality control and data cleaning are being undertaken in collaboration with the World Register of Marine Species (WoRMS).

See http://www.iobis.org/about/statistics for the current statistics in OBIS. Currently, only half of the estimated 115,000 valid species that have an occurrence data in OBIS have more than three points. The number of valid species represents 'only' half of the expected 230,000 marine species of which about 221,000 are currently listed in WoRMS.

FishBase is a global species database of fish species (Froese and Pauly, 2000). It contains comprehensive species data, including information on taxonomy, occurrence and

---

[14] ***Glossary:***

"Biodiversity data" -There is a current trend to use the locution "biodiversity data" as a strict synonym of "point data". Beyond the communication effect of the buzz word "biodiversity", we think that it is quite misleading, and we urge colleagues to use that locution only for datasets that cover or potentially cover the three levels of biodiversity, genes, species and ecosystems.

Point data - A point data is a minimum information triplet comprising a taxon name, a location (locality name and/or geo-coordinates), and a date.

Occurrence data - An occurrence data is a minimum information doublet comprising a taxon name and a geographic area (from a continent/ocean down to a point data), usually with a third information, the occurrence status (endemic, native, introduced, etc.). The major difference with a point data is the usual absence of the time dimension (but it can be indicated as well), or the implicit understood as the historic period.

geographical distribution, biometrics and morphology, behavior and habitats, ecology and population dynamics as well as reproductive, metabolic and genetic data. It also features a number of tools such as faunal checklist, identification keys and field guides, trophic pyramids and fishery statistics among others. FishBase is linked to other databases and global initiatives such as the Catalog of Fishes, GenBank, LarvalBase, GBIF, OBIS, FishBoL and the IUCN Red List. It currently has information for over 32,800 species, 303,100 common names, 49,800 references from the literature, and 56,100 pictures.

Catalog of Fishes as of web published version 19 May 2014: The Catalog of Fishes (CofF) of W.N. Eschmeyer (California Academy of Science) is the world taxonomy and nomenclature authority database for fishes. The Catalog contains around 60,000 names of fishes and 34,000 references to date. It is updated every four to eight weeks.

SeaLifeBase is patterned after FishBase and maintains an information system for all other aquatic organisms, estimated at around 400,000 species. Nonfish marine organisms, numbering roughly 200,000 species, are the current target of the project, with focus on metazoans first followed by marine plants. Key information include data on life history, trophic ecology, and marine biodiversity lists with the goal of making available the biological information necessary to conduct biodiversity and ecosystem studies.

World Database on Protected Areas as of web published version November 2013: WDPA is the most comprehensive global database on terrestrial and marine protected areas, and is a joint project of IUCN and WCMC. Protected area coverage and statistics are generated assessing progress towards international biodiversity protection targets. Here, only the global data set for marine protected areas was used.

4.4.2 Results

The results are discussed in three levels of organization: ecosystem, species and population/genetic level. For the analyses on the ecosystem and species levels, we used point data from GBIF/OBIS rounded to Half-Degree Cells (HDC) with 0.5ºlatitude x 0.5º longitude resolution and without time dimension. We also used FishBase (for occurrence data, species traits and population/genetic data) and SeaLifeBase (occurrence data), the IUCN-WCMC World Database on Protected Areas (marine protected areas), and the web version of Echmeyer's Catalog of Fishes (for data on the number of fish species described). Results of targeted web searches were also used for identifying potential, unexploited and inaccessible sources of data.

Gaps on a Ecosystem level: Comparison of country occurrence data between OBIS/GBIF and FishBase/SeaLifeBase [15]

At the level of ecosystem, particular geographic areas are the national territories and their marine extensions, the Exclusive Economic Zones (EEZ). Species are not constrained by administrative boundaries, but all conservation measures even if regulated at regional or global levels are enforced on the ground by countries and their administrative subdivisions. It is then crucial that governments and their administrations are well aware of which species their territories host.

[15] FishBase and SeaLifeBase data, as published on the web under the version February 2014, for more information see fishbase.org and sealifebase.org.

To estimate the knowledge in that domain, we have compared for marine vertebrates (fishes, reptiles, and marine mammals) how many occurrence data are reported for countries in the literature (as recorded quite completely in FishBase and SeaLifeBase), with how many country records can be inferred from HDC occurrence data, which is the scale at which the AquaMaps are elaborated. The number of countries per species in the two cases should be equal.

Fishes

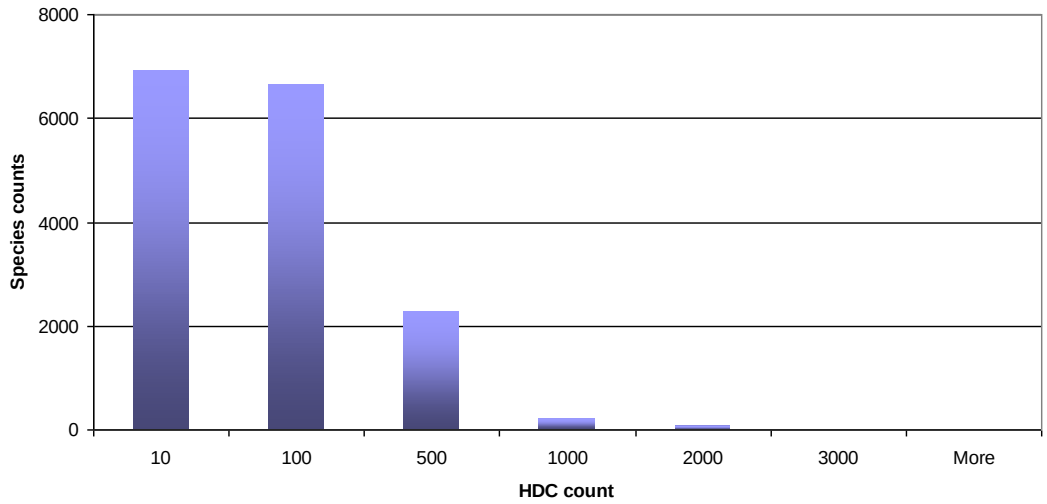


Fig. 21: Number of occurrence data by Half-Degree Cell (HDC) per species computed from GBIF/OBIS/ FishBase data for marine fishes.

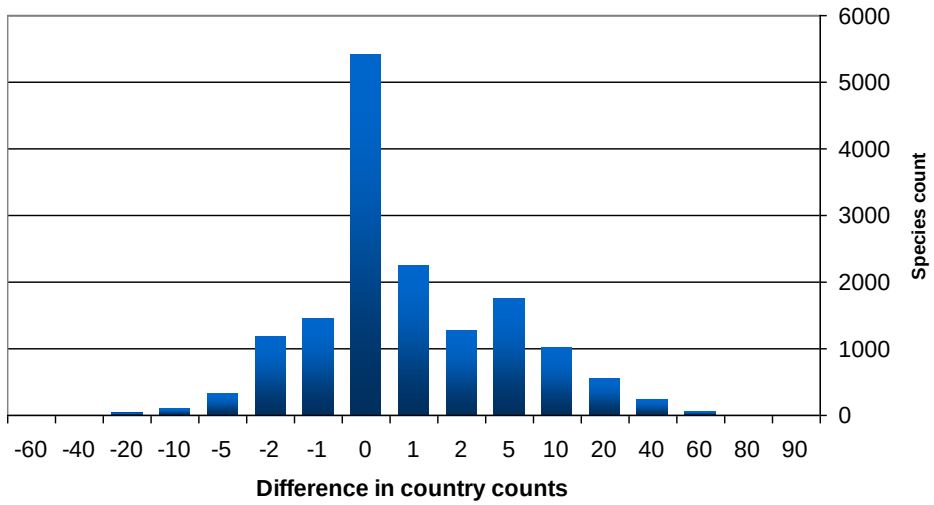


Fig. 22: Gap in the number of countries covering the native range of marine fish species. The gap is expressed as the difference in the number of countries occupied by a species based on occurrence points and the number of countries where a species naturally occurs based on the literature as documented in FishBase. The x-axis shows the range of difference (axis values represent upper limits) while the y-axis shows the number of species falling within a range. Positive values in the x-axis represent cases where species country count based on point data are higher than species country counts in FishBase. Negative values represent cases where country counts in FishBase are higher than species country counts based on occurrence point data..

The Fig.21 summarizes the data with regards to the number of occurrence data. Of approximately. 17,100 fish species occurring in marine waters (incl. 780 diadromous

species), point data are available from GBIF/OBIS for about 16,100 species (94%). On the average, there are seven point data per HDC per species, while the maximum point data for a species is 21,000. Large data records can typically be attributed to the industrial commercial species that are surveyed intensively by fisheries institutes and departments around the world (but data may be protected or fuzzed). So a few species, especially cod, tunas, and the like, may have thousands of records over the past 50 years.

All ca. 13,600 species over ca. 16,100 (84%) with a difference between -5 (more country occurrences by species from literature than from point data) and 5 (the reverse) can be considered well documented at country scale (Fig. 22). There are a number of small countries where the species are not reported from the literature or where point data are missing which explains the differences.

However we did not expect that the majority of larger differences (ca. 1,800 over 2,000) would be about more country occurrences from point data than from literature (Fig.22). The analysis of the species involved shows that it is in large majority oceanic species, either mesopelagic (Myctophidae), bathypelagic or bathydemersal, and a few epipelagic. It means that large ichthyofaunas, in particular those associated to countries, primarily cover coastal fishes, and that the oceanic species are ignored when they do not reach the coasts. This is of concern in the view of the increase of the non-regulated exploitation of high seas, seamounts in particular; that is, even when these waters are within an EEZ, their biodiversity is not well known. There are a few cases where a former species with widespread distribution has recently been split in two or several species but the point data were not corrected.

The 590 species without allocated countries are the oceanic species living only in the middle of the Atlantic-, Pacific- and Indian Ocean, and do not occur in any EEZ. These are mostly deep sea species. Note: Encyclopedia of Life (EoL) has a Biosynthesis Project for establishing the list of all deep sea fishes which started through a workshop gathering most of the specialists in the domain. Unfortunately, the country distribution is not considered as a priority; and, the project is not yet finished as the position of the principal actor is currently vacated. Funding to fill the position to finish this work could be helpful.

Only 22 species of reptiles over 91 have geo-referenced point data. All marine turtles (7 of 7), saltwater crocodiles (3 of 3), and the marine iguana (1 of 1) are represented, while only 11 sea snakes over 80 species are covered.

For reptiles, the figure is comparable to fishes, but at a lesser degree (Fig. 23). It is mainly six of the seven marine turtles that are more reported in countries by point data than by literature. As they are emblematic conservation species in many countries, it seems that we need to check SeaLifeBase and complete the country list from unseen literature.

The case of marine turtles will be treated with marine mammals that show the same pattern. For iguanas the match is perfect. For crocodiles, the match could be better but do not represent a concern.

Reptilia

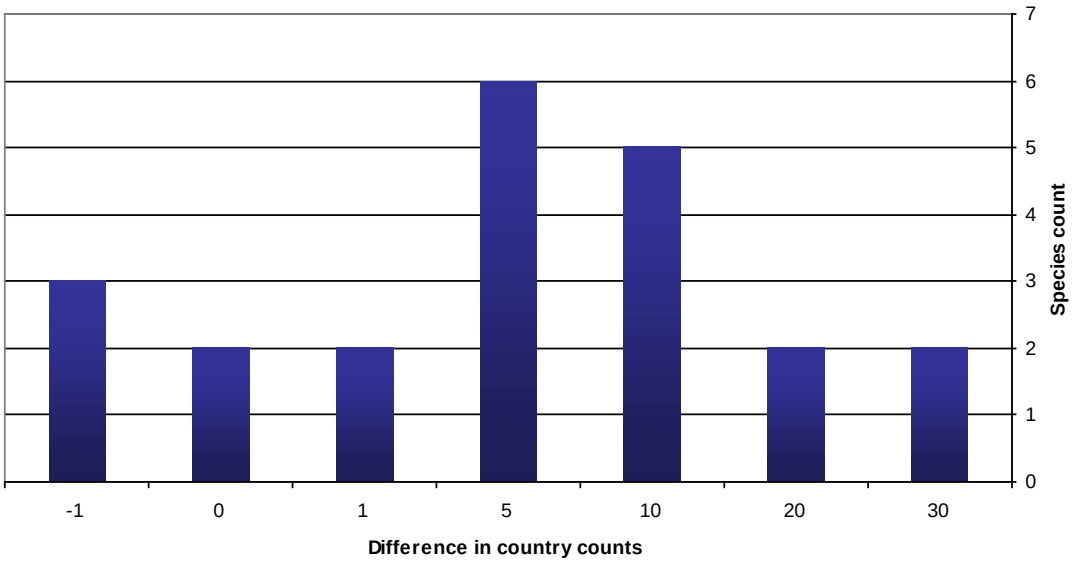


Fig. 23. Gap in the number of countries covering the native range of marine reptiles. The gap is expressed as the difference in the number of countries occupied by a species based on occurrence points and the number of countries where a species naturally occurs based on the literature as documented in SeaLifeBase. The x-axis shows the range of difference (axis values represent upper limits) while the y-axis shows the number of species falling within a range. Positive values in the x-axis represent cases where species country count based on point data are higher than species country counts in SeaLifeBase. Negative values represent cases where country counts in SeaLifeBase are higher than species country counts based on occurrence point data.

Boa constrictor has been spotted in several countries on beaches and consequently reported from the supra-littoral zone of marine areas, but obviously, point data could hardly be in the real marine water area. *Acrochordus granulatus*, is a freshwater and brackish water species occasionally found at sea, so it has very few marine point data. These two species are not a concern but they are representative of those species living at the edge of the marine life zone, like those freshwater species occurring also in brackish water. Caution must be exercised when performing analysis where these species should or should not be included knowingly.

For other sea snakes the concern is less the country record difference than the large gap of geo-referenced point data for 84% of species (69 over 80). These snakes are not collected by usual oceanographic cruises but through dedicated campaigns conducted by specialists knowing where to find and how to catch them, which limits the number of available data. However catches and observations are reported in the literature.

Mammalia

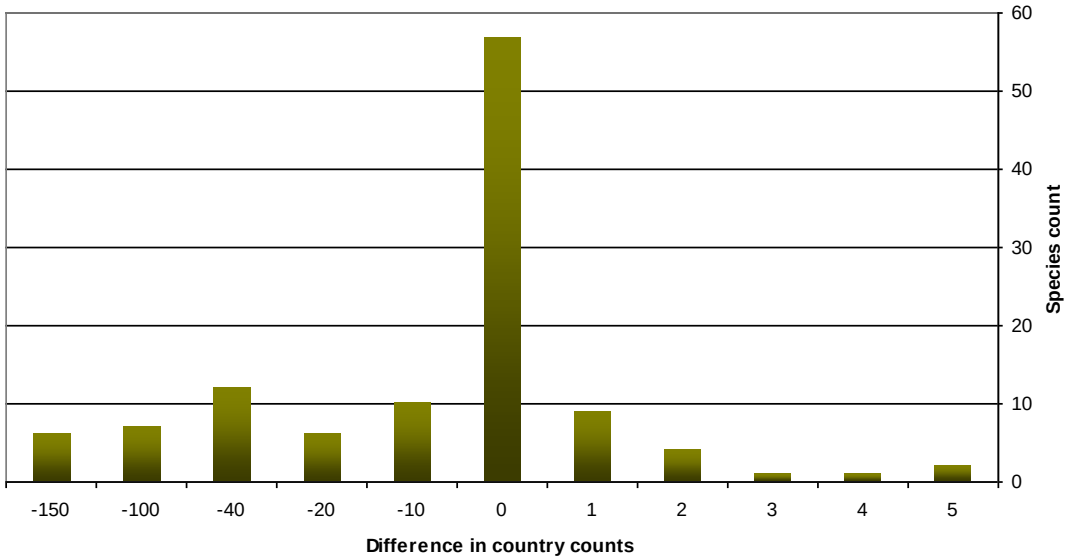


Fig. 24: Gap in the number of countries covering the native range of marine mammals. The gap is expressed as the difference in the number of countries occupied by a species based on occurrence points and the number of countries where a species naturally occurs based on the literature as documented in SeaLifeBase. The x-axis shows the range of difference (axis values represent upper limits) while the y-axis shows the number of species falling within a range. Positive values in the x-axis represent cases where species country count based on point data are higher than species country counts in SeaLifeBase. Negative values represent cases where country counts in SeaLifeBase are higher than species country counts based on occurrence point data.

Although presenting a better match than sea turtles on the positive differences, marine mammals present an unbalanced figure on negative differences; i.e., there are more country records from literature than deduced from point data (Fig. 24). Like turtles, marine mammals are the subject of many surveys by scientific institutions and conservationist NGOs, so it is doubtful that there is a lack of point data for 41 marine mammal species over 142 for so many countries. It means that the data for these species, mainly resulting from visual surveys and not from museum collections (with respect to their size), are not served to aggregators yet.

This analysis could be repeated with Large Marine Ecosystems (LME) or with other system of ecoregions (e.g., WWF-TNC MEOW[16]). For LMEs, data are available for vertebrates in FishBase and in SeaLifeBase, but not complete enough in SeaLifeBase and WoRMS to draw conclusions for the invertebrates (just like for countries above). For ecoregion systems, there is usually no complete allocation of species by ecoregion. This allocation can be done by overlaying point data or published distribution polygons to the ecoregion layer. However, while having an independent allocation from the literature allows quality control cross-checking, we would lose the independence of two datasets to be compared.

For very large areas like realms or FAO areas, the results are much closer to a perfect match, which means that global analyses at those scales can be performed with the existing data.

[16] http://www.worldwildlife.org/publications/marine-ecoregions-of-the-world-a-bioregionalization-of-coastal-and-shelf-areas

We used the half-degree cell occurrence data for marine fishes (see previous section) which we filtered using the World Database on Protected Areas MPA layer (IUCN and UNEP-WCMC, 2013). This includes areas that have a special status but that are not managed as conventional marine protected areas (MPAs): fishing restricted areas, listed natural sites to be monitored, international convention areas (e.g. OSPAR). We have also used half-degree cells that may include an MPA where some species may not find the correct habitat. Fig. 25 shows a summary map of these data.

About 10,500 marine fish species have occurrence data in marine protected areas globally. This covers approximately 61% of over 17,000 fish species occurring in the world's oceans. In European waters, around 800 marine fishes have been recorded from within MPAs, covering roughly 63% of over 1,200 fish species known from seas around Europe.

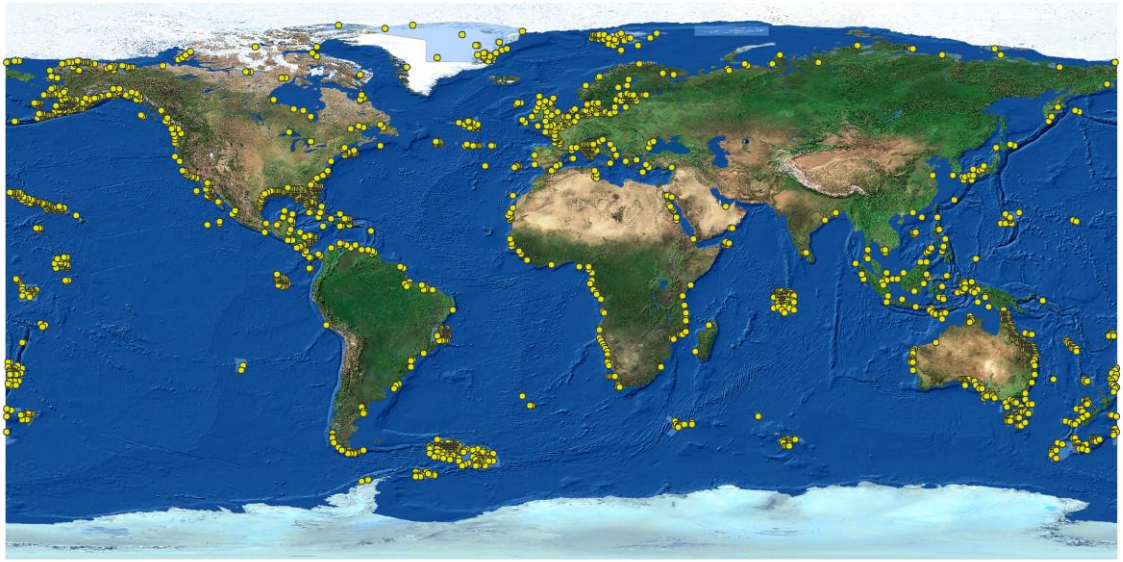


Fig. 25: Occurrences of fish species recorded from MPAs (IUCN and UNEP-WCMC, 2013).

We also compared marine fishes in Europe with and without occurrence data in MPAs and examined their data coverage in terms of commercial importance, IUCN Red List and habitat (generally considered to be the adult feeding and breeding environment). A summary of the comparisons of data coverage is provided in Fig.26a-26c.

Most species that are of importance to European fisheries (i.e., highly commercial, commercial, minor commercial and subsistence) have been found within MPAs (Fig. 26a). About 76% of these species (368 out of 483) have occurrence points within MPAs. On the other hand, for species that are currently of no commercial interest, 60% (141 out of 232) have occurrence points within MPAs.

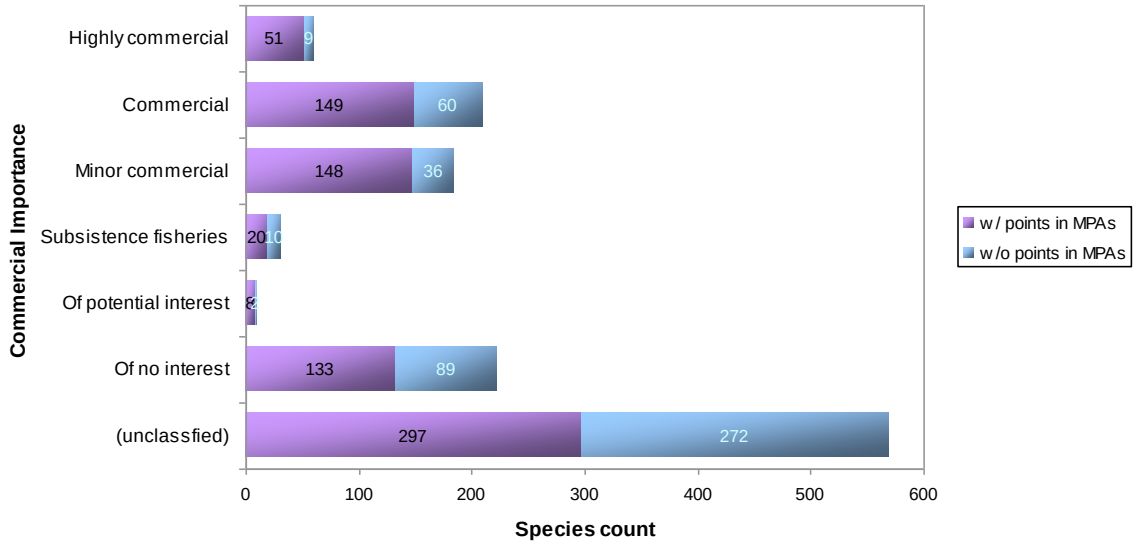


Fig. 26a: Comparison of data coverage on commercial importance for marine fishes with and without occurrence points in marine protected areas in Europe.

About 569 species have not been recorded in fisheries statistics and are currently regarded as 'unclassified.' This group could be combined with species classified as 'of no interest' to come up with a total for species that are currently do not contribute to fisheries. That would allow comparison between the numbers occurring in MPAs for species that are of importance to fisheries and those are not. Comparing these two groups, about 76% of species of importance to fisheries and 55% of species that currently do not contribute to fisheries have occurrence points within MPAs.

About 64% of the species categorized as 'threatened' under the IUCN Red List (critically endangered 60%; endangered 69%; vulnerable 79%) have occurrence points in MPAs (Fig. 26b). For species that have been found to be data deficient for IUCN threat evaluation, 41% (32 of 79 species) have occurrence points within MPAs. Around 921 species are either not yet evaluated or not categorized. Of these, roughly 63% have occurrence points in MPAs.

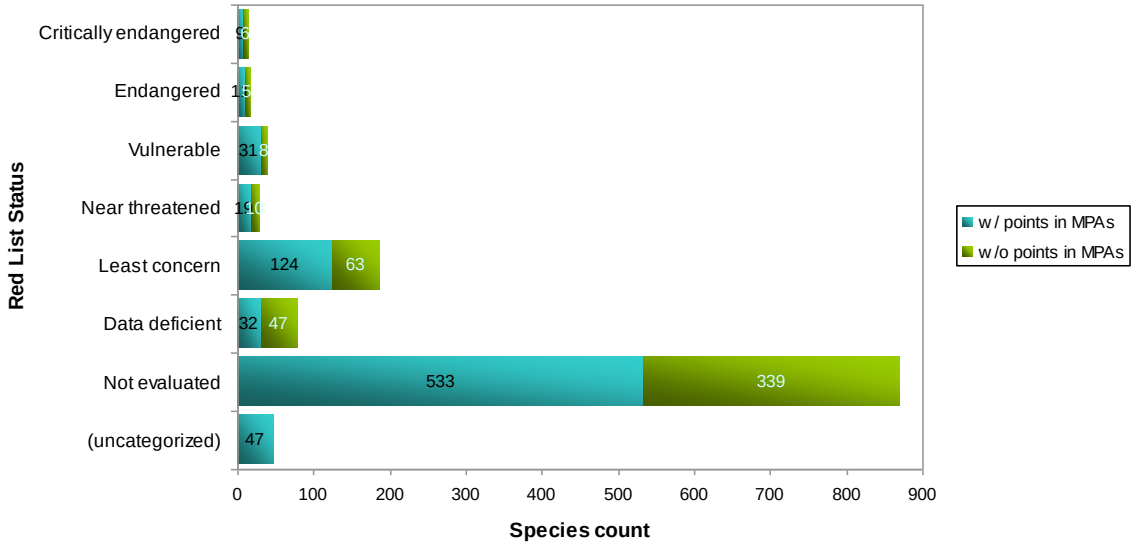


Fig. 26b: Comparison of data coverage on IUCN Red List threat status for marine fishes with and without occurrence points in marine protected areas in Europe.

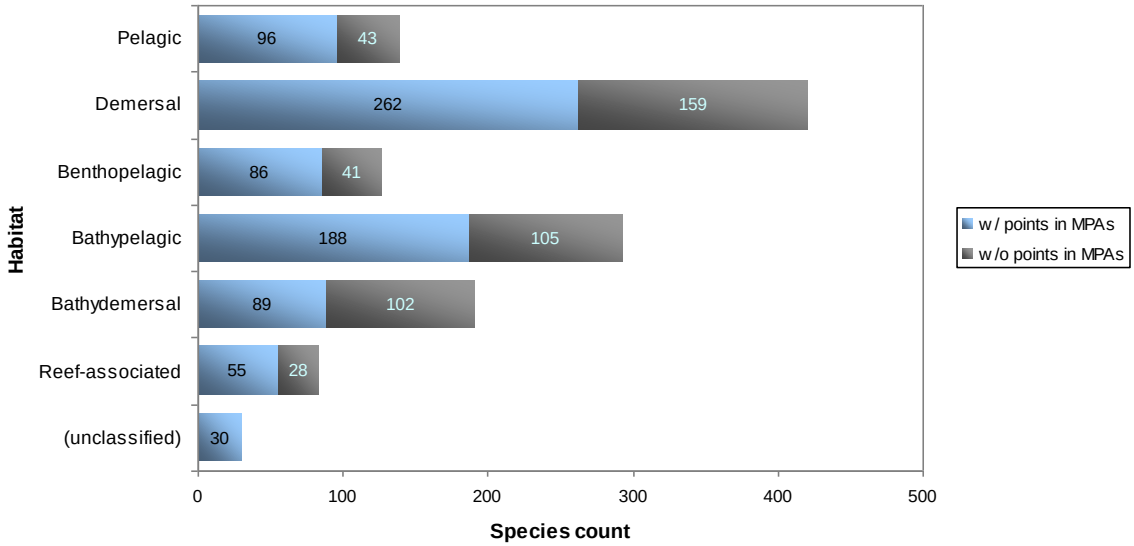


Fig. 26c: Comparison of data coverage on IUCN Red List threat status for marine fishes with and without occurrence points in marine protected areas in Europe.

On average 62.5% of the species, which can be classified by habitat type, have geo-referenced occurrences in MPAs (Fig. 26c). The species are spread across different habitat types and have the following composition: pelagic 69%, demersal 62% benthopelagic 68%, bathypelagic 64%, bathydemersal 47%, and reef-associated 66%. The relatively low number of bathydemersal species with occurrence points in MPAs may be due to the relatively shallow depths covered by MPAs, and probably needs to be investigated further. Overall, occurrence points in the MPAs cover demersal species (50%) slightly more than pelagic species (46%).

Gaps on a species level: Analysis of gaps in knowledge of species traits:

FishBase database (as of 31 January 2014) was queried to examine information gaps in key tables covering data on life history traits, ecology and genetics of marine fishes, as well as picture collections. Table 8 gives a summary of data available for a number of important traits for fishes, and include the total number of records per topic, the number of species with records, and the percentage of this number with respect to the total number of species in the database.

Topics	No. of Spp. (1)	% Spp. (2)	No. of Records
Reproduction	6,359	37%	6,377
Spawning	2,057	12%	4,124
Larvae	2,409	14%	2,421
Eggs	2,632	15%	2,639
Maturity	1,920	11%	5,400
Maximum length	4,898	29%	10,172
Growth	1,551	9%	7,106
Mortality	1,551	9%	7,106
Length-weight-relationship	2,757	16%	8,406
Ecology	6,979	41%	6,989
Food	5,484	32%	36,583
Diet	1,881	11%	49,747
Predator	1,686	10%	5,328
Genetics	1,058	6%	2,953
Pictures	11,124	65%	38,109

Table 8: Gap analysis of traits in FishBase for marine fishes . (1): number of species with at least one record for the corresponding topic; (2): percentage over 17,191 marine species.

FishBase has gathered fish data from over 27,000 references to date. Table 8 list some of the key topics covered by the database. In general, despite the amount of literature covered, data on most traits cover less than 16% of marine fish species with the exception of reproduction, maximum length, ecology and food items which cover about 30%-40% of species. There is good coverage for fish pictures which currently cover about 65% of marine fishes.

Examination of number of records against the number of species covered provide useful information with regards to the sparseness and biases in data coverage in the literature. For instance, there is effectively one record per species for information on reproduction, larvae and eggs. The other traits have around three records per species on average. Data on diet, on the other hand, is substantial with nearly 50,000 records yet covers only about 1,800 species, or 11% of marine fishes. From the biodiversity point of view, data is sparse with respect to the number of fish species covered and the amount of data on traits currently available.

One possible solution to get estimations for missing traits is to use the Bayesian approach which has started under the European project ECOKNOWS to estimate length-weight relationships (Froese *et al*., 2014). A summary of these efforts will be given for the final report from the reports from ECOKNOWS (ending in August 2014).

One possible solution for filling the gaps would be to engage students in marine biology to study less documented species instead of repeating their university work always on the same species (or at least in addition after training on well-known species). This would require a minimum of organization in universities but it could yield rapidly a number of data that could be published over the Internet.

One possible solution for increasing the number of occurrence data and habitat record is to engage in citizen science. A number of experiments have already proven that good data may be accumulated in such a way (Bodilis et al., 2014; Edgar et al., 2014), even if there are limitations using this approach to elaborate or test scientific hypotheses (Bernard et al., 2013; Dickinson et al., 2010) that however may be overcome by statistical treatments (Bird et al., 2014).

Gaps on a species level: Identifying potential, unexploited and inaccessible sources of data: trawling surveys, scuba diving visual census

Trawling surveys are conducted by governmental or scientific institutes for discovery of new fishing grounds or for evaluation the status of exploited stocks. Visual census surveys are conducted by scientific institutes in shallow waters where other sampling methods are inefficient or when non-destructive methods are required.

The aim of the work is to establish if reports and associated data are made freely available and easily accessible through the web.

Trawling surveys

An advanced search with Google on "data repository" (all these words) and "trawling survey" (this exact word or phrase) yielded only 26 links. All the links were explored to search for downloadable data.

It seems that very few trawling survey dataset are available on the web. The main reason is that these data are under strict IPR, and countries refuse to share what they consider as strategic economic data. WorldFish had an experience building a database (TrawlBase) with a number of ASEAN countries. It cannot be used for analyses as each country must give its agreement to use its data even if fuzzed. This important source of data, because it usually includes by-catch, is thus unavailable for biodiversity studies except for information produced by the fishery institutes or departments that have conducted the survey, and have disseminated fuzzed data as maps or highly aggregated tables only. The same is true for the European programme MEDITS.

If the word "repository" is removed, then the search yields 12,400 links (yet with many duplicates, e.g., the Guinean Trawling Survey). A subsample will be analyzed, excluding the 72 links as above.

Visual census surveys

An advanced search with Google on "data repository" (all these words), visual census survey (this exact word or phrase) or "underwater", "Scuba diving" "skin diving" (any of these words) yielded only 2 links, and without 'repository', 4,740.

Gaps on a population/genetic level: Primary analysis on gaps at stock/population level - elaboration of a proposal for future projects

The aquatic genetic resources are not document at global level like for crops, cattle, poultry, and forestry. For more than 15 years when this question was raised by WorldFish to FAO, there is a lack of political will to engage the world community to accumulate data at population level.

FishBase and SeaLifeBase have been structured to record such information, but the lack of funded project explains the very low numbers for only 135 species presented in Table 9 compared to a total of about 7,500 fish species that are used by humans in a way or the other (food, aquarium, fish game and even for spa).

Rank	No. of Stocks	No. of Species

Cultures strain	74	8
Hybrid	10	6
Subspecies	51	30
Wild stock/population	270	91
Total	**405**	**135**

Table 9: Number of records in FishBase for stocks/population/strain ranks.

The stock information were mainly collected for European marine fish stocks for the European project ECOKNOWS. The aquacultured strains are more about developing country commodities (tilapias in particular).

Gaps on a population/genetic level: Species that lack barcodes in the Barcode of Life Data System

As of 1st of March 2014, there are 10,185 fish species (that have at least one barcode in the BoLD system (for 94,836 specimens, about 10 specimens per species on the average).

There are already a number of families where the mitochondrial cytochrome c oxidase 1 gene (CO-I) barcode does not work for various reasons:

- Acipenseridae (sturgeons): the lineage is quite ancient and it seems that the evolution of CO-I has been reduced since the radiation in XXX (period/age).
- Cichlidae (cichlids): in the great East African Lake, the radiation is too recent and explosive (flocks of species), mutations are not fixed yet.

Other markers exist but may have not been studied yet for these groups.

Gaps on a population/genetic level: Species still to be described

The literature on this topic is quite abundant. For the marine species Costello et al. (2010)

As of February 2014, there are 33,065 valid described species recorded in the Catalog of Fishes (CofF), whereas there are about 32,800 species recorded in FishBase. The difference is due to the delay between the acquisition of literature and its full exploitation for morphology, identification keys, pictures, distribution, biology, ecology and occurrence data compared to CofF which is limited to taxonomy and nomenclature, and short textual distribution statements. The trend of fish species description since Linnaeus 1758 is given on Fig. 27.

At the turn of the millennium in 2001, Eschmeyer and Froese presented a poster at the 10th European Congress of Ichthyology in Prague where they predicted a total number of extant finfish species about 35,000 but already the number of 33,000 was surpassed in November 2013 (estimation N. Bailly from various sources).

As we can see on Fig. 28, since that prediction, the number of species described per year increased significantly. Moreover the tendency of the curve is slightly geometric (see the cumulative curve Fig. 29), not even linear, which means that nobody is able today to compute a reasonable estimation of the total number of extant finfish species.

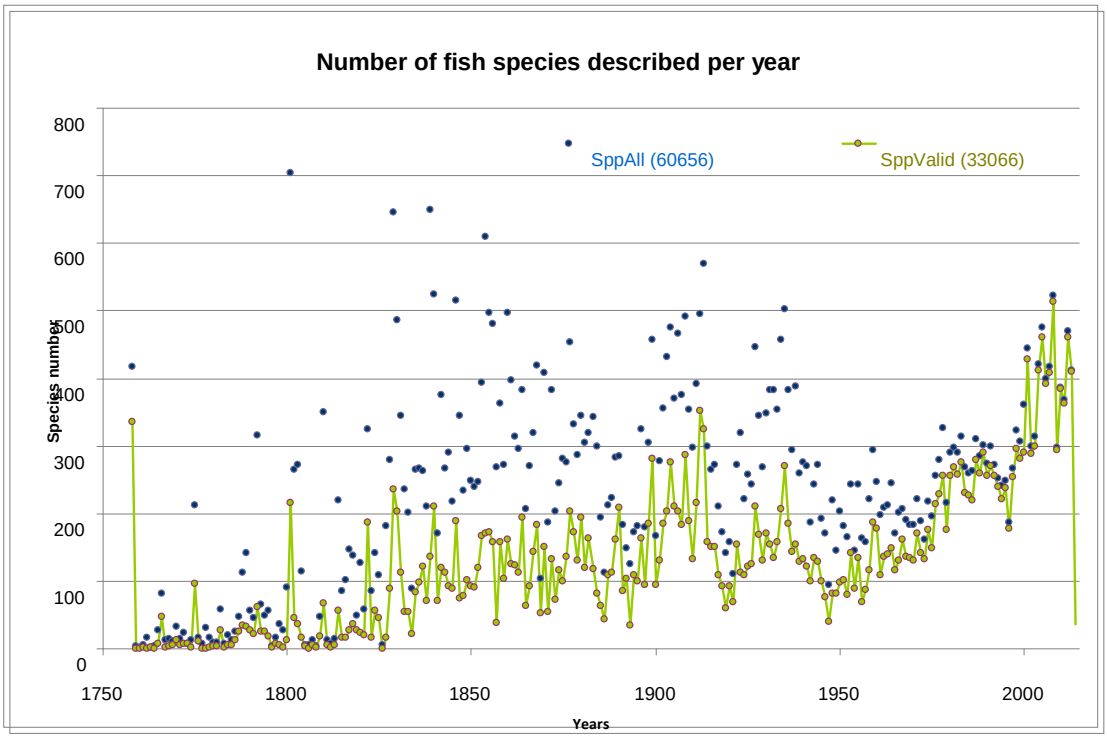


Fig. 27. Number of new species described per year since Linnaeus (1758) up to 2014 [Source of data: Eschmeyer – Catalog of Fishes – web version 05 February 2014.]. 2014: already 56 new species are described as of 1st March 2014, source: Mikšik and Schraml - Welt Der Fische).

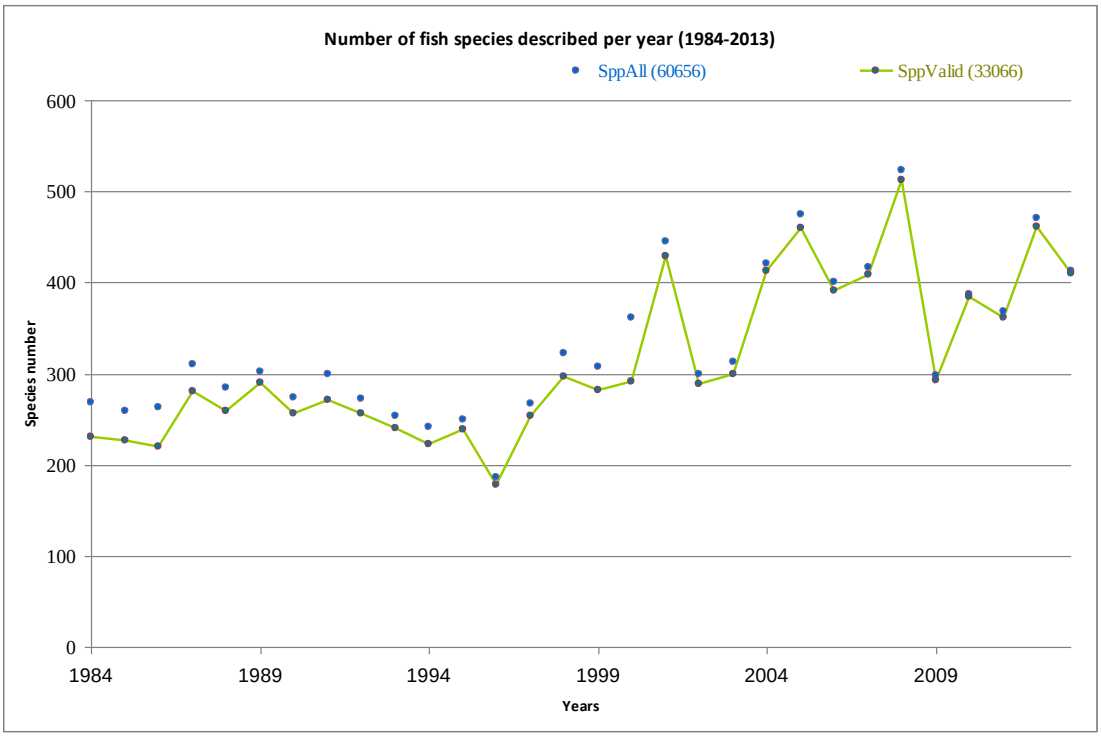


Fig. 28: Number of new fish species described per year between 1984 and 2013 (the past 30 years) [Source of data: Eschmeyer – Catalog of Fishes – web version 05 February 2014.]

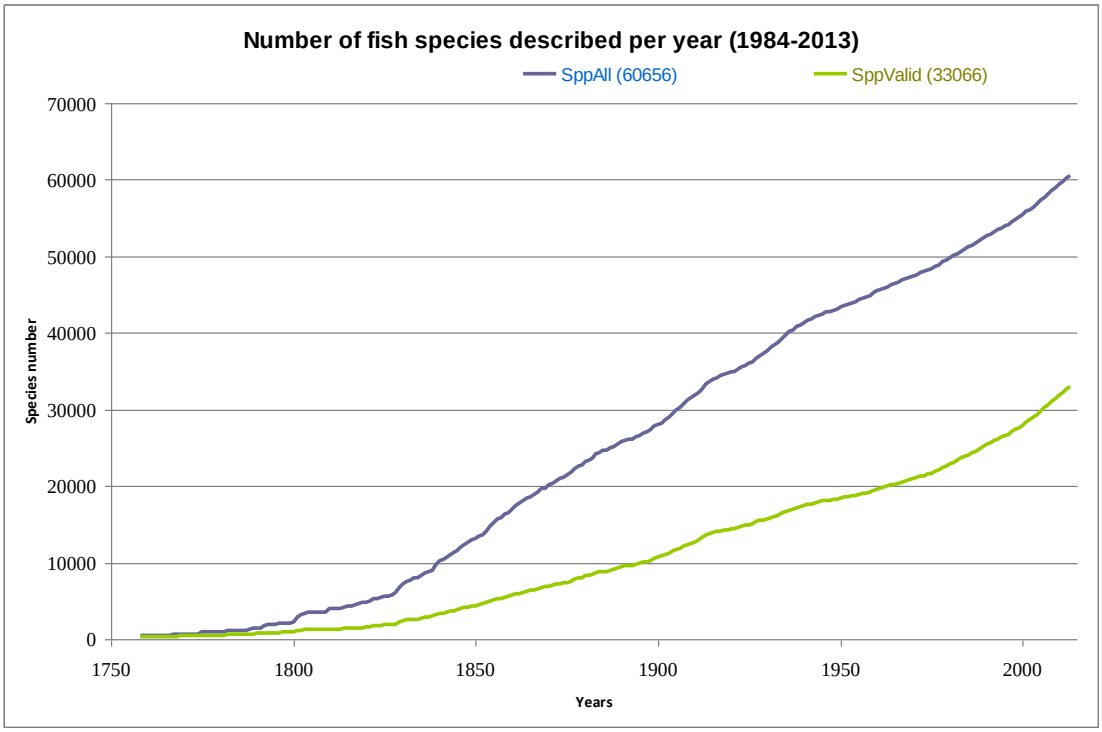


Fig. 29: Cumulative number of fish species described since Linnaeus (1758) up to 2014 [Source of data: Eschmeyer – Catalog of Fishes – web version 05 February 2014.]

Recent experiments in freshwaters of Brazil and Kenya yielded three times more species with barcoding approach than compared with what is known. In marine environment in Info-Pacific, a study on *Trimma* genus (Gobiidae) yielded 1.8 more species with the same methods. Another current trend is to elevate the Red Sea populations at species rank to separate them from the Indian Ocean populations; already 20 of those were recently described in the past three years.

However these potential new species are not all yet confirmed by further analysis. And there are debates about the species concept and what should be the barcode difference threshold for a species. Currently, some 2,000 fish barcodes over approximately 95,000 barcoded specimens are unnamed in the barcode BoLD system.

In the current state of knowledge, it is difficult to imagine that all large marine species will be split into three different species each, or that there will be as many new discoveries. The use of depths sounds promising, but there are only few recent expeditions and less and less specialists, so a very little percentage of new species per year are described from deep sea. However as an example for Ophidiiformes, there could be as many as 100 new species awaiting their description (Nielsen and Møller, 2013, pers. comm.). Too little is known about the larval cycle and the capacity of dissemination of the deep sea species.

Thus, supposing there are three times more species in total than currently accepted as valid, the number of finfish species would reach approximately 100,000 maximum. One prediction that is most reasonable is that eventually there will be more freshwater than marine species due to the isolation of basins/catchment compared to the marine environment, which facilitates speciation. It is confirmed by the increasing ratio of freshwater species described per year in the past 20 years (from 45 to 55% on the average).

4.4.3 Target High-level Questions

1. Can we identify status and trends of [European] species? Can we identify status and trends of biodiversity taking interspecific phylogenetic or intraspecific genetic diversity into account? Can we assess the risk of extinction?

- Data on traits (ecological, life-history, morphological etc.) of species

There is currently a trend in developing databases for traits and ecosystem services for marine species. Unfortunately, there isn't any aggregator system which can accumulate all the information made available although TRAITBANK was designed by EOL to play this role at the global scale (http://eol.org/info/516). At the EU scale, such an initiative has been undertaken by the EMODnet Biology project, coordinated by VLIZ. Other examples include : (1) FishBase for finfishes (www.fishbase.org); (b) PolyTraits for annelids (polytraits.lifewatchgreece.eu); (c) SeaLifeBase (www.sealifebase.org) for all non finfish marine organisms.

These databases are far from complete, and it is unclear what part is still unknown (i.e., species never studied) and how much of the published part is not yet encoded in the database. Estimations are time-consuming. The problem is aggravated by the following facts: (a) the auto-ecological work has not been funded unless for the species with a certain economic value; (b) the fundamental taxonomic work was never a priority for funding both from the EU and the state research funding instruments; (c) the quality of data is not always at the level that is needed to perform any analysis at the EU scales since every researcher makes choices about the traits to be included in each of the database and the ecosystem services they are connected to.

One option to temporarily fill the gaps is to use the Bayesian statistical approach as has recently been experimented under the EC FP7 ECOKNOWS project in FishBase for traits necessary for stock assessments (Froese *et al*., 2014).

One issue that has prevented data exchange in that domain is the lack of proper ontology: how to store information about species traits for a whale and for a crab in the same database. A previous effort led by GBIF in 2007 failed to achieve this goal, but a second attempt by the EC FP7 project EModNet has recently established a 4-Level hierarchical system that seems to accommodate this trait diversity. It remains to implement that ontology in the various databases and to develop the finest levels for each taxonomic group.

- Data on phylogeny / genetic diversity of species

Phylogeny has made recent progress at all levels thanks to the development of sequencing methods and of the bioinformatics domain. The animal phylogeny at phylum level is currently the object of many recent publications, but there are still some issues debated (Coelenterata hypothesis or not?). Data are made more and more available so there is no concern here except to support that domain.

But at genetic variability level, there is a global lack of repository. See Question 6 for the current issue on the Aquatic Genetic Resources. All of the sequences produced at the species (e.g. metazoa) or operational taxonomic unit level(e.g. archaea and bacteria) are currently stored in GENBANK, which does not support all those functionalities in order for the data to be available for other aggregators. Also, there is no information on the meta-data for their provenance.

As the metabarcoding and metagenomic data are being produced at big volumes, the above problems will be increasing. On the other hand, the new datasets will largely assist our trend analyses at large scales since these datasets are more easy and relatively less expensive than the "conventional" methods of sampling and identifying species.

- Occurrence / abundance data over time

Datasets are far from being complete enough to perform analyses on many species (see the statistics page of OBIS). The analysis of the OBIS data at the EU scale has been published by Vandepitte Et al. (2011).

A massive experiment had been attempted during the MarBEF NoE, which resulted in a series of databases from various ecosystem components, such as meio- and macrofauna. The main results of the analyses of these databases have been published in a Special Issue by Somerfield et al. (2009).

- Current Red List status of the species

The status of threat of marine species is globally assessed by IUCN under the Global Marine Species Assessment programme. The number of species assessed increases regularly. However, besides the vertebrates, the number of groups to be assessed is still high.

In addition, country or regional assessments (under the IUCN hat) are performed. The situation for users may be confusing as a species can be regionally endangered, while be of no concern globally. An example is the fish *Sciaena umbra* in the general Red List database: No result is given because the species has only been regionally assessed at the Mediterranean Sea level and is assessed as threatened on a national level (in Turkey).

- Data on major threats to European species

There is a good knowledge of potential threats to species (in particular, exploitation for food, insidious pollutions, oil spills, etc.), however there are new threats (e.g., plastic microparticules) that should be the object of detailed studies.

Groups with uncertain taxonomy are also less studied from that point of view.

2. Can we assess the status and trends of [European] ecosystems and ecosystem services?

Like in many cases, there is enough data to demonstrate in general/theory which are the general status, trends and threats of/to ecosystems and their mechanism.

But detailed data are usually missing at local scales to assess the ecosystem and their trends, or at least the density in space and time is so low that even with performing statistical tool, the range of results is too wide or uncertain to give clear indications to policy makers (e.g. Beaumont et al. 2007, Galparsoro et al. 2014).

The only way then is to apply the precautionary principle based on the few examples, studied in details. However, due to the complexity of the studied systems, each particular case may or could require adaptations of the solutions for better and sustainable exploitation.

One exception to all the questions below may be the fishery/aquaculture domain. However, raw data are usually protected and not disseminated, or is only available as fuzzed and/or aggregated.

A. Lists of species and the ecosystem services they perform

There are only a few studies. MARBEF Theme 3 has summarized the list of services that could be expected from marine ecosystems (Heip et al. 2009).

B. Occurrence data for relevant ecosystem services

A very few studies have mixed socio-economics and biodiversity data.

C. Comparable geo-referenced occurrence (abundance) data over time

Except for fisheries, no other data, or long-term series are linked to ecosystem services.

D. Can we infer ecosystems from occurrence data or do we need independent data on ecosystems and their composition under ideal / natural conditions?

Ecosystems are by definition the interactions between the living organisms and the communities they form with their environment. It is not really safe to infer ecosystems from the species occurrences data because many of the key-player species in a certain ecosystem can also be present in another one with a slightly different role. Also, the mechanisms by which the species come together and form communities may be different in the different ecosystems or even in habitats within ecosystems. Therefore, the safer way is to map the ecosystems and ground-truth the findings of the mapping techniques (multiple site-scan, etc.).

Hence a logical answer is that we need data on both.

E. Does sufficient taxonomic data exists (regarding number of species / species names, estimation of number of dark taxa etc.)

There are still gaps of taxonomic knowledge for a number of groups (e.g., Nematoda), usually those with species of small size, living in cryptic habitats, and requiring some technology for taxonomic analyses. These gaps are under constant analysis by WoRMS, PESI and especially under ERMS for Europe.

The finer the scale the patchier the faunistic and floristic lists become. The complete list of species in the Mediterranean, for instance, may be well known, but at the scale of a bay, a lagoon or a small island, there are too few validated lists, and regular monitoring.

So the trend at sea scale are relatively well known, but not at fine scale.

F. Data on major threats to ecosystem functioning in Europe

In general, there is enough data to demonstrate in general which are the threats to ecosystem and their mechanism.

But locally, there is generally not enough data to demonstrate that a particular threat is running in a given location (except the places where the threat has been study and that were used as an example for general conclusion).

In essence, the type of threat and their mechanism are well known, but data to demonstrate joint effect of two or more threats are lacking. as usually threats were studied separately due to the complexity of systems. Non-linear cumulative effects are almost unknown.

3. Are we closing the biodiversity knowledge gap (poorly known organisms, ecosystem services, areas)?

In general, the answer is negative. There are few exceptions localized in time and space such as fish and fisheries datasets, the North Sea ecosystems, the Channel ecosystems, the ATBI and LTER sites established in Europe.

A. Trends in accumulation of occurrence data (of different quality) over time with respect to taxonomic groups, geographic areas, ecosystem services, genetical information, etc.

B. Lists of species and the ecosystem services they provide if analyzing ecosystem services knowledge gaps

The number of such studies in the marine environment is very low, partly because the field is relatively new.

An exception could be the Ecosystem Approach to Fisheries, which actually encompasses socio-economic studies especially for small-scale fisheries (in coral reef ecosystems in developing countries, see WorldFish). FAO is currently engaging in many efforts to promote that approach and the iMarine infrastructure developed under the eponym FP7 European project is being developed with socio-economic oriented Virtual Research Environments.

Aquaculture is obviously another domain where data exist. Although it seems at the edge of biodiversity studies, marine farming has deep impact on ecosystems and their exploitation.

However, a few attempts on mapping and assessing ecosystem goods and services have only recently occurred in the literature (e.g. Salomidi et al. 2012, Galparsoro et al. 2014)

C. Improvement of the quality of occurrence data (removal of duplicates, validation etc).

OBIS has engaged in such efforts, and basically the marine biology community relies on these efforts. Several cleaning packages exist, but are not well advertised enough or are too focused on one group or one region.

GBIF has also developed a tool for cleaning data in the context of the ViBRANT (e-Infrastructures) project, the data quality improvement module (GBIF), a report of which can be found at: http://www.slideshare.net/DavidRemsen/tdwg-1remsen

There are individual researchers that have taken the work to clean data. However, they fail to report back to original providers, and international aggregators. One technical reason is the lack of Global Unique Identifiers for point data records. For two years, GBIF has tried to use its own identifiers as a potential solution, but it is not yet advertised and still under experimentation.

Progress were made during the EC FP& pro-iBiosphere, but what remains to be done is the actual implementation of possible solutions, and/or for GBIF to publish its solution.

D. Improved quality of the taxonomic information (building a global registry of species names, compatibility problems between CoL and GBIF classifications)

This issue is being addressed by the Global Name Architecture where GBIF and CoL are engaged, and by WoRMS for the marine environment. Pieces of the puzzle are there but the procedural links between them remain to be established.

4. Are we filling the gaps in historical knowledge (in relation to available historical data in collections, literature and non-mobilized digital datasets) so we can evaluate long-term trends?

A. Trends in accumulation of historical occurrence data (of different quality) according to different time spans (long-term distribution data at least with the beginning of the 1980ies).

B. Estimates of the total amount of available historical data in collections, literature, and non-mobilized digital datasets

The two points are treated together.

Previous European projects like BioCISE, ENSHIN and BioCASE have addressed that issue (even if not necessarily focused on marine data). BioCASE holds a metadatabase. However, assembling metadata was a major effort from the project partners, with medium success. Analyzing collection content from a metadata point of view (how many specimens, geographic and period coverages, etc.) is time-consuming in the absence of a computerization strategy. Even with well-organized manuscript catalogues, collection curators estimate that they have little time to answer questionnaires, and update the information regularly. One of the informal conclusions of BioCASE was that assembling an updated metadatabase consumes too much time, effort, and/or funding, unless there is a strong incentive for curators to be involved. The same result was found by the BioFresh project for its metadatabase about datasets. But then as datasets were mostly computerized, the project partners could cope with the absence of responses from colleagues by elaborating a minimum metadata by themselves to be just checked by data owners.

So all analyses that could be done would have a very wide uncertainty range. See the recent analysis of EurOBIS (2011).

Many historical data remain to be digitized, not only from the (main) museum collections, but from the literature, including from the narrative part of expeditions like done by Jackson et al (2011), Palomares et al (2006), Holm et al (2010), and other potential sources.

The primary potential source of data remains the museum collections where a massive digitization effort remains to be implemented (in collaboration with CETAF), not only encoding the manuscript catalogues, but also allocating geo-coordinates to locality name with a precision marker. Better tools and procedures, involvement of students and professors, recognition of the work, mobilization of citizen science is needed. Particular effort must be done on geo-referencing the types. Combined with the barcoding results, it facilitates taxonomic and nomenclatural decisions.

Involving citizen science could be a means to massively digitize such data, but proper incentive (both for citizen and piloting specialists), tools, organization, protocols/procedures, and quality control remain to be elaborated.

The second source is the gray literature stored in marine stations. Thousands of thesis and technical reports hold data. Even if the results were later published in the international literature, raw data were rarely included in low page number papers. The librarians of marine stations and their organizations (e.g., IAMSLIC) should be mobilized to build catalogues of gray literature with a proper indexation about raw data.

Semi-automatic text mining, the third source, is in its infancy but together with the Biodiversity Heritage Library, it may yield a large number of data, including that for traits. However, it is not expected that the density in space and time per species will be that high. Very often, one experimentation is conducted in one narrow location in a restricted time period. Note that data mining could be used to detect datasets, not only to flag potential data themselves.

A fourth source could be long-term data series, but they are actually very few in the world and in Europe (e.g., the zooplankton analysis of points A and B in the Bay of Villefranche-sur-Mer, France, Mediterranean).

A fifth source are the landing statistics for fisheries but only for a limited number of species. Moreover, often data are aggregated, and the exact origin of the catch is fuzzed or unknown. It is important to point out that official fishery statistics are “landing” statistics, not “catch” statistics. They do not account for by-catch, and obviously for unreported and illegal catches. However the Sea Around Us project, and now FAO, are developing methodologies to reconstruct catch statistics since the 1950s. There is a current debate to know if landing/catch statistics can be a measure of biodiversity, at least biomass abundance and species richness. While recent trawling surveys are difficult to get, the older ones are less protected, although raw datasets may be lost or, at least, are far from being digitized.

A sixth source could be the analysis of old pictures and stories accumulated by people as a citizen science activity. This requires specialists with good knowledge of local faunas and flora. Crowd sourcing digitization results in massive high quality data when combined with sample collection activities. Lastly, a carefully designed and applied rewarding system that provides accreditation to the citizen scientists is instrumental for the success of the citizen scientists action.

More than specific gaps as listed above, we have tried to summarize them under a general Gap/Recommendation section below.

5. Can we identify trends in the spread and effects of alien and invasive species [in Europe]?

A. Data on traits (ecological, life-history, morphological etc) of species

Many efforts are done at global (GISIN: Global Invasive Species Information Network, http://www.gisin.org) and European levels (DAISIE: Delivering Alien Invasive Species Inventories for Europe, http://www.europe-aliens.org/; CIESM Atlas of Exotic Species in the Mediterranean (www.ciesm.org/online/atlas/). Some countries, such as Greece have designed and developed specific repositories (ELNAIS, EASIN).

These information systems take the information from or link towards other species databases (like FishBase for fishes). However, hampered by the luck of a constant state (or regional) funding, it is hard for these databases to be thoroughly updated, and interoperable.

B. High-resolution occurrence / abundance data over time

C. Occurrence / abundance data over time

The two points are treated together below.

Although more data than for native species are available on the average due to the potential economic impact of non-indigenous species, it is still difficult to get quality data as there is no running central monitoring and repository at European level. However, the situation is better for the Mediterranean with the CIESM atlas of exotic species, although repository and access to raw data are not properly set up.

Opportunistically, and sometimes due to the interest of one researcher, high density data may have been collected for a given species but data remain difficult to access. . High density data can also be collected and made available in those countries which are placed in the crossroads of the invasive species, when they enter a new region. Such an example is Israel in the Eastern Mediterranean Sea.

D. Data on major routes and vectors of penetration of alien species in Europe

In general, the potential pathways are fairly well known or even successfully inferred for most of the non-indigenous species: escapes from importation for aquaculture and public aquariums, and species unknowingly transported with them, ballast waters, and the Suez canal for the Mediterranean (Lessepsian migration). Natural and climate change induced distribution range extensions also occur (through Gibraltar Strait mainly from the Atlantic to the Mediterranean; from South to North in Northeast Atlantic even at mesopelagic depth).

These information are documented in the information systems mentioned above. Only recently have thorough reviews started to published on the issue (e.g. Katsanevakis et al. 2014a)

E. Data on most invaded ecosystems

The most invaded ecosystems in the European seas are waters around main commercial harbors (due to ballast waters and fouling issues), and the eastern basin of the Mediterranean (due to Lessepsian migrants). The amount of data available for the former is highly variable depending on the proximity of marine stations and universities, and the political will to monitor the non-indigenous species. For the latter, see above the CIESM atlas of exotic species. However, massive substitutions of the seagrasses have also been documented in the entire basin of the Mediterranean Sea (see next paragraph).

F. Data on the ecological and economic impact of alien species to European ecosystems

Data are available when the socio-economic impact is/was high, especially on fisheries (e.g., Mnemopsis blooms in the Black Sea). The case of the impact of *Caulerpa taxifolia* on *Posidonia oceanica* seagrass bed ecosystem was also exhaustively documented. However, the impact of non-indigenous species over (yet) non-important ones is scarcely studied (e.g., the impact of Lessespsian Siganidae grazing over the *Cystoseira* bush).

Data are not gathered and are scarcely available. A recent and comprehensive review on the issue has been published by Katsanevakis et al. (2014b), in which both positive and negative impacts are analyzed, based on literature resources.

6. Can we assess the effect of [European] marine protected areas on the conservation of biological diversity?

A. High-resolution occurrence data over time (monitoring data) for protected areas and control areas.

The situation is highly variable across all MPAs in Europe. Some like Port-Cros in France and Zakynthos in Greece have been monitored for a long time, in particular for targeted species such as the dusky grouper *Epinephelus marginatus* (studies conducted with the help of the NGO Groupe d'Étude du Mérou in France, www.gemlemerou.org). But as noted above no exhaustive survey across all benthic groups exists, and the density of survey is at most annual but more often at several years interval.

Regionally, data are sufficient as the positive impact of closed MPAs over targeted populations within the area and also just outside could be demonstrated (e.g., Cerbère\Banyuls-sur-Mer).

Locally however, data may be lacking for proper decisions.

As a reminder (see section on Gaps on a Ecosystem level: Species in European marine protected areas above), about 10,500 marine fish species in the world (over ca. 17,100 in total, 61%) have occurrence data in at least one MPA. Data are still too scarce for the majority of invertebrates, but a detailed analysis by group should be done to precisely assess the situation.

At the European level, a dashboard should be established to follow up the progress (Marbef, Euromarine, JRC by extending their information system DOPA to marine areas, EEA) in partnership with EurOBIS.

B. Data on biodiversity changes (see above)

When monitoring exists, changes in flora and fauna of MPAs can be analyzed over time in correlation with changes in environmental parameters. Even if the environmental parameters are not measured at the local level, global datasets with proper estimation algorithm may provide good proxies for the recent decades.

However, as noted above, the density of occurrence data may not be enough locally for the majority of species to enable drawing reliable conclusions.

Gaps Identified and Recommendations

	Gaps	Recommendations
1	Oceanic deep sea data and information is not well synthesized in secondary literature. Their presence in country EEZ although evidenced by point data is not well documented in synthetic checklists made available to policy makers.	Knowledge about oceanic deep seas should be reported in integrated syntheses and high level technical/management reports targeted to biodiversity management purposes.
2	There are few sea snake species with geo-referenced records in global aggregators.	The herpetological community dedicated to sea snakes must make a huge effort to computerize collections, records from the literature, and assign geo-coordinates to point data recorded only with a locality name.
3	Visual surveys like census data for marine mammals are not available for global aggregators.	OBIS must make efforts to attract potential providers of marine mammals point data.
4	General lack of biological and ecological species traits	Now that a proper ontology seems to be established, make a major effort on data encoding. GSDs in WoRMS and SeaLifeBase should be supported to encode data for many groups to be prioritized, as well as FishBase and PolyTraits to complete their efforts. Marine biology students should be involved as part of their training (but citizen science does not seem to be a good means for data encoding, but may be learned by natural history society members). Focus first on maximal size, size at first maturity, maximal weight, depth range, habitats.
5	General lack of point/occurrence data: historical and legacy data digitized and made available; long-term monitoring	A massive effort is to be done to encode existing data, and to develop long-term series and regular monitoring. Marine biology students, as part of their training, and citizen science must and could be efficiently mobilized.
6	Complete threat assessment for all marine species and a link between assessments at different scales	Support IUCN to complete the GMSA, and to link better with country and regional assessment.
7	Except for fisheries, an almost complete gap of data about marine ecosystem services	A major effort should be made to assess marine ecosystem services in Europe properly and at a medium scale (country, sub-basins).

8	Complete taxonomic lists	Support WoRMS but moreover, the taxonomic community that collaborates with WoRMS – the so-called taxonomic editors and the Global Species Databases that they maintain.
9	Lack of local faunistic and floristic checklists	Develop the citizen science approach, in particular with the sport and touristic diving domain. Monitoring of certain areas could be also made by successive generations of marine biology students.
10	Lack of local data for analyzing local threat	Facilitate the development of Small and Medium Enterprises that could apply locally protocols developed in general by the marine ecology research domain.
11	Lack of Global Unique Identifiers for point data records	The Biodiversity Informatics community needs urgently to solve this long-standing issue now.
12	Implement the Global Name Architecture procedures between several components	GBIF should be the institution responsible for this implementation.
13	Lack of monitoring of digitization efforts	The BioCASE metadatabase about European collections could be revived (with proper marker of life zones), but only if there is a strong incentive for institutions and curators to respond to questionnaires – or to elaborate mechanisms for automatic metadata production and harvesting.
14	Lack of proper work organization in Europe to achieve the digitization of historical biodiversity data, primary point and occurrence data	There has been already a number of European organizations/ initiatives/projects that could work together to organize and monitor the historical data digitization: Euromarine, EurOBIS, CETAF, Marbef, EEA, JRC, LifeWatch, BHL Europe, Seadatanet, EMOdNet, etc. A complete strategy must be elaborated between all of them, and be implemented by staff 100% dedicated to one given task (e,g,. gray literature; collections; marine expeditions; old literature; …). Such an initiative has been undertaken by the recently started project SYNTHESYS3 (EU 7th FP). To boost digitization in many places, there could be task forces, e.g., by taxonomic group, visiting institutions to organize the digitization and train staff locally.

15	Lack of regular updates of and inconsistencies between non-indigenous information systems	Organize and support financially the regular update of non-indigenous information systems, in particular from the literature. From a European point of view, it must be made clear what the information system of reference is: DAISIE, the existing national information systems, the global system GISIN, or something else? Whichever is chosen, there is a need to develop mechanism for exchanging information in real time (alerts at minimum) both at procedures/protocols and technology points of view.
16	Lack of data compilation about the socio-economic impacts of non-indigenous species in marine life zone in Europe	Socio-economic data should be collected at the same time as biodiversity data, and vice versa. There is an important need for a mindset shift leading to more interdisciplinary projects. But also, each domain being epistemologically and academically independent, there is a need that each domain defines what would be the minimum dataset to be collected by the other domain when projects cannot be multi-disciplinary.
17	Apart from targeted and emblematic species, exhaustive local data are not available in many protected areas, from no data at all up to stored in electronic files but not shared.	MedPan and other MPA networks must be more effective in organizing a cost-effective common repository for their data to be shared through OBIS/GBIF.

4.4.4 Literature cited

Beaumont, N.J., Austen, M.C., Atkins, J.P., Burdon, D., Degraer, S., Dentinho, T.P., Derous, S., Holm, P., Horton, T., van Ierland, E., Marboe, A.H., Starkey, D.J., Townsend, M., Zarzycki, T. (2007) Identification, definition and quantification of goods and services provided by marine biodiversity: Implications for the ecosystem approach. *Marine Pollution Bulletin*, 54: 253-265

Bernard A.T.F., Götz A., Kerwath S.E., Wilke C.G. (2013) Observer bias and detection probability in underwater visual census of fish assemblages measured with independent double-observers. *Journal of Experimental Marine Biology and Ecology*, 443: 75-84, 10.1016/j.jembe.2013.02.

Bird T.J., Bates A.E., Lefcheck J.S., Hill N.A., Thomson R.J., Edgar G.J., Stuart-Smith R.D., Wotherspoon S., Krkosek M., Stuart-Smith J.F., Pecl G.T., Barrett N., Frusher S. (2014) Statistical solutions for error and bias in global citizen science datasets, *Biological Conservation*, 173:144-154. 10.1016/j.biocon.2013.07.037

Bodilis P., Louisy P., Draman M., Arceo H.O., Francour P. (2014) Can citizen science survey non-indigenous fish species in the eastern Mediterranean Sea? *Environmental Management*, 53:172–180.

Costello MJ, Coll M, Danovaro R, Halpin P, Ojaveer H, et al. (2010) A census of marine biodiversity knowledge, resources, and future challenges. PLoS ONE 5(8): e12110. doi:10.1371/journal.pone.0012110

Dickinson, J., Zuckerberg, B., Bonter, D. (2010) Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41: 149-172. 10.1146/annurev-ecolsys-102209-144636.

Edgar, G.J., Stuart-Smith, R.D. (2014) Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data*, 1[14007]: 1-8. 10.1038/sdata.2014.7

Froese, R., Pauly, D. (2000) FishBase 2000: concepts, design and data sources. ICLARM, Los Baños, Laguna, Philippines. 344 p.

Froese, R., Thorson, J. T. and Reyes, R. B. (2014), A Bayesian approach for estimating length-weight relationships in fishes. *Journal of Applied Ichthyology*, 30: 78–85. doi: 10.1111/jai.12299

Galparsoro, I., Borja, A., Uyarra, M.C. (2014) Mapping ecosystem services provided by benthic habitats in the European North Atlantic Ocean. *Frontiers in Marine Science* 1: 1-14.

Gledhill, D., Hobday, A., Welch, D., Sutton, S., Lansdell, M., Koopman, M., Jeloudev, A., Smith, A., Last, P. (in press) Collaborative approaches to accessing and utilising historical citizen science data: a case-study with spearfishers from eastern Australian. Marine & Freshwater Research [Prepubl. http://www.publish.csiro.au/view/journals/dsp_journals_pip_abstract_scholar1.cfm?nid=126&pip=MF14071].

Heip, C., Hummel, H., van Avesaath, P., Appeltans, W., Arvanitidis, C., Aspden, R., Austen, M., Boero, .F, Bouma, T.J., Boxshall, G., Buchholz, F., Crowe, T., Delaney, A., Deprez, T., Emblow, C., Feral, J.P., Gasol, J.M., Gooday, A., Harder, J., Ianora, A., Kraberg, A., Mackenzie, B., Ojaveer, H., Paterson, D., Rumohr, H., Schiedek, D., Sokolowski, A., Somerfield, P., Sousa-Pinto I., Vincx, M., Węsławski, J.M., Nash, R.

(2009) *Marine Biodiversity and Ecosystem Functioning*. Printbase, Dublin, Ireland, 100p.

Holm P, Marboe AH, Poulsen B, MacKenzie BR (2010) Marine Animal Populations: A New Look Back in Time. *Life in the World's Oceans*. Wiley-Blackwell. 1–24.

IUCN, UNEP-WCMC (2013) The World Database on Protected Areas (WDPA). November release. Cambridge (UK): UNEP World Conservation Monitoring Centre. URL: www.protectedplanet.net

Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A, Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B., Lenihan,, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J., Warner, R.R. (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science* 293:629-637. doi: 10.1126/science.1059199

Katsanevakis, S., Coll, M., Piroddi, C., Steenbeek, J., Ben Rais Lasram, F., Zenetos, A., Cardoso, A.C. (2014a) Invading the Mediterranean Sea: human-shaped biodiversity patterns. *Frontiers in Marine Science* (in press).

Katsanevakis, S., Wallentinus, I., Zenetos, A., Leppäkoski, E., Çinar, M.E., Oztürk, B., Grabowski, M., Golani, D., Cardoso, A.C. (2014b) Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. *Aquatic Invasions* (in press)

Miyazaki, Y., Murase, A., Shiina, M., Naoe, K., Nakashiro, R., Honda, J., Yamaide, J., Senou, H. (2014) Biological monitoring by citizens using Web-based photographic databases of fishes. *Biodiversity Conservation*, 23:2383–2391. 10.1007/s10531-014-0724-4.

Palomares, M.L.D., Mohammed, E., Pauly, D. (2006) European expeditions as a source of historic abundance data on marine organisms: a case study of the Falkland Islands. *Environmental History* 11 (4): 835-846 doi:10.1093/envhis/11.4.835

Salomidi, M., Katsanevakis, S., Borja, A., Braeckman, U., Damalas, D., Galparsoro, I., Misfud, R., Mirto, S., Pascual, M., Pipitone, C., Rabaut, M., Todorova, V., Vassilopoulou, V., Vega Fernandez, T. (2012) Assessment of goods and services, vulnerability, and conservation status of European seabed biotopes: a stepping stone towards ecosystem-based marine spatial management. *Mediterranean Marine Science* 13: 49-88.

Somerfield, P., Vanden Berghe, E., Arvanitidis, C. (2009) Large-scale studies of the European benthos: the MacroBen database. *Marine Ecology Progress Series* Vol. 382: 221–224 (doi: 10.3354/meps08045).

Stelzer, K., Heyer, K., Bourlat, S., Obst, M. (2013) Application of Ecological Niche Modelling and Earth Observation for the risk assessment and monitoring of Invasive species in the Blatic Sea. MarCoast II - Marine and Coastal Environmental Information Services Ballast Water Option; Report, 57p.

Vandepitte, L., Hernandez, F., Claus, S., Vanhoorne, B., De Hauwere, N., Deneudt, K., Appeltans, W., Mees, J. (2011) Analysing the content of the European Ocean Biogeographic Information System (EurOBIS): available data, limitations, prospects and a look at the future. *Hydrobiologia, 667*: 1-14 (dx.doi.org/10.1007/s10750-011-0656-x).

4.5 FOCUSED-REVIEW OF GAPS IN SPECIFIC DATABASES - MARINE AND COASTAL DATA HOLDINGS OF UNEP-WCMC

4.5.1 Taxonomic groups/realm

Marine and coastal data holdings

4.5.2 Data sources analyzed

The "data source" analyzed here are the **marine and coastal data holdings of the UNEP World Conservation Monitoring Centre** (UNEP-WCMC), based in Cambridge (UK). UNEP-WCMC is the specialist biodiversity assessment arm of the 'United Nations Environment Programme'. UNEP-WCMC's mission is to provide authoritative information about biodiversity and ecosystem services in a way that is useful to decision-makers who are driving change in environment and development policy. In this context, UNEP-WCMC curates and/or distributes a number of global spatial datasets of biodiversity importance.

4.5.3 Results

Coverage of the data source

UNEP-WCMC curates and/or distributes 31 marine and coastal datasets, which fall into eight categories: *biogenic habitat*, *species habitat*, *species distribution*, *biodiversity metric*, *area of biodiversity importance*, *biogeographic classification*, *ecological status and impact*, *administration* (Table 10). The datasets are global in geographic extent, and 22 are of relevance to European Seas. As some European Member States have territories located outside of "mainland Europe" (e.g. British and French Overseas Territories), also listed are datasets that are not directly relevant to European Seas, but might be relevant to these territories.

Spatial data are held in a variety of formats, primarily vector (polygon, polyline, point, grid), but also raster (e.g. geotiff). Data originate from various sources: some are collation of national/regional subsets (e.g. *Global Distribution of Seagrasses, 2005*), whilst others are model outputs (e.g. *Global Patterns of Marine Biodiversity, 2010*) or are derived from satellite imagery (e.g. *Global Distribution of Islands OSM, 2013*).

Outline of gaps and biases (e.g. spatial, taxonomic, temporal) and data quality

All the datasets curated and/or distributed by UNEP-WCMC have detailed, ISO 19115-compatible, metadata sheets that provide dataset-specific background information, including *citation*, *creation methodology*, *lineage*, *maintenance frequency*, and *quality, limitation(s) and fitness for use*. The purpose of these metadata sheets is to ensure optimum use of the data, in the light of their known biases and other known and potential weaknesses.

The *Global Distribution of Coral Reefs (2010)* is known, for instance, to contain (partly) overlapping polygons, meaning that a dissolve operation within a Geographic Information System software is needed before surface area calculations are carried out. This dataset moreover does not provide spatial information on the distribution of individual coral reef species: the dataset is limited to providing information on the 'presence of reef', and it should not be assumed that coral reefs are "absent" elsewhere. Other datasets, such as the *Global Distribution of Cold-water Corals (2005)*, show spatial biases (in this case a high density of reefs in the North Atlantic Ocean), as a result of survey efforts by data contributors. A couple of datasets, such as the *Global Distribution of Seagrasses (2005)* and the *World Database on Protected Areas (2014)*, have both point and polygon subsets, which need to be used in combination, keeping in mind that the point subsets do not necessarily have associated surface areas (particularly relevant when the aim is to calculate surface areas). Some datasets such as the *Global Distribution of Sea Turtle Nesting Sites* (1999) and *of Feeding sites* (1999) are no longer maintained and must hence be used with caution.

4.5.4 Data accessibility

A number of datasets part of UNEP-WCMC's data holdings can be viewed and/or downloaded from the Ocean Data Viewer (UNEP World Conservation Monitoring Centre 2014) at http://data.unep-wcmc.org. In parallel, Web Map Services for most of these datasets can be accessed on ArcGIS Online (http://wcmc.io/58c2), for use in web-mapping applications such as SeaSketch.org. Other datasets are kept on file and interested users are encouraged to contact UNEP-WCMC (marine@unep-wcmc.org) to access them. Table 10 provides details on dataset-specific access, and interested users should refer to available individual metadata sheets for use restrictions, especially if there is a business or commercial element to the work undertaken.

4.5.5 Trends in accumulation of occurrence data / integration of historical data

UNEP-WCMC does not monitor trends in the accumulation of occurrence data in its data holdings. For some datasets (e.g. *Global Distribution of Seagrasses, 2005*; *Global Distribution of Saltmarsh, 2014*), this may be inferred from the information contained in the dataset's attribute table (if published data sources are listed).

4.5.6 General recommendations and prioritization for closing the gaps

UNEP-WCMC is undertaking work to address known issues and gaps in the datasets that it curates and/or distributes. The *World Database on Protected Areas (2014)* is continually being updated (monthly releases) based on submissions from national governments. Supplementary occurrence data are being sought from national- and regional-level organisations to fill spatial gaps in three biogenic habitat datasets (e.g. *Global Distribution of Seagrasses*, *Global Distribution of Saltmarsh*, *Global Distribution of Cold-water Corals*). UNEP-WCMC is currently investigating the possibility of validating remaining un-validated portions of the *Global Distribution of Coral Reefs (2010)*, so as to create a trustworthy dataset that can be used as baseline for investigating changes due to climate and other human-induced impacts.

UNEP-WCMC has also created an online validation tool so as to make use of citizen-science to improve selected spatial datasets (http://validation.unep-wcmc.org). Using this tool, users can validate and/or edit the boundaries of coral reefs, based on local knowledge or the underlying satellite imagery.

Finally, UNEP-WCMC recently collaborated with Dr. K. Kaschner (AquaMaps; Albert-Ludwigs-University of Freiburg, Germany) so as to raise awareness of the usefulness of 'species distribution modelling' in filling in spatial gaps in our knowledge of where species are more or less likely to be found. The known and probable distributions of 10 marine mammal species were modelled using the AquaMaps approach (Kaschner et al. 2014). Expert-reviews of each map can be found in Annex 3 of Martin et al. (2014).

4.5.7 Literature cited

Kaschner, K., Rius-Barile, J., Kesner-Reyes, K., Garilao, C., Kullander, S.O., Rees, T., Froese, R. (2014) AquaMaps: Predicted range maps for aquatic species. Version 08/2013.

Martin, C.S., Fletcher, R., Jones, M.C., Kaschner, K., Sullivan, E., Tittensor, D., Mcowen, C., Geffert, J., Bochove, J. van, Thomas, H., Blyth, S., Ravillious, C., Tolley, M., Stanwell-Smith, D. (2014) Manual of marine and coastal datasets of biodiversity importance. May 2014 release. http://wcmc.io/MarineDataManual. UNEP World Conservation Monitoring Centre, Cambridge (UK)

UNEP World Conservation Monitoring Centre (2014) Ocean Data Viewer, http://data.unep-wcmc.org.

Table 10. Marine and coastal data holdings of the UNEP World Conservation Monitoring Centre (UNEP-WCMC). Relevance of the various datasets to European Seas is indicated. Coloured shading is used to indicate that:

- the dataset can be viewed and/or downloaded from UNEP-WCMC's *Ocean Data Viewer* (ODV; http://data.unep-wcmc.org) and related *Data Download* page (DDP; http://datadownload.unep-wcmc.org/datasets),
- information about dataset access can be sought from UNEP-WCMC (at marine@unep-wcmc.org).
-

Category	Dataset title	Contact organisation	ID	Metadata	Data access	EU seas
Biogenic habitat	Global Distribution of Coral Reefs (2010)	UNEP-WCMC	WCMC-008	y	ODV	
	Global Distribution of Coral Reefs - 1 Km Data (2003)	UNEP-WCMC	WCMC-009	y	ODV	
	Global Distribution of Cold-water Corals (2005)	UNEP-WCMC	WCMC-001	y	ODV	y
	Global Distribution of Mangroves USGS (2011)	UNEP-WCMC	WCMC-010	y	ODV	
	World Atlas of Mangroves (2010)	UNEP-WCMC	WCMC-011	y	ODV	
	Global Distribution of Mangroves (1997)	UNEP-WCMC	WCMC-012	y	ODV	
	Global Distribution of Seagrasses (2005)	UNEP-WCMC	WCMC-013-014	y	ODV	y
	Global Distribution of Saltmarsh (2014)	UNEP-WCMC	WCMC-027	y	Contact UNEP-WCMC	y
Species habitat	Global Distribution of Marine Turtle Nesting Sites (1999)	UNEP-WCMC	WCMC-007	y	DDP	y
	Global Distribution of Marine Turtle Feeding Sites (1999)	UNEP-WCMC	WCMC-006	y	DDP	y
Species distribution	Global Distribution of Northern Fur Seals (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-001	y	Contact UNEP-WCMC	
	Global Distribution of Hawaiian Monk Seals (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-002	y	Contact UNEP-WCMC	
	Global Distribution of Grey Seals (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-003	y	Contact UNEP-WCMC	y
	Global Distribution of Hector's Dolphins (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-004	y	Contact UNEP-WCMC	
	Global Distribution of Northern Bottlenose Whales (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-005	y	Contact UNEP-WCMC	y
	Global Distribution of Sperm Whales (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-006	y	Contact UNEP-WCMC	y
	Global Distribution of	Albert-Ludwigs-	Kaschner-	y	Contact	y

Category	Dataset title	Contact organisation	ID	Metadata	Data access	EU seas
	Bowhead Whales (2013)	University of Freiburg	008		UNEP-WCMC	
	Global Distribution of Sei Whales (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-009	y	Contact UNEP-WCMC	y
	Global Distribution of Atlantic Spotted Dolphins (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-011	y	Contact UNEP-WCMC	y
	Global Distribution of Melon-Headed Whales (2013)	Albert-Ludwigs-University of Freiburg	Kaschner-012	y	Contact UNEP-WCMC	
Biodiversity metric	Global Patterns of Marine Biodiversity (2010)	UNEP-WCMC	WCMC-019	y	ODV	y
	Global Map of Shannon's Index of Biodiversity (2014)	Ocean Biogeographic Information System, Intergovernmental Oceanographic Commission (UNESCO)	OBIS-001	y	ODV	y
	Global Map of Hurlbert's Index of Biodiversity (2014)	Ocean Biogeographic Information System, Intergovernmental Oceanographic Commission (UNESCO)	OBIS-002	y	ODV	y
	Global Seagrass Species Richness (2003)	UNEP-WCMC	WCMC-015	y	ODV	y
	Global Marine Turtle Species Richness (2002)	UNEP-WCMC	WCMC-003		Contact UNEP-WCMC	
Area of biodiversity importance	World Database on Protected Areas (2014)	UNEP-WCMC	WCMC-016	y	Protected Planet[17]	y
Biogeographic classification	Marine Ecoregions of the World (2007)	UNEP-WCMC	WCMC-017	y	ODV	y
	Pelagic Provinces of the World (2012)	UNEP-WCMC	WCMC-018	y	ODV	y
Ecological status and impact	SeagrassNet: Global Seagrass Monitoring Network (2013)	Washington State Department of Natural Resources, Aquatic Resources Division	WaDNR-001	y	Contact UNEP-WCMC	y
Administration	Global Distribution of Islands IPBoW (2010)	UNEP-WCMC	WCMC-005	y	Contact UNEP-WCMC	y
	Global Distribution of Islands OSM (2013)	UNEP-WCMC	WCMC-031	y	Contact UNEP-WCMC	y

[17] www.protectedplanet.net

4.6 AVAILABILITY OF FRESHWATER BIODIVERSITY DATA

4.6.1 Introduction

Freshwater ecosystems face the highest species extinction rates, yet freshwater biodiversity data availability is very poor. While only giving a rough indication of data availability, a simple keyword search on the Global Biodiversity Information Facility portal yielded 45 freshwater versus 315 marine datasets[18]. Similarly, the 8.9 million freshwater occurrence records we harvested for the data portal in December 2011 only represent 2% of all records available through the GBIF network at that time.

The realisation that data to improve our understanding of distribution patterns of freshwater organisms are largely unavailable, was a key issue that the EU FP7 BioFresh project (Biodiversity of Freshwater Ecosystems: Status, Trends, Pressures, and Conservation Priorities) wanted to address. During the project (November 2009-April 2014), BioFresh constructed a central freshwater biodiversity information and data platform (http://www.freshwaterbiodiversity.eu/), which aims to improve the discoverability of freshwater biodiversity resources and make them publicly available.

This section includes a short overview of the gap analysis performed during BioFresh and experience gained during the project and follow-up activities. The recommendations focus on those relevant in the context of this EU BON report.

4.6.2 Short summary of the gap analysis conducted in the framework of the BioFresh project

The gap analysis on freshwater biodiversity data conducted in the framework of the BioFresh project focussed specifically on the needs of the scientists involved in the project. The analyses performed in the project ranged from contemporary species distribution modelling on global, European, catchment and point locality scale to future scenarios under climate change and a wide range of environmental stressors.

Gaps for scientists. Ten out of 16 requests for freshwater biodiversity data are related to (expert curated) distribution ranges or require the translation and expert curation of point data for generating a complete picture of a species range. Six requests indicated the need for point data originating from complete surveys, thus providing information on the presence/absence of a specific species on a sampling locality. Three of these requests had the strict requirement for environmental data collected during sampling.

Gaps for policy makers. At a later stage in the project, we also conducted a gap analysis specifically targeting the data and information needs expressed by policy makers. To capture their input, we both contducted a survey and summarised the discussions held during the Water Lives symposium in January 2014. With regards to the needs of scientists vs. policy makers, it is important to highlight that the former are looking for primary data or raw data products, while the latter require polished information and knowledge products. In terms of outcomes, the BioFresh data portal (http://data.freshwaterbiodiversity.eu/) primarily targets scientists, while the Global

[18] (http://www.gbif.org/dataset/search?q=freshwater&type=OCCURRENCE vs. http://www.gbif.org/dataset/search?q=marine&TYPE=OCCURRENCE - 06/06/2014).

Freshwater Biodiversity Atlas (http://atlas.freshwaterbiodiversity.eu/) envisages a much wide audience including policy makers.

The survey respondents and discussion participants highlighted on one hand the need for establishing socio-economic justification for biodiversity conservation, while on the other hand valuing the link between biodiversity and ethics. They also stressed the need for providing solutions and indicated the preference of being presented with contrasting scenarios reflecting a range of policy decisions rather than having to interpret model uncertainty per se. While policy makers require clear and concise messages, they clearly appreciated face-to-face interactions with scientists (as organised during the Water Lives symposium) and encouraged investing in strong scientific advocacy. Finally, the value of more elaborate teaching materials targeting specific interest groups (e.g. freshwater ecology for dam engineers, water managers) was stressed.

4.6.3 Information on freshwater datasets and their availability

As mentioned earlier, the BioFresh project started out of the realisation of poor data availability of freshwater biodiversity data, which was confirmed during a range of analyses executed in the course of the project.

An important activity to document the existence and availability of freshwater biodiversity related datasets, is to gather such information in a (meta)database. Such an effort was initiated with the construction of the BioFresh metadatabase. At present, this metadatabase (http://data.freshwaterbiodiversity.eu/metadb/) contains 251 datasets, contributed and/or validated by 154 individual users. Freshwater datasets relevant in the context of EU BON will be incorporated in this metadatabase and further advertised through this network.

4.6.4 Freshwater occurrence data

By building a network and web-infrastructure for publishing and centralising global freshwater biodiversity data, the BioFresh data portal was conceived as a thematic node of the GBIF network.

In addition to these technical developments, BioFresh invested considerably in active data mobilisation. At least 2 million occurrence data records were mobilised during the project, with the majority of the datasets still being processed for public release through the GBIF network. Major datasets which may be relevant in the framework of EU BON include a systematic effort to complete the gap in data availability for fish occurrences for several European countries (9 digitisation projects, covering 13 countries), and the compilation of distribution data for European caddisflies (Trichoptera) and stoneflies (Plecoptera). For caddisflies for example, a comprehensive dataset on their distribution was previously unavailable. In total, 66 contributors made almost 600.000 new occurrences available.

Needless to say, the amount of available data (and metadata) is still extremely limited. Several BioFresh partners have committed resources to continue the efforts to integrate and mobilise freshwater data, but further attention for freshwater data mobilisation from different actors in the field is highly needed (e.g. through supporting GBIF nodes and thematic initiatives such as BioFresh). In addition, we believe that a more systematic approach in publishing biodiversity data is needed from the institutes and organisations involved in (freshwater) biodiversity monitoring. Approaching these parties and setting up an exchange of expertise and guiding them

to set up a data publishing workflow could be an action to be initiated under the umbrella of EU BON, e.g. after carrying out some initial case studies in collaboration with the test sites.

4.6.5 Taxonomic checklist on freshwater species

At present, the Freshwater Animal Diversity Assessment (http://fada.biodiversity.be/) database contains species names for roughly one third of the estimated 150.000 freshwater species. Through consultations with taxonomic experts during the BioFresh project, the species checklists, which are used as a taxonomic backbone for the BioFresh data portal, could be extended with over 30,000 new names, particularly for macro-invertebrate species. Further extension of this database is on-going through a collaborative project with the World Register of Marine Species (WoRMS) and through collaboration with the EU BON partners involved in the harmonisation of the European taxonomic backbone. Nevertheless, this freshwater specific taxonomic database clearly has some catching up to do with initiatives such as WoRMS in order to reach a reasonable level of completeness.

4.6.6 Freshwater trait data

So far, we have not conducted a systematic inventory or gap analysis of trait data specific for freshwater organisms. The following list reflects the main databases we are aware of;

- freshwaterecology.info provides information on autecological characteristics, ecological preferences and biological traits as well as distribution patterns of more than 12.000 European freshwater organisms belonging to fish, macro-invertebrates, macrophytes, diatoms and phytoplankton.
- As elaborated in the chapter on marine species (4.4), FishBase (http://www.fishbase.org/) forms an indispensable source of information on fish species, also in the freshwater realm.
- Various datasets include information on traits for specific organism groups, e.g. breeding info for birds in the atlas website from the European Bird Census Council (http://www.ebcc.info) and depth and occurrence data for Amphibia on Amphibiaweb (http://amphibiaweb.org) but are to our knowledge not necessarily integrated in a central trait network or database.

4.6.7 Recommendations with regards to gaps for scientists

To a large extend, the experience gained during and the gap analysis performed within the BioFresh project, and follow up activities under EU BON, support a wide range of the recommendations provided throughout this report. The following overview focusses on the main recommendations which are relevant in the context of this report.

- There is a general need to actively encourage biodiversity data holders to make their data available. This can be achieved through;
    - the continued support to initiatives such as GBIF and its national and thematic nodes

    - elaborating an IPR policy on aggregated data and standardised embargo for EU and nationally funded research
    - require the submission of data associated with scientific publications both by the scientific journals (e.g. the Pensoft Biodiversity Data Journal), and the encouragement of (meta)data papers, as by researchers who publish in traditional journals who have not yet implemented such facility (cfr. sequence submission to GenBank)
    - support efforts for the digitisation of legacy literature and automating the mark-up of recent works.
- Scientists indicated a clear need for presence/absence data or abundance data. Making such information more readily available requires that the current standards and tools used for publishing primary biodiversity data, notably GBIF's Integrated Publishing Toolkit (IPT), are improved for dealing with data from sampling or monitoring campaigns. This recommendation aligns well with the on-going efforts to facilitate sharing of sample data, which are lead by GBIF under the umbrella of EU BON.
- The fact that scientists highlighted value of standardised sampling, and sampling different organism groups and environmental variables simultaneously, reflects the need to coordinate and standardise sampling efforts at various scales. Initiating such a concerted effort could be part of the activities of the EU BON testing sites.
- Scientists indicated the need for supporting spatial data products e.g. water temperature maps, stream flow modification. The exchange and generation of such products could be established through the collaboration with geographers and remote sensing experts. Biodiversity scientists should clearly express the data needs to these communities.
- Based on the need for data reflecting the complete distribution range of species, which underlines the indispensable value of the IUCN RedList assessments, we see both a requirement for
    - building a capacity to automate the translation of point data into catchment level information and facilitate expert, and
    - supporting species experts and coordinating bodies to carry out the validation work.
- Along the same lines, we observed a clear need for supporting expert databases, esp. those focussing on taxonomic checklists and species traits.

4.7 FOCUSED-REVIEW OF GAPS IN SPECIFIC DATABASES: GAP ANALYSIS ON POLLINATOR SPECIES (HYMENOPTERA: APOIDEA: ANTHOPHILA)

4.7.1 Introduction

This preliminary gap analysis shows the data availability for wild bee species and points out some gaps of current available biodiversity information of bee species on a European scale. As other studies show, the gaps and differences among common sources of distribution information could be quite substantial (e.g. Duputié et al 2014) and here we tested the data available for European bee species. The bee species were selected based upon occurrence records on a national scale of the Checklist of Western Palaearctic Bees (provided by M. Kuhlmann, NHM-London).

Bees are crucial for maintaining important ecosystem services, particularly for pollination services. A study estimates the yearly annual economic value for providing pollination services to be €153 billion for the year 2005 (Gallai et al., 2009) . The role of bees as pollinators became even more important for the agricultural production within the last decades, as the global agricultural area depending on pollination services expanded significantly. For example, the area of agricultural production depending on pollination services increased by over 300%. At the same time, the pollination-dependent agricultural production expanded from 3.6 % to 6.1% (in % of the whole agricultural production, from 1961-2006, Aizen and Harder, 2009).

Pollination from bees and other insects like hoverflies, butterflies or wasps, has not only a high economical but also a high ecological value. Pollination is crucial for wild flowering plants, and it is estimated that 88% of the angiosperms are pollinated by animals, (i.e. over 300 000 species, Ollerton et al., 2011). The percentage is particularly high in the tropical communities, whereas in the temperate zone, the percentage is slightly lower (94 respectively 78%).

There are many species involved in the pollination of agricultural and wild plants, and there is a high diversity in bee species. We find over 3,351 bee species in the Western Palaearctic realm. On an extended Pan-European scale (see below), there are around 2,546 bee species with known occurrences listed in our dataset.

Particularly the recent declines of pollinating species stress the need for high-quality datasets to detect such changes. Studies show that significant declines in pollinator abundance could be detected in some well surveyed countries, for example in regions in Britain and the Netherlands which was, at least in Britain, accompanied by a decline of plants reliant on insect pollinators (Biesmeijer et al., 2006). However, such conclusions are mostly based on the distribution and population estimates of *Apis mellifera* and global assessments are not taking into account the wild pollinators or feral honeybees (Aizen and Harder, 2009).

To include wild pollinator species in the current and future assessments, datasets are needed for also assessing their former and current distribution and their trends. In this preliminary gap analysis we evaluated current bee species distribution data on a Pan-European scale.

4.7.2 Methods

For evaluating the quality of data on pollinators, we used distribution information on European bee species from various sources (European Union data on Habitats Directive Article 17 reporting, Global Biodiversity Information Facility, Checklist datasets, Atlas of European Bees). We have chosen data from the data-mediating portal Global Biodiversity Information Facility (GBIF) for a first evaluation of available information on bee species. To compare the quality of data shared through GBIF, we also used the Checklist of Western Palaearctic Bees (hereafter called Checklist) provided by Dr. Michael Kuhlmann from the Natural History Museum in London. The checklist contains, as mentioned above, 3,351 bee species in total and 2,546 species with occurrence records in an extended Pan-European Context (including Turkey, countries of the Caucasus and others, see Annex B for a list of the countries).

For the comparison we analyzed occurrence records on a national level for 46 European countries, as the Checklist data are based on country presence/absence, taking into account the geographical borders of (Pan)-Europe. When using GBIF-mediated point-data, the occurrences were assigned to European countries. Point occurrences that were located within small regions and states, like Åland Island, Guernsey or San Marino, were assigned to the adjacent larger country/surrounding country. This allocation was needed, as the Checklist provides occurrence records only on a national scale and only for the main 46 countries. The Checklist data was validated by experts and was used as reference data for the comparison with GBIF-datasets. For a comparison regarding the number of specimen records/occurrence records the data from the expert database "Atlas of the European Bees" was used, a project where many experts are involved that also contributed to the Checklist data. Occurrences of the checklist data for species in Yugoslavia were only counted as number of occurrences in the region "Balkan Peninsula". Also, these occurrences were only used if there was not already a presence mapped in one of the recent Balkan states that are former Yugoslavian countries (like Serbia, Croatia etc.). Occurrence records of the GBIF-mediated data were obtained by an export of the database. However, as GBIF constantly integrates new datasets, there might be slight differences when comparing results of this current analysis to the most recent version of the datasets shared through GBIF.

For a comparison of available data sources on a European scale, we evaluated the country-specific occurrence records of 1245 European bee species, which are all species from the Checklist that could be matched with GBIF-records. For a first comparison of other available data sources on a European scale and in Denmark, we evaluated species of the Checklist that occur at least in 30 European Countries. That approach resulted in a selection of 84 bee species which all could be linked to accepted GBIF species names. Data from the Atlas of European Bees came from the web page of the Atlas (Rasmont & Iserbyt, 2014). For 59 of the sample species, data could be found on the web page, however two species were listed under another name.

4.7.3 Results

Datasets of the European Union (species listed in the Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora)

Currently, there are no bee species listed in the habitats directive, neither in the Annex II or the Annex IV species. For Annex II species (species requiring designation of Special Areas of Conservation) and Annex IV (species in need of strict protection) European member states have

to report their distribution and trends in species abundances. This means for the case of European bee species that no data is collected on a European scale, despite the fact that they are important e.g. for pollination in the wild and crops for agricultural production.

Comparison of GBIF-mediated data versus the Checklist of Western Palaearctic Bees

The evaluation of GBIF-data and the Checklist of Western Palearctic Bees shows that there are significant differences between the information on species distributions contained in both data sources. For the comparison, the 1245 bee species of the Checklist were used where at least one occurrence record in GBIF exists. For the other 1301 species, either no occurrence records exist or the name could not be found/matched in the GBIF database. The Checklist data could be used for evaluating gaps and limitations of GBIF-data as they are a validated by a whole set of European experts. The datasets shared in GBIF contain remarkable numbers of distribution data on the selected bee species and give the impression that plenty of data is already available. However, there are significant gaps in data, particularly for some specific countries and regions in Europe.

Gaps in GBIF- data on a country-level scale

Generally, for 48.9% of the species that occur in Pan-Europe occurrence records can be found in GBIF. This means that for 1'245 species occurrence records are available out of 2'546 species that are ocurring in Europe according to the Checklist data. So before analysing the differences between GBIF and Checklist data more specifically, it can be stated that there are quite obvious gaps, as currently for nearly half of the species no observation or specimens records exist.

The analysis of all 1245 selected bee species shows that there are no GBIF-data available for on average seven countries per species. There is a high variation in gaps among the different European countries. Some countries are not well covered by GBIF-data regarding the surveyed country-specific bee occurrences. Table 11 (and Fig. 30, see below) shows the ten countries with largest gaps in GBIF-data and the ten countries with smallest gaps.

Table 11: 10 countries with smallest gaps in GBIF occurrence information (green) and 10 countries with largest gaps in GBIF data. Gaps were detected by evaluating lacking species records in GBIF compared to Checklist occurrence records.

United Kingdom	1.	Armenia	1.
Ireland	2.	Belarus	2.
Sweden	3.	Azerbaijan	3.
Germany	4.	Georgia	4.
Finland	5.	Latvia	5.
Netherlands	6.	Liechtenstein	6.
Norway	7.	Albania	7.
Austria	8.	Montenegro	8.
Spain	9.	Malta	9.
Belgium	10.	Moldova	10.

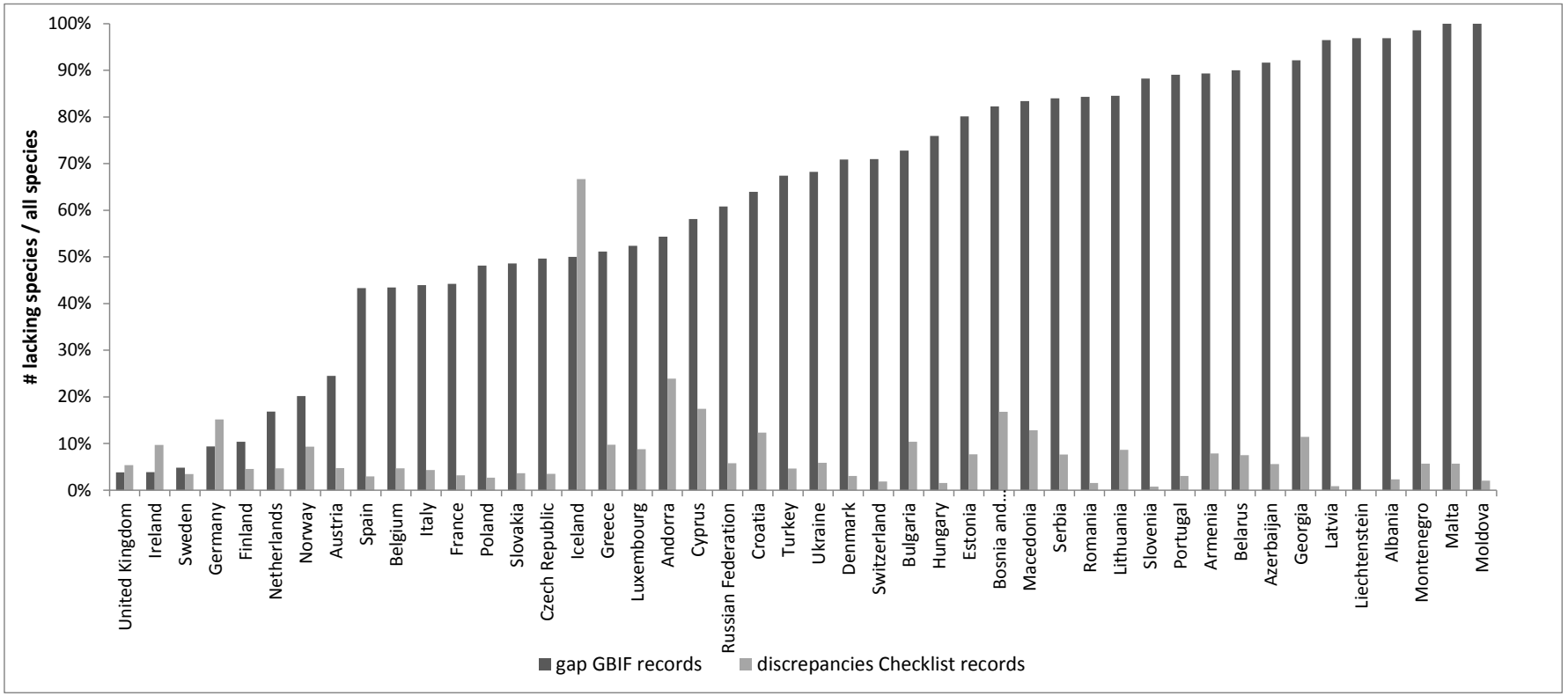


Fig. 30: GBIF gaps in the countries regarding bee species occurrence data. Gaps are determined by calculating the ratio between the number of lacking bee species occurrence records (dark grey: GBIF gaps) and the number of all bee species that are noted in the checklist to occur in a given country (baseline is the Checklist data for a given country for 1245 bee species with GBIF records ~ 48,9% of bee species in Europe). Also the discrepancies of checklist records are noted, i.e. species records of GBIF where no validated country occurrences are recorded yet (light grey: Checklist discrepancies).

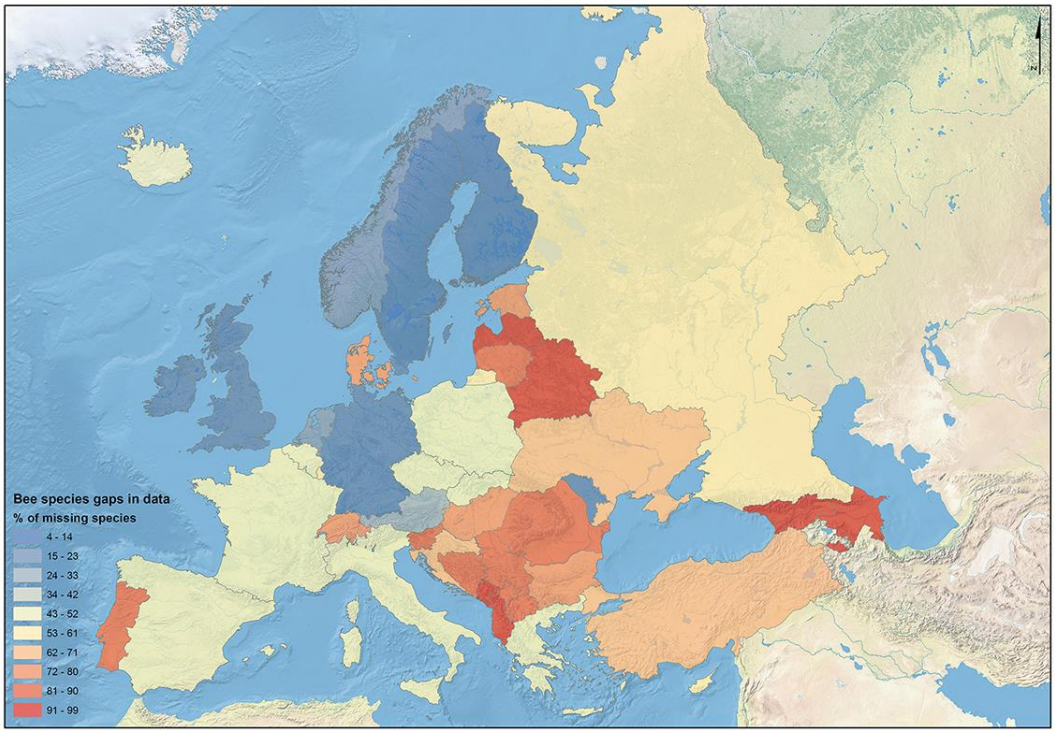


Fig. 31: Map visualizing the GBIF gaps in the countries regarding bee species occurrence data. Gaps are determined by calculating the ratio between the number of lacking bee species occurrence records (Blue: small gaps, red: large gaps). Not included are bee species where no occurrence records exist in GBIF.

Fig. 31 also shows the gaps of datasets shared through GBIF. However, GBIF contains species occurrence records for species where there is no validated occurrence for this particular species in the Checklist data yet.

As the figure shows, some countries like U.K., Ireland, Sweden, Germany and Finland are well covered by the GBIF-data, but there are large gaps in other countries like Albania, Montenegro and Moldova. Fig. 31 shows a map representation of the results regarding the gaps in GBIF-data. Please note that the figure shows only the gaps for the species that could be linked to GBIF data, there are in addition other gaps for species where no GBIF data exists but which are not part of this analysis. But GBIF also contains occurrence data of additional species where the Checklist indicates an absence of the species in the particular country. Such countries with additional GBIF-data are for example Germany, Ireland, Greece or Bosnia and Herzegovina (see Fig. 30). However, the additional country occurrences ("Checklist discrepancies", see Fig. 30) of bee species due to GBIF occurrence records need to be validated in each case by experts as there are many cases of misidentifications or nomenclatural problems that lead to these discrepancies. The GBIF-provided occurrence records could potentially be used for the specific expert datasets. However, as GBIF contains different kinds of datasets in terms of quality, such in-depth quality checks are urgently needed.

Also, when analyzing the data from a regional perspective, there are quite some significant differences. The highest number of species not covered in a region at all by GBIF-data is **Eastern Europe and the Caucasus.** Here, 68% of the species are not covered with occurrence data shared in GBIF. Other high ranking regions regarding GBIF lacunae are the **Balkan Peninsula**. More records in GBIF datasets are available in **Central Europe; least gaps** are in **Western Europe** and **Scandinavia**.

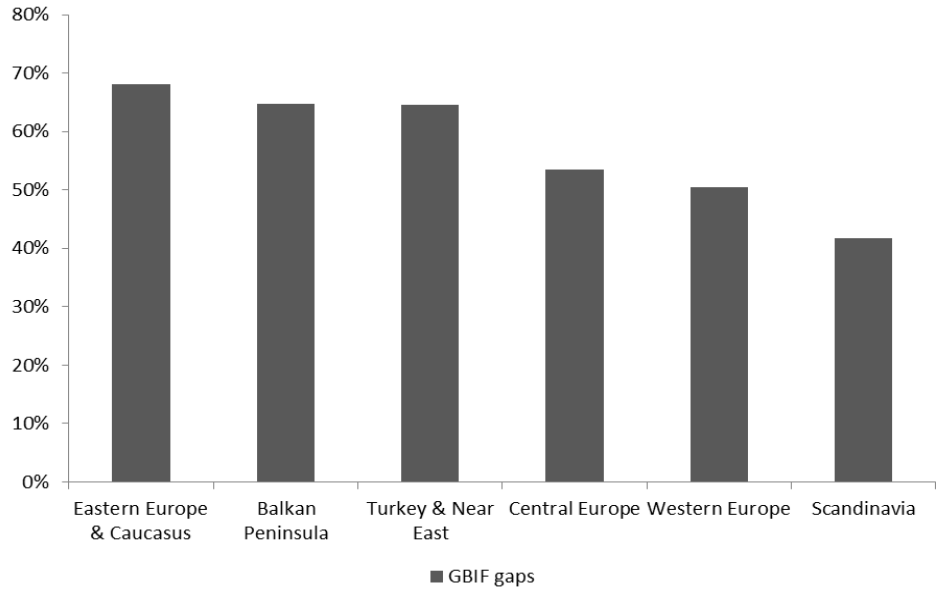


Fig. 32: Gaps across European regions in GBIF data. Gaps are determined by calculating the ratio between the number of lacking bee species occurrence records (dark grey: GBIF gaps) and the number of all bee species for the region (baseline is the Checklist data for a given country for 1245 bee species with GBIF records ~ 48,9% of bee species in Europe).

Example – data differences in time, and data for a country:

For eight selected bee species, we looked at 1) the total number of data shared through GBIF on two different dates, in November 2013 and February 2014; and 2) GBIF records for Denmark (see Table 12 below).

Obviously, the number of GBIF data records at a certain time is a "snapshot" of data mobilisation and GBIF data indexing. For one species, *Bombus pascuorum*, there is a huge difference in the total number of occurrences between the two dates. The exact reason for this difference is unknown and was probably caused by removal of doublets in datasets which by error had been indexed/registered twice in the GBIF system. For *Halictus scabiosae* the number of data was reduced by more than half from November to February. On the GBIF-portal, it is possible to track activities of datasets being cleaned and republished etc. Here, it appears that a certain dataset underwent cleaning by the provider, related to georeferencing etc., and seems to have been republished. This, and the example above with *B. pascuorum*, underpins the importance of being aware that the pool of GBIF-data is a very dynamic structure, reflecting activities of the more than 600 data-publishers sharing over 15,000 datasets.

Hence it is important, if possible, to investigate a given data-compilation at different dates, before drawing final conclusions. Currently, GBIF is developing a system to provide a timestamp and a unique reference ID to a given instance of filtering and downloading data for e.g. a research project – to keep a snapshot of how data "looked" at the time of download for a certain purpose.

From the table below, it is also obvious, that there is an urgent need for mobilisation of occurrence data for bees in Denmark (see the complete list in Annex C). Data for Denmark

are known to exist, and specific datasets are planned to be shared in the future (including the Citizen science initiative Fugle & Natur, http://www.fugleognatur.dk/ and a PhD-project on bees in agricultural landscapes by Isabel Calabuig). Furthermore, in addition to species where there currently are no data shared in GBIF, for 45 % of the 77 species (see Annex C), there is a remarkable difference between the total number of occurrences in Europe and the number of georeferenced occurrences. This indicates that quite a large number of GBIF records lack actual point locations.

For the 283 bee species that occur in Denmark, Fig. 33 shows the current availability of data (accumulated in range-classes) for Denmark, and for Europe as a whole. In general, Fig.33 shows that for most of the wild bee species of Denmark, there are few or no occurrence records available for Denmark in GBIF. Hence, for 206 species in Denmark, there are no records at all and there are no georeferenced records for 267 of the 283 species. Looking at all available occurrence records for Denmark, for 46 of the species, only one record is available, and for 16 species, there are between 11 and 100 records available. Looking at occurrence records with a georeference, nine species in Denmark have one georeferenced record, six species have between 2 and 10 georeferenced records. Only for few species there are a larger number of records (> 100 records) available, i.e. for six species regarding records at all and for only one species regarding georeferenced records. For Europe as a whole, the majority of the 283 Danish species have number of occurrences in European countries that fall within the ranges 11-100 or 101-1,000. Only two species on the checklist for Denmark have no records at all for Europe as a whole.

Table 12: A selection of eight bee species, GBIF georeferenced records, total number of occurrences and GBIF data records in Denmark.

Scientific name (click on name for link to GBIF portal)	**GBIF data records georeferenced and (total number of occurrences) Date: 20131112**	**Date: 20140203**	**GBIF data records in Denmark (comments on actual occurrence in DK)**
Andrena flavipes	7.778 (10.445)	7.895 (10.582)	0 (Common in DK)
Andrena fulva	4.360 (4.596)	4.464 (4.709)	0 (Common in DK)
Andrena hattorfiana	4.405 (4.597)	4.746 (4.930)	0 (Relatively rare in DK)
Anthophora pubescens	19 (23)	8 (13)	0 (Not on Danish species checklist)
Bombus pascuorum	43.975 (80.821)	43.710 (47.824)	0 (Common in DK)
Halictus scabiosae	267 (395)	94 (153)	0 (Not on Danish species checklist)
Lasioglossum clypeare	5 (8)	5 (8)	0 (Not on Danish species checklist)
Melecta luctuosa	103 (131)	107 (132)	0 (Rare in DK)

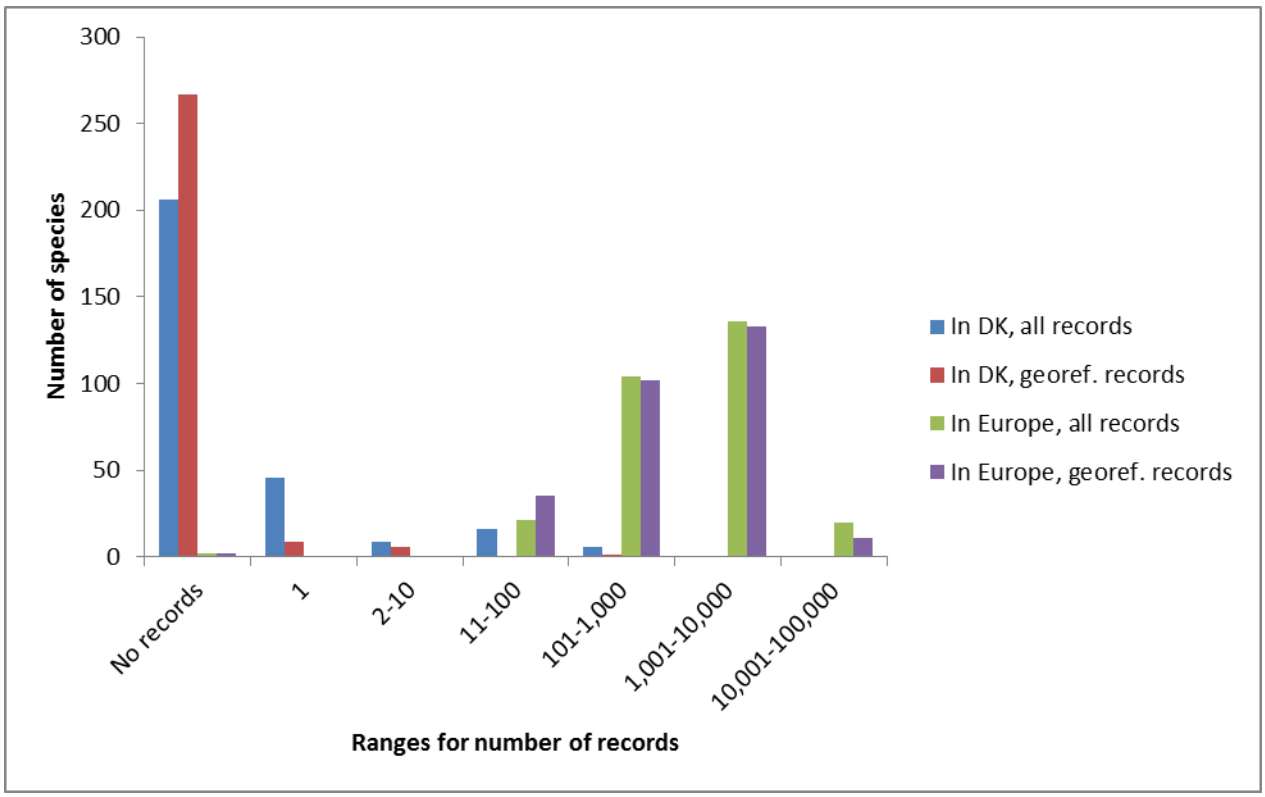


Fig. 33: Overview of number of available occurrence records for 283 bee species available in GBIF (occurrence records range-classes) and number of species which have occurrence records that fall into one of these classes. The occurrence record classes indicate the number of available occurrence records, separately for species country occurrences at all (i.e. all records), and for the number of records with exact point location (georeferenced records).

Comparison Atlas of European Bees versus GBIF data

For 59 of the most common bee species, data could be found on the web page of the Atlas of European Bees. When comparing the number of specimen records shared in GBIF with the available specimen records in the Atlas data, in **57 cases** the Atlas data contains many more specimen records (with geo-referenced location and collection date) than GBIF does. In only **two cases**, GBIF contains more specimen records than the Atlas.

As the figure shows, the difference can be quite significant (see some examples below, Fig. 34a and 34b). Overall, for the selected species, GBIF mediates **346,692** specimen records and observation data with known locations. In the atlas of European bees there are **1.030.803** records, which means **287%** more records.

The difference varies for 57 cases where the Atlas of European bees contains between **113%** and **2013%** more records. On average the Atlas contains **297% more records than GBIF.** For the two cases where GBIF contains more records, GBIF contains **12%** respectively **13%** more records. However, it is not clear if or which of the Atlas data are already shared through the GBIF-portal as well.

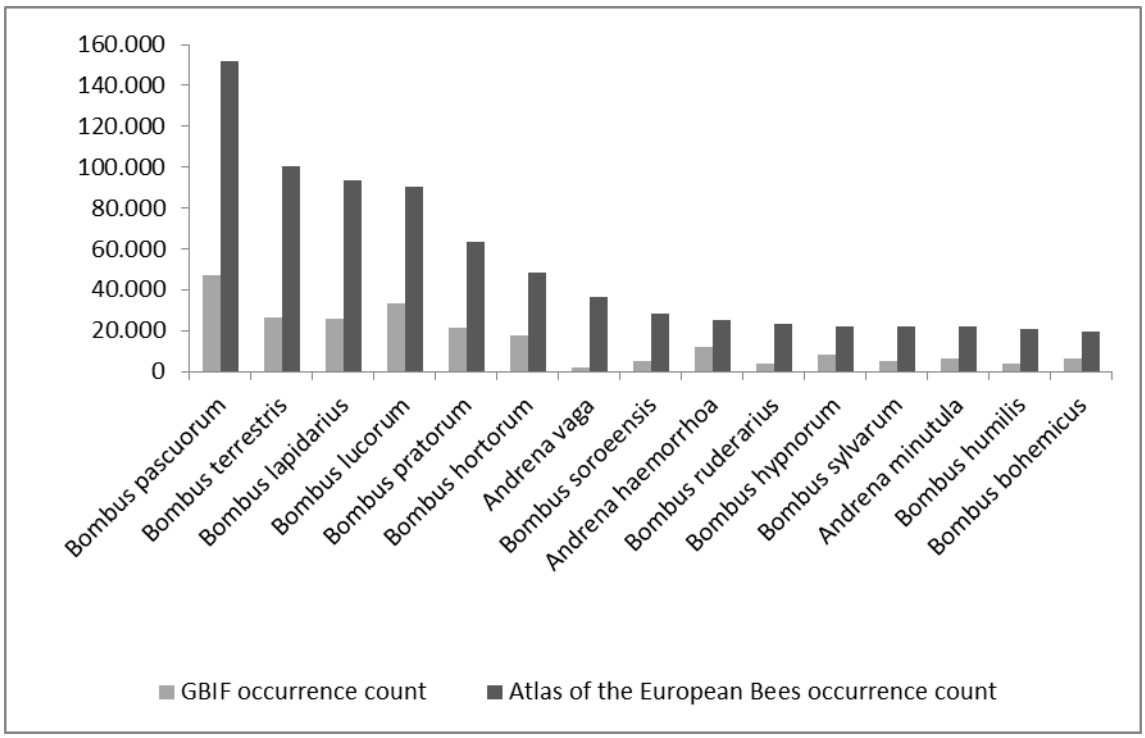


Fig.34a: Comparison of the number of occurrence records for 15 bee species among data from GBIF and from the Atlas of European bees (light grey: GBIF records, dark grey: Atlas of the European bees data).

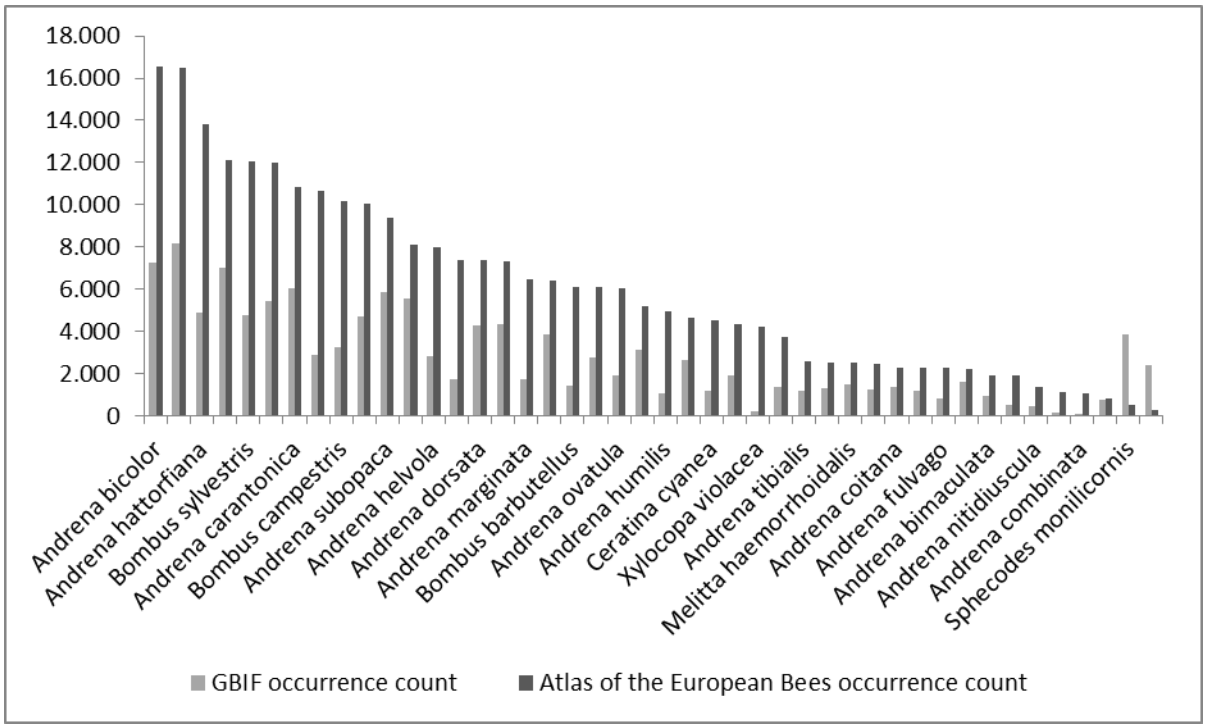


Fig. 34b: Comparison of the number of occurrence records for 22 bee species among data from GBIF and from the Atlas of European bees (light grey: GBIF records, dark grey: Atlas of the European bees data).

4.7.4 Recommendations based on the results of the gap analysis

Due to the results, several recommendations can be made based upon our findings:

- For data on European bee species, there are different data sources that contain relevant biological information on bees. However, these datasets are not all yet integrated in or shared through one common repository.
- As a common repository, GBIF could serve as a free, online available source. Efforts should be intensified to aggregate all available information by using this infrastructure.
- The analysis also showed that GBIF might contain datasets with doublets or inaccurate data. However, it is important here to state, that GBIF serves as a mediator of data provided and shared by its member institutions (countries). GBIF itself does not generate data. Thus, the data providers in GBIF need to implement specific quality checks of their data, to secure the high quality standards.
- On a GBIF participant level, it should be communicated that there is a lack of valuable data for an important group of organisms. GBIF participant nodes could prioritise to mobilise datasets on bees and other pollinators.
- It is important to note that lacunae in GBIF-data not only is a matter of mobilising data, but also reflects the current distribution of participant countries in GBIF. Some European countries are not members, and some countries are (associate) members but do not have the resources to share data in GBIF.
- The European Union currently does not consider the role of pollinators appropriately in their current policy of biodiversity data collection. For example no bee species are part of the Annex II (species requiring designation of Special Areas of Conservation) or Annex IV (species in need of strict protection). The European policy should also safeguard that data on wild bee species will be collected in the future and that the current ongoing activities regarding a European red list of bees are maintained and supported.
- Datasets that are important for scientific research should be made publicly available with unrestricted and online access to the datasets, which is currently not the case for some European bee species databases like for the Atlas of European bee datasets. However, there is also the need to find ways of benefit-sharing for the parties or individuals that collected the data over the years in a cost- and time-consuming way. To promote the free sharing of data, there is the strong need to find incentives for the involved people (publications, financial contribution). In addition to that, there is a need for the development and recommendations for fair and best practice rules for sharing the data. One way to secure the fair use of data is to draft so-called data sharing agreements that define the ways and principles of sharing and determining the benefits for data providers.
- Lack of integration of the available data sources will very likely cause a severe bias in status- and trend analyses of wild bee species. Future modelling of bee species distributions and the impact of change drivers will greatly be improved by integrating these additional sources.

4.7.5 Literature:

Aizen, M.A., Harder, L.D., 2009. The global stock of domesticated honey bees is growing slower than agricultural demand for pollination. Current biology : CB 19, 915-918.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A.P., Potts, S.G., Kleukers, R., Thomas, C.D., Settele, J., Kunin, W.E., 2006. Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. Science 313, 351-354.

Calabuig, I. & H.B. Madsen, 2009. Annotated checklist of the Bees in Denmark – Part 2: Andrenidae (Hymenoptera, Apoidea). – *Entomologiske Meddelelser* 77 (2): 83-113. Copenhagen, Denmark. ISSN 0013-8851. In Danish with UK abstract and legends. Available online at: http://entomologiskforening.dk/tidsskrift/manuskript/2009_Calabuig-Madsen_Ent-Med_Andrenidae_Optimised.pdf/

Dupont, Y. L. & H.B. Madsen, 2010. Bumblebees (in Danish, "Humlebier"). Natur & Museum 2010: No. 1.

Duputié, A., Zimmermann, N. E., Chuine I. (2014): Where are the wild things? Why we need better data on species distribution. Global Ecology and Biogeography 23, 457-467.

Gallai, N., Salles, J.-M., Settele, J., Vaissière, B.E., 2009. Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. Ecological Economics 68, 810-821.

Madsen, H.B. & I. Calabuig, 2008. Annotated checklist of the Bees in Denmark – Part 1: Colletidae (Hymenoptera, Apoidea). – *Entomologiske Meddelelser* 76 (2): 145-163. Copenhagen, Denmark. ISSN 0013-8851. In Danish with UK abstract and legends. Available online at: http://entomologiskforening.dk/tidsskrift/artikler/aargang-2008-bind76-2/entmed2008_76-2_madsen_calabuig.pdf

Madsen, H.B. & I. Calabuig, 2010. Annotated checklist of the Bees in Denmark – Part 3: Melittidae & Megachilidae (Hymenoptera, Apoidea). – *Entomologiske Meddelelser* 78 (2): 73-99. Copenhagen, Denmark. ISSN 0013-8851. In Danish with UK abstract and legends. Available online at: http://snm.ku.dk/english/staffsnm/staff/?pure=files%2F33528268%2F2010_Madsen_Calabuig_Ent_Med_Megachil_Melitt_Fullversion.pdf

Madsen, H.B. & I. Calabuig, 2011. Annotated checklist of the Bees in Denmark – Part 4: Halictidae (Hymenoptera, Apoidea). – *Entomologiske Meddelelser* 79 (2): 85-115. Copenhagen, Denmark. ISSN 0013-8851. In Danish with UK abstract and legends. Available online at: http://entomologiskforening.dk/tidsskrift/artikler/aargang-2011-bind79-2/entmed2011_79-2_85-115.pdf/

Madsen, H.B. & I. Calabuig, 2012. Annotated checklist of the Bees in Denmark – Part 5: Apidae (Hymenoptera, Apoidea). – *Entomologiske Meddelelser* 80 (1): 7-52. Copenhagen, Denmark. ISSN 0013-8851. In Danish with UK abstract and legends. Available online at: http://entomologiskforening.dk/tidsskrift/artikler/aargang-2012-bind80-1/entmed2012_80-1_07-52_madsen-og-calabuig.pdf/

Madsen, H.B. & Y.L. Dupont, 2013. Wild bees (in Danish, " Vilde Bier"). Natur & Museum 2013: No. 1.

Ollerton, J., Winfree, R., Tarrant, S., 2011. How many flowering plants are pollinated by animals? Oikos 120, 321-326.

Rasmont P. & Iserbyt I. 2010-2013. Atlas of the European Bees: genus Bombus. 3d Edition. STEP Project, Atlas Hymenoptera, Mons, Gembloux.

Sangild, S. (Editor), chapter on bees by H.B. Madsen & I. Calabuig, 2007. Insects in Colours (in Danish, "Insekter i farver"). Politikens Forlag, ISBN-13: 978-87-567-7273-0

Annex B: List of Pan-European Countries analyzed in the study

Albania, Andorra, Armenia, Austria, Azerbaijan, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Moldova, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom

Annex C: List of Danish bee species with GBIF records

Of the 283 bee species on the Danish checklist, only 77 species have GBIF data records (total count 2889), of which only 16 species have georeferenced records (total count 304) (counts from June 2014).

For the same 77 species, GBIF data records in Europe are listed. For roughly 45 % of these species, there is a remarkable difference between the total number of occurrences in Europe and the number of georeferenced occurrences.

	GBIF data records in Denmark			**GBIF data records in Europe**	
Scientific name for the 77 Danish bee species with GBIF data records (out of in all 283 species on the Danish checklist)	**Total number of occurrences**	**Geo-referenced**	**Comments on actual occurrence in DK***	**Total number of occurren-ces**	**Georeferenced**
Andrena barbilabris (Kirby, 1802)	1	0	Occurs sporadically	4610	4405
Andrena bicolor Fabricius, 1775	1	0	Common	7491	7372
Andrena carantonica Pérez, 1902 – see note #1 below	1	0	Common	6264	6193
Andrena clarkella (Kirby, 1802)	1	1	Common	3948	3907
Andrena fucata Smith, 1847	1	0	Relatively rare	3336	3305
Andrena fulva (Müller, 1766)	2	2	Common	4685	4610
Andrena fuscipes (Kirby, 1802)	1	0	Relatively rare	3085	3018
Andrena nigroaenea (Kirby, 1802)	1	0	Common	7193	7073
Andrena subopaca Nylander, 1848	1	0	Common	6281	6137
Andrena varians (Kirby, 1802)	1	0	Rare	624	573
Anthidium manicatum (Linnaeus, 1758)	1	0	Common	2486	2251
Anthidium punctatum Latreille, 1809	1	0	Relatively rare	1310	1295
Anthophora furcata (Panzer, 1798)	1	0	Common	2033	1963
Anthophora quadrimaculata (Panzer, 1798)	1	0	Common	1175	1153
Apis mellifera Linnaeus, 1758	1	0	Common	9422	9163
Bombus barbutellus (Kirby, 1802)	23	0	Rare	2748	1492
Bombus bohemicus Seidl, 1837	188	4	Common	12812	7195
Bombus campestris (Panzer, 1801)	8	0	Rare	6650	3452
Bombus cullumanus (Kirby, 1802)	1	0	Probably extinct	306	90
Bombus distinguendus Morawitz, 1869	73	0	Occurs sporadically	5720	3909
Bombus hortorum (Linnaeus, 1761)	334	1	Common	33603	21010

Bombus humilis Illiger, 1806	16	0	Rare	11659	4055
Bombus hypnorum (Linnaeus, 1758)	225	4	Common	15295	9538
Bombus jonellus (Kirby, 1802)	11	0	Occurs sporadically	15884	12871
Bombus lapidarius (Linnaeus, 1758)	16	1	Common	34450	28023
Bombus lucorum (Linnaeus, 1761)	62	0	Common	48182	38841
Bombus muscorum (Linnaeus, 1758)	45	0	Relatively rare, local	11288	6855
Bombus norvegicus (Sparre-Schneider, 1918)	1	0	Common	1215	686
Bombus pascuorum (Scopoli, 1763)	797	277	Common	87856	55738
Bombus pomorum (Panzer, 1805)	78	0	Probably extinct or occurs sporadically	1930	86
Bombus pratorum (Linnaeus, 1761)	124	0	Common	40944	25912
Bombus quadricolor (Lepeletier, 1832)	4	0	Probably extinct	1194	318
Bombus ruderarius (Müller, 1765)	76	0	Relatively rare	12337	4392
Bombus ruderatus (Fabricius, 1775)	37	0	Probably extinct or occurs sporadically	4826	1001
Bombus rupestris (Fabricius, 1793)	63	0	Common	7004	3156
Bombus soroeensis (Fabricius, 1777)	72	0	Relatively rare, local	18826	6104
Bombus subterraneus (Linnaeus, 1758)	26	0	Rare	3805	1755
Bombus sylvarum (Linnaeus, 1761)	402	0	Relatively rare	10915	5629
Bombus sylvestris (Lepeletier, 1832)	24	0	Common	9108	5652
Bombus terrestris (Linnaeus, 1758)	56	3	Common	35810	28001
Bombus vestalis (Geoffroy, 1785)	3	0	Common	8464	5490
Bombus veteranus (Fabricius, 1793)	65	0	Rare	3570	475
Chelostoma rapunculi (Lepeletier, 1841)	1	0	Common	1638	1325
Colletes cunicularius (Linnaeus, 1761)	3	0	Common	3150	3096
Colletes daviesanus Smith, 1846	1	0	Common	3868	3715
Colletes impunctatus Nylander, 1852	1	0	Not known	304	289
Dasypoda hirtipes (Fabricius, 1793) See note #2 below	1	0	Common, local	198	83
Halictus tumulorum (Linnaeus, 1758)	1	0	Common	11270	11098
Hylaeus annularis (Kirby, 1802)	1	1	Not known	2243	2118
Hylaeus confusus Nylander, 1852	1	1	Relatively common	6203	6009
Hylaeus pectoralis Förster, 1871	1	1	Rare, local	924	917
Lasioglossum albipes (Fabricius, 1781)	1	0	Common	8885	7906

Lasioglossum calceatum (Scopoli, 1763)	1	0	Common	17411	13711
Lasioglossum lativentre (Schenck, 1853)	1	0	Relatively common	1951	1751
Lasioglossum leucopus (Kirby, 1802)	2	0	Common	10318	9272
Lasioglossum leucozonium (Schrank, 1781)	1	0	Common	8753	6402
Lasioglossum morio (Fabricius, 1793)	1	0	Common	14396	11985
Lasioglossum villosulum (Kirby, 1802)	1	0	Not known	8372	6358
Megachile apicalis Spinola, 1808	1	0	Probably extinct or occurs sporadically	164	72
Megachile lagopoda (Linnaeus, 1761)	1	0	Relatively rare	1471	1398
Megachile leachella Curtis, 1828	1	0	Common locally	154	100
Megachile willughbiella (Kirby, 1802)	4	3	Common	4710	4570
Melitta haemorrhoidalis (Fabricius, 1775)	1	0	Common	1535	1477
Melitta leporina (Panzer, 1799)	1	1	Common	1971	1881
Nomada ferruginata (Linné, 1767)	1	1	Common	299	275
Nomada flava Panzer, 1798	1	0	Common	3763	3671
Nomada flavoguttata (Kirby, 1802)	1	0	Common	4760	4681
Nomada flavopicta (Kirby, 1802)	1	0	Common	1122	1073
Nomada goodeniana (Kirby, 1802)	1	0	Common	4425	4335
Nomada marshamella (Kirby, 1802)	2	1	Common	6020	5919
Nomada similis Morawitz, 1872	1	0	Rare	55	24
Osmia aurulenta (Panzer, 1799)	1	0	Common	1745	1582
Panurgus banksianus (Kirby, 1802)	1	0	Common locally	1690	1667
Sphecodes crassus Thomson, 1870	1	0	Not known	2711	2654
Sphecodes ephippius (Linné, 1767)	1	0	Not known	4829	4722
Sphecodes geoffrellus (Kirby, 1802)	1	0	Common	5910	5877
Sphecodes miniatus Hagens, 1882	2	2	Common	730	667

* Nomenclature and occurrence of species in Denmark is according to: Sangild (Ed.), 2007; Madsen & Calabuig, 2008-2012; Calabuig & Madsen, 2009; Dupont & Madsen, 2010; Madsen & Dupont, 2013; http://www.fugleognatur.dk/

#1: Due to former name-confusions, synonyms etc., in Denmark, *Andrena carantonica* Pérez, 1902 has also been recorded under the names *Andrena trimmerana* (Kirby, 1802), *Andrena scotica* Perkins, 1919 or *Andrena jacobi* Perkins, 1921. Therefore, the GBIF-portal was checked for records of these species names, occurring in Denmark. No such records were found.

#2: Although the Danish official checklist uses *Dasypoda hirtipes* (Fabricius, 1793), the accepted name according to GBIF is *Dasypoda altercator* (Harris, 1780)

4.8 GENERAL REVIEW OF GAPS IN EUROPEAN MONITORING SCHEMES - ASSESSMENT OF THE EUMON DATABASE

Material and Methods

The EuMon project (Schmeller et al. 2006) started a survey on biodiversity monitoring practices across Europe in 2005 (data used here were extracted in April 2010). For reaching representatives of stakeholder groups involved in monitoring activities (governmental and non-governmental bodies), we distributed announcements of the questionnaire through emails, letters and at conferences to over 1600 individuals and through several national and international mailing lists (including national and regional ornithological organizations, national ringing offices, EBCC). We invited all recipients to forward our invitation to their colleagues. We asked respondents to provide data online. The questionnaire was designed to assess how biodiversity monitoring schemes were carried out and what the motivation was to launch that scheme. For general background information we inquired, e.g., the official title of the scheme, the institution it was located at, and the principal coordinator of the scheme. In a second part of the questionnaire, we focused on the design and methodology of a scheme. We asked information on field and sampling methods, and associated statistical considerations. For the taxonomic gap analysis, we compiled the data in a database, which can be reached at http://www.surveygizmo.com/s3/1304066/Global-Survey-of-National-Biodiversity-Monitoring-Schemes. We used the DaEuMON data from over 600 monitoring programs to compile information on species monitoring by country. We performed a taxonomic gap analysis by species groups, but despite the fact that the EuMon survey is the first large-scale survey of its kind, it may suffer from biases in taxonomic and geographic coverage. The taxonomic bias was assessed by searching the Zoological Records and Google Scholar for references to monitoring. The search query used was (monitoring AND species group AND europ. AND biodiversity). We computed the bias as logit(observed) - logit(expected), where the observed values are the values from our database and the expected values the total of institutions contacted, respectively the records from Zoological Records or Google Scholar, meeting our search criteria. However, note that two databases may suffer from the same type of biases as our survey – differential inclinations of monitoring schemes to publish their results – and the biases in our data may not differ much from usual publication biases.

In DaEuMon fishes are strongly underrepresented and raptors and waterbirds overrepresented, while birds in general show no bias based on the Web of Science, but an overrepresentation in regard to records in Google Scholar (Fig. 35). To determine the species numbers per species group present in one country, we solicited a range of different sources, including the national reports in EUROBAT, Amphibia Web, Fishbase, Nation Master (Birds and mammals). We then compared how many species were covered by the monitoring schemes in DaEuMon and calculated the proportion of coverage in the EU countries (or candidate countries). We could not find records on species numbers for all species groups and countries, e.g. missing butterfly numbers of Croatia, Iceland, and Norway.

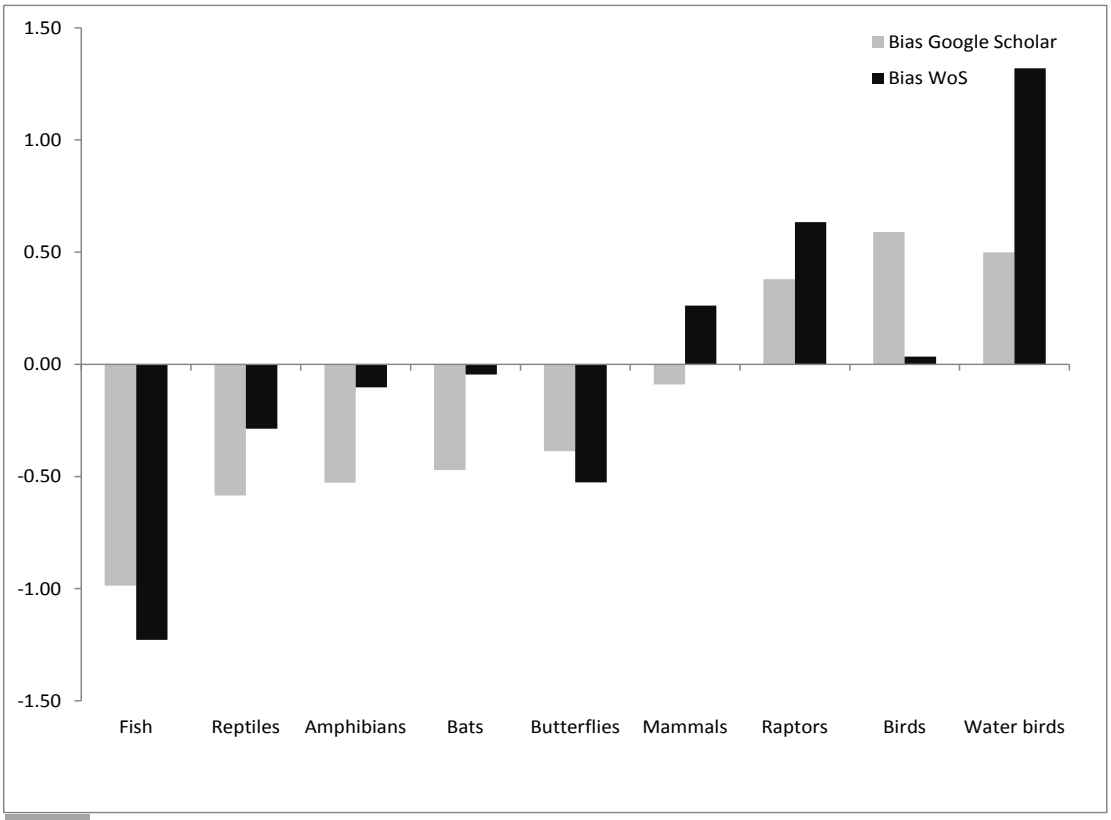


Fig. 35: Bias in the taxonomic coverage in the monitoring program database of the project EuMon

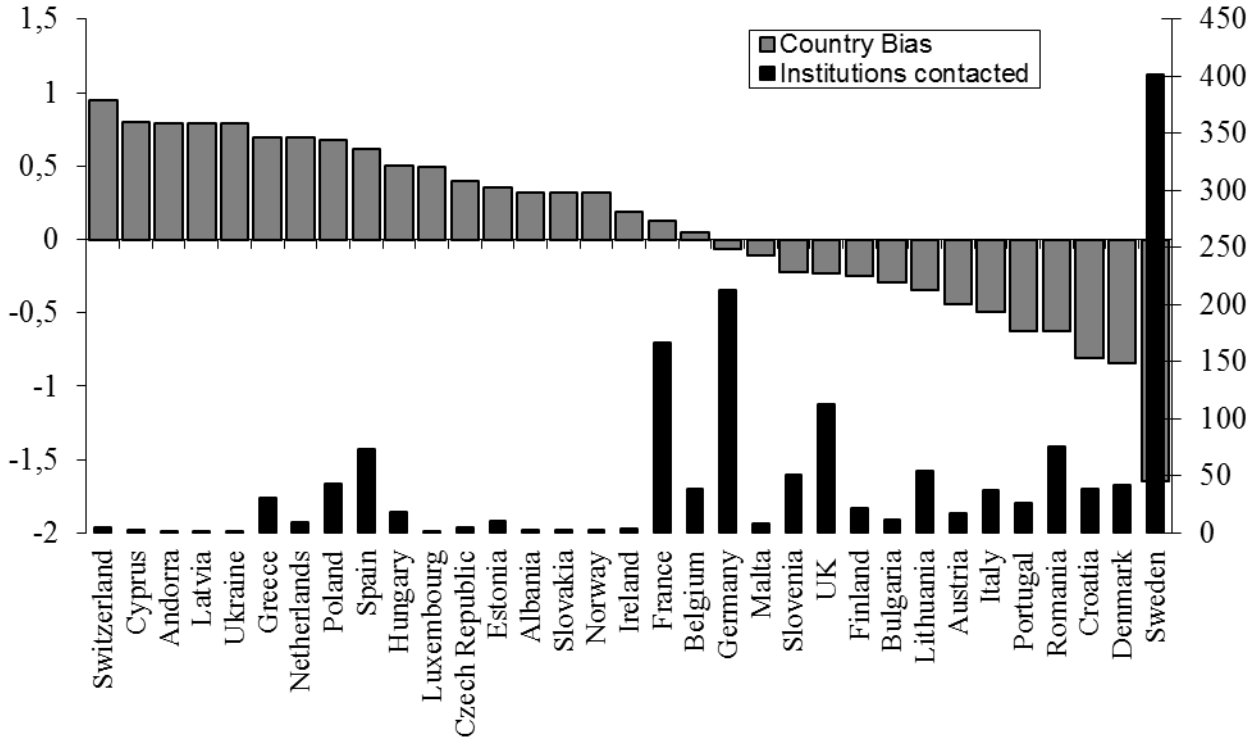


Fig. 36: Bias by country

Another problem is the country bias in DaEuMon (Schmeller, 2012;Schmeller, 2009). Hence, in some cases, we may underestimate the coverage due to a lack of schemes in our database. In butterfly species, we found a higher coverage (>100%) than species. This might be due to an underestimation of species due to not updated records, erroneous entries in our database or an mismatch of names when combining information from different monitoring schemes. However, in these cases, we assumed that the coverage would be 100%.

Table 13: Taxonomic coverage by species group and country.

Country		butterflies	amphibians	fish	Birds	Mammals	Bats	reptiles	Raptors
Austria		40.6%	0.0%	0.0%	0.4%	1.2%		0.0%	3.0%
Belgium		34.1%	47.4%	2.1%	53.9%	19.0%	50.0%	0.0%	0.0%
Bulgaria		0.0%	33.3%	0.0%	0.0%	16.0%	39.4%	36.8%	0.0%
Croatia			6.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Cyprus		0.0%	0.0%	0.0%				0.0%	0.0%
Czech Republic		118.6%	0.0%	0.0%	56.1%	0.0%	0.0%	0.0%	0.0%
Denmark		100.0%	0.0%	0.0%	63.8%	9.3%	0.0%	0.0%	21.2%
Estonia		0.0%	83.3%	0.0%	90.7%	18.5%	91.7%	83.3%	100.0%
Finland		111.8%	0.0%	0.0%	72.0%	3.3%	0.0%	0.0%	44.1%
France		57.4%	23.8%	0.0%	95.4%	21.5%	58.8%	0.0%	25.0%
Germany		44.9%	81.8%	208.3%	73.3%	26.3%	50.0%	50.0%	21.2%
Greece		0.0%	0.0%	9.8%	1.2%	0.0%	0.0%	2.9%	0.0%
Hungary		5.9%	105.9%	39.0%	54.8%	25.3%	57.1%	70.6%	11.4%
Iceland				0.0%	0.0%	0.0%			0.0%
Ireland		113.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Italy		106.1%	0.0%	0.0%	60.0%	0.0%	0.0%	0.0%	0.0%
Latvia		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Lithuania		0.0%	0.0%	71.9%	72.1%	55.9%	100.0%	0.0%	3.8%
Luxembourg		0.0%	0.0%	0.0%			0.0%		10.0%
Netherlands		90.9%	94.1%	0.0%	114.6%	14.5%	30.4%	43.8%	0.0%
Norway			0.0%	108.7%	124.1%	18.5%	0.0%	0.0%	6.3%
Poland		13.6%	94.4%	72.1%	117.2%	83.3%	104.8%	75.0%	86.5%
Portugal		0.0%	0.0%	0.0%	38.3%	0.0%	0.0%	0.0%	0.0%
Romania		0.0%	0.0%	0.0%	42.8%	0.0%	0.0%	0.0%	0.0%
Slovakia		0.6%	0.0%	0.0%	0.5%	0.0%	0.0%	0.0%	3.2%
Slovenia		0.0%	47.1%	113.4%	41.8%	36.0%	90.0%	0.0%	23.5%
Spain		0.0%	21.6%	0.0%	88.3%	0.0%		1.5%	71.4%
Sweden		0.0%	0.0%	0.0%	6.6%	0.0%	0.0%	0.0%	0.0%
United Kingdom		0.0%	0.0%	0.0%	85.6%	18.0%	0.0%	0.0%	17.6%

The highest mean taxonomic coverage was achieved in Poland and Germany (Fig. 37). However, the latter was mainly driven by an overcoverage of 208% for fishes. The countries with zero taxonomic coverage are underrepresented in our database (Fig. 36).

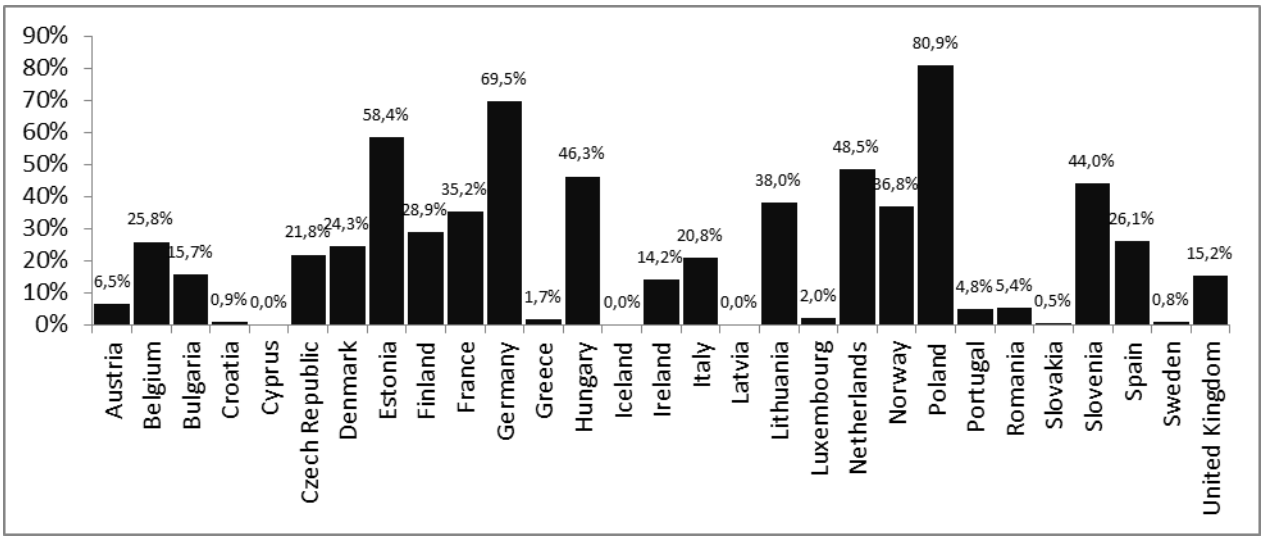


Fig. 37: Mean taxonomic coverage per country across all species groups.

Across Europe birds in general are the best covered species group (48.1%), followed by butterflies, bats and amphibians (Fig. 38).

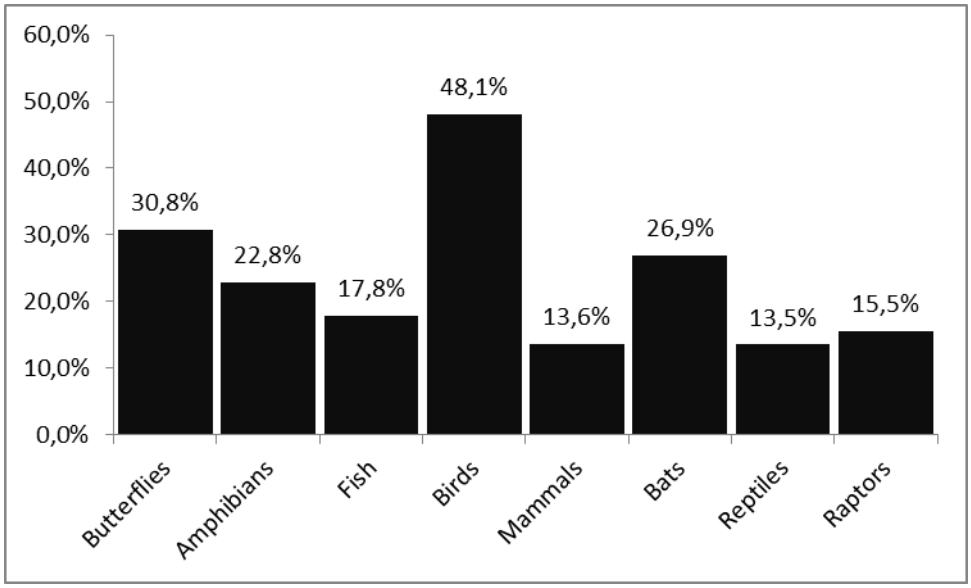


Fig. 38: Mean taxonomic coverage by species group across European Countries

To what extent each of the species groups is spatially well covered is difficult to determine. However, the EuMon database suggests that most countries achieve a 100% or near to 100 % spatial representation of their monitoring schemes.

4.9 FOCUSED-REVIEW OF GAPS IN A SPECIFIC MONITORING SCHEME: ATLAS OF EUROPEAN BREEDING BIRDS (VERSION 1&2) AND THE PAN EUROPEAN COMMON BIRD MONITORING SCHEME

The European Bird Census Council (EBCC) promotes bird monitoring and atlas work across Europe and joins efforts carried out at national level to build a robust strategy to determine bird species distribution and population trends at a European level. Up to date, the most relevant examples of this EU-wide strategy are the Atlas of European Breeding Birds and the Pan-European Common Bird Monitoring Scheme.

4.9.1 Introduction - Short overview of the Atlas of European Breeding Birds

The first comprehensive Atlas of European Breeding Birds was the first major initiative of the EBCC. The final product was an impressive book with more than 900 pages with maps of distribution, accompanying text and information on the population size estimates for key countries where it is present (Hagemeijer and Blair 1997). The atlas data has been used by a wide variety of researchers and conservationists for purposes ranging from estimating hotspots of species occurrence to predicting the effects of climate change. As an example of its wide interest, a recent review by Tulloch et al. (2013) showed that this has the highest Google Scholar citation rate among all bird atlases in the world.

A total of 339 386 species records from 497 European bird species was collected in 3 959 50x50 km squares (UTM grid). This dataset can also be viewed through the internet by means of an interactive portal (http://s1.sovon.nl/ebcc/eoa/) currently hosted by SOVON Dutch Centre for Field Ornithology (Fig. 39).

Fig 39. Example of the EBCC Atlas data for the Sardinian Warbler *Sylvia melanocephala* as shown in the on-line portal. As in the book, the dots on the map refer to six different categories of information on breeding evidence and abundance.

4.9.2 Coverage of the dataset

The area covered in the European Breeding Bird Atlas included all of Europe, including The European part of Russia and Kazakhstan, Madeira, the Azores, Iceland, Svalbard, Franz Josef Land and Transcaucasia) although not Turkey, Cyprus and the Canary Islands.

It integrated mainly field data for the period 1985-1988, but this depended a lot on the situation of each country and for some of them data recorded in the 1970s (in Russia even earlier) were also included.

4.9.3 Outline of gaps and biases

The European Breeding Bird Atlas collected data for the whole of Europe but not with the same intensity everywhere. No data was recorded in a high number of squares, especially in Russia, and coverage was also considered poor or incomplete in many other squares (Fig. 40)

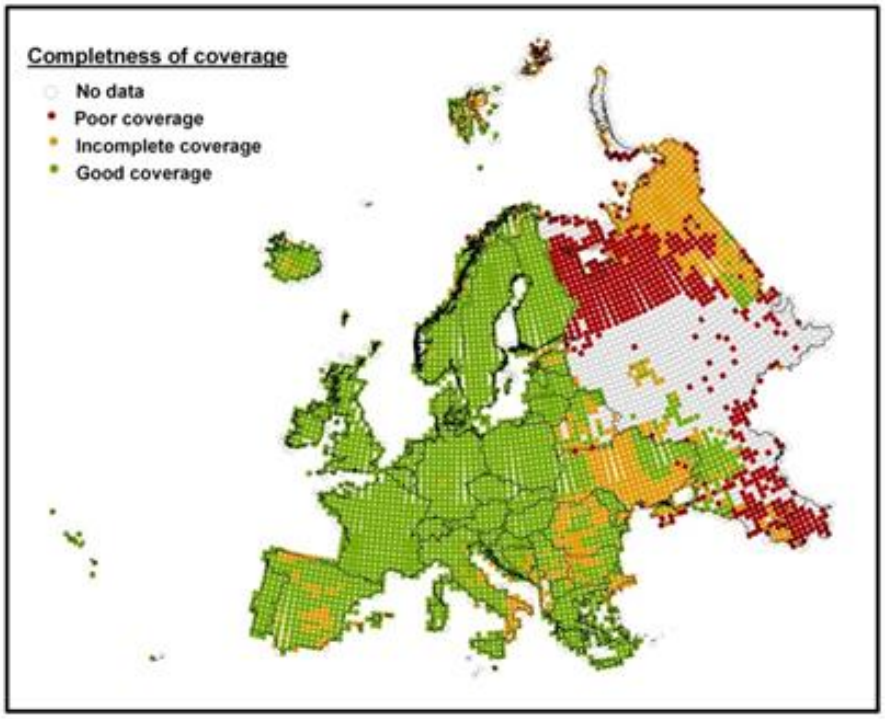


Fig. 40. Completeness of coverage at each of the 50x50 km squares of the European Breeding Bird Atlas (Hagemeijer and Blair 1997).

Regarding temporal biases, not all available data comes from the standard study period (1985-1988). In some cases, even data from the 1950 or 1960 was included (see Table 14 below)

Table 14: Study period for the EBCC Atlas data

Country	Years of fieldwork	Country	Years of fieldwork
Albania	85 and 93	Italy	79-92
Armenia	85-93	Latvia	85-88
Austria	85-88 (some 89/90)	Lithuania	85-88
Azerbaijan	94 (few species only)	Luxembourg	76-90
Belarus	72-92	Macedonia	85-92
Belgium	85-89	Malta	85-88
Bosnia and Hercegovina	85-92	Moldova	86-90
Bulgaria	80-89	Montenegro	85-92
Croatia	85-88	Netherlands	84-90
Czech Republic	85-88 (some 89)	Norway	50-89 (-94)
Denmark	85-88	Poland	86-93 (some 94/95)
Estonia	85-88	Portugal	78-89 (-95 in Azores)
Faeroe Islands	81-89	Romania	77-92
Finland	86-90	Russia	63-94
France	80-92	Serbia	85-92
Georgia	92	Slovakia	85-88
Germany	79-90	Slovenia	79-88
Great Britain	85-88 (some 89)	Spain	70-92 (includes Andorra)
Greece	81-90	Sweden	86-91
Hungary	79-91	Switzerland	85-88 (some 89/90 included Liechtenstein)
Iceland	85-95	Turkey	88-95 (only European part)
Ireland	85-88	Ukraine	80-93

In principle there were no evident biases with respect to the different avian orders or families and data for all bird species was collected everywhere.

It is important to highlight that there are limitations in data quality in this atlas that should be taken into account for further research. Essentially these are:

- There was no prescribed set of fieldwork methodologies for establishing presence or absence.
- There was no universal set of methods of preparing the fieldwork results for collation onto the reporting forms.
- Differences in the quality of field ornithology and in the relative numbers of observers will result in varying data quality.
- Some data are partly extrapolated data
- Some data come from literature (mainly in arctic areas).

4.9.4 Data accessibility

The reference website for the project is http://www.ebcc.info/index.php?ID=5. Use of the data is administered via the EBCC Executive Committee and the data extraction and handling is currently done by staff at SOVON in the Netherlands and Catalan Ornithological Institute in Spain, according to agreed rules. There are countless possibilities for using this valuable dataset, and those interested should contact the EBCC Atlas data provider about the conditions for obtaining the data. Requests will be reviewed by the Executive Committee and there are usually costs for its provision to cover data handling.

4.9.5 Trends in accumulation of occurrence data / integration of historical data

This atlas itself does not include any data that allow trend analyses, but see EBBA2 (below)

4.9.6 Recommendations for closing the gaps

The new European Breeding Bird Atlas (EBBA2) is a project promoted by the EBCC and its partners to update the ground-breaking first atlas, whose data are now 30 years old (http://www.ebcc.info/new-atlas.html). This new atlas attempts also to cover many of the geographical gaps of information of the first atlas, in particular in the East of the continent. The efforts currently conducted in countries such as Russia in close cooperation with the EBCC are expected to be especially important to cover these gaps.

The fieldwork period is focused between 2013 and 2017. In a context of over 50 countries (now covering also all Turkey, the Canary Islands and Cyprus) and 10 000 000 km^2, situations are extremely diverse, from countries with intensive atlas work projected within this period to huge remote regions for which ornithological exploration represents a noticeable challenge even today.

Fig. 41: Map of the area covered by the EBBA2 project, which includes 52 countries and 11 million km^2. In addition to the area covered in the first European atlas, EBBA2 includes the whole of Turkey, Cyprus and the Canary Islands.

The EBBA2 methodology attempts to achieve four aims: 1) To document breeding evidence for all bird species at a resolution of 50x50 km, 2) To estimate abundance for all bird species at a resolution of 50x50 km; 3) To determine the changes in bird species distribution at a resolution of 50x50 km since the 1980s and 4) To model fine-grained distribution for as many bird species as possible and project it at a resolution of 10x10 km (Herrando et al. in press). Differently from the first European atlas, in EBBA2 it is expected a more important role for modelling. In particular, this is expected to be important for: 1) covering the gaps of information at a resolution of 50x50 km, and 2) predicting the species occurrence at 10x10 km resolution.

4.9.7 Literature

Hagemeijer, E.J.M., Blair, M.J. (editors)., 1997. The EBCC Atlas of European Breeding Birds: their distribution and abundance. T & A.D. Poyser, London.

Herrando, S., Voříšek, P., Keller, V., 2014. The methodology of the new European breeding bird atlas: finding standards across diverse situations. Bird Census News (in press).

Tulloch, A.I.T., Possingham, H.P., Joseph, L.N., Szabo, J., Martin, T.G., 2013. Realising the full potential of citizen science monitoring programs. Biological Conservation 165, 128-138.

4.10 GENERAL REVIEW OF GAPS IN NUCLEOTIDE SEQUENCE DATABASES

4.10.1 Introduction - Short overview of the datasource

We analyzed nucleotide sequence data in the International Nucleotide Sequence Databases (INSD: GenBank, ENA, DDBJ)[19] with a focus on fungal taxa and the formal fungal barcode, internal transcribed spacer (rDNA ITS) region (Schoch et al., 2012), that in most cases is the marker of choice for the exploration of fungal diversity in biological samples like soil, water, air, tissue, etc.

As of March 2014 there were 390 858, 83 446, and 203 694 rDNA ITS sequences deposited in INSD of fungal, animal, and plant origin, respectively. For animal species the Cox1/COI (689 109 sequences), and for plant species rbcL (70 070 sequences) and matK (68 860 sequences) are more commonly utilized as barcodes for identification purposes and in taxonomic studies.

Two problems are particularly acute in the pursuit of satisfactory taxonomic assignment of newly generated fungal ITS sequences: (i) the lack of an inclusive, reliable public reference data set and (ii) the lack of means to refer to fungal species, for which no Latin name is available in a standardized stable way. International community of mycologists developed the UNITE[20] database for molecular identification of fungi that attempts to solve the problems referred above - all public rDNA ITS sequences are clustered on different similarity thresholds into *Species Hypotheses* (SH). All SH-s are given a unique, stable name of the accession number type, and they are open to third-party annotation to improve their metadata and identifications (Kõljalg et al., 2013).

In this analysis we give an overview of the representation of fungal species based on nucleotide sequence data available in INSD, and compare INSD ITS sequences with full species names against SH-s provided by the UNITE community.

4.10.2 Coverage of the dataset

INSD fungal dataset covers all publicly available fungal rDNA ITS sequences. As of June 2014, the number of sequences deposited in INSD was 417 987. For 49,6% of sequences the country of origin was specified (in total: 201 distinct countries).

For further analysis we generated quality-filtered (chimeric, low quality, and overly short sequences excluded) dataset consisting of 276 898 sequences. In this quality-filtered dataset where third-party annotations (e.g. adding country and geo-coordinates, habitat, isolation source, and identifications) were carried out, 65,7% of records had country of origin, and 15,9% of records geo-coordinates specified. In table 15 the number of sequences and species in INSD and UNITE SH system are shown by continents.

[19] http://www.insdc.org

[20] http://unite.ut.ee

Table 15: Continent / rDNA ITS based fungal species in INSD and UNITE databases (UNITE SH version 6, 98,5% threshold for all tables in this document).

Continent name	INSD sequences [1]	INSD species [2]	QF sequences [3]	UNITE SH species [4]
Europe	73515	7414	57171	14686
Asia	49006	5150	39671	10008
North-America	89346	5541	57483	14328
South-America	15051	1920	12213	3526
Australasia	12410	2054	9275	3566
Africa	8296	1320	7408	2595
Antarctica	1424	262	1176	407
Unspecified	168939	14415	92501	19393
Total (unique)	417987	22873	276898	54540

[1] Number of sequences in INSD
[2] Number of species names in INSD fungal classification (species names like Amanita sp. 1, etc. are excluded)
[3] Number of all quality filtered (QF) sequences included in UNITE SH-s
[4] Number of 98.5% UNITE SH-s

Table 15 shows that although for 34% of sequences the origin of country is unknown, there is a sampling bias towards North-America and Europe with South-America, Australasia and Africa being clearly under-represented.

We also looked at the proportions of shared species (SH-s represented by more than one sequence in UNITE system) between all continents. Fig. 42 illustrates that all continents have similar proportion of species unique to this area but the number of unique species in each continent differs greatly (e.g. 667 in South-America vs 2 643 in Europe). With 20% of all sequences originating from Europe, and another 20% from North-America, in Fig. 42 most of the continents share greater proportion on species with Europe. On the other hand, it can be observed that a low proportion of species present in Europe are shared with South-America, Australasia and Africa – areas that are characterized by scarce sampling and represented by less sequence data.

In INSD dataset 41,7% of sequences were identified to species level whereas 28,6% were identified only on kingdom level (as Fungi sp.) In Table 16 and Fig. 43 the number of species in distinct phyla are compared between different data sources (INSD full species names, UNITE *Species Hypotheses*, Index Fungorum).

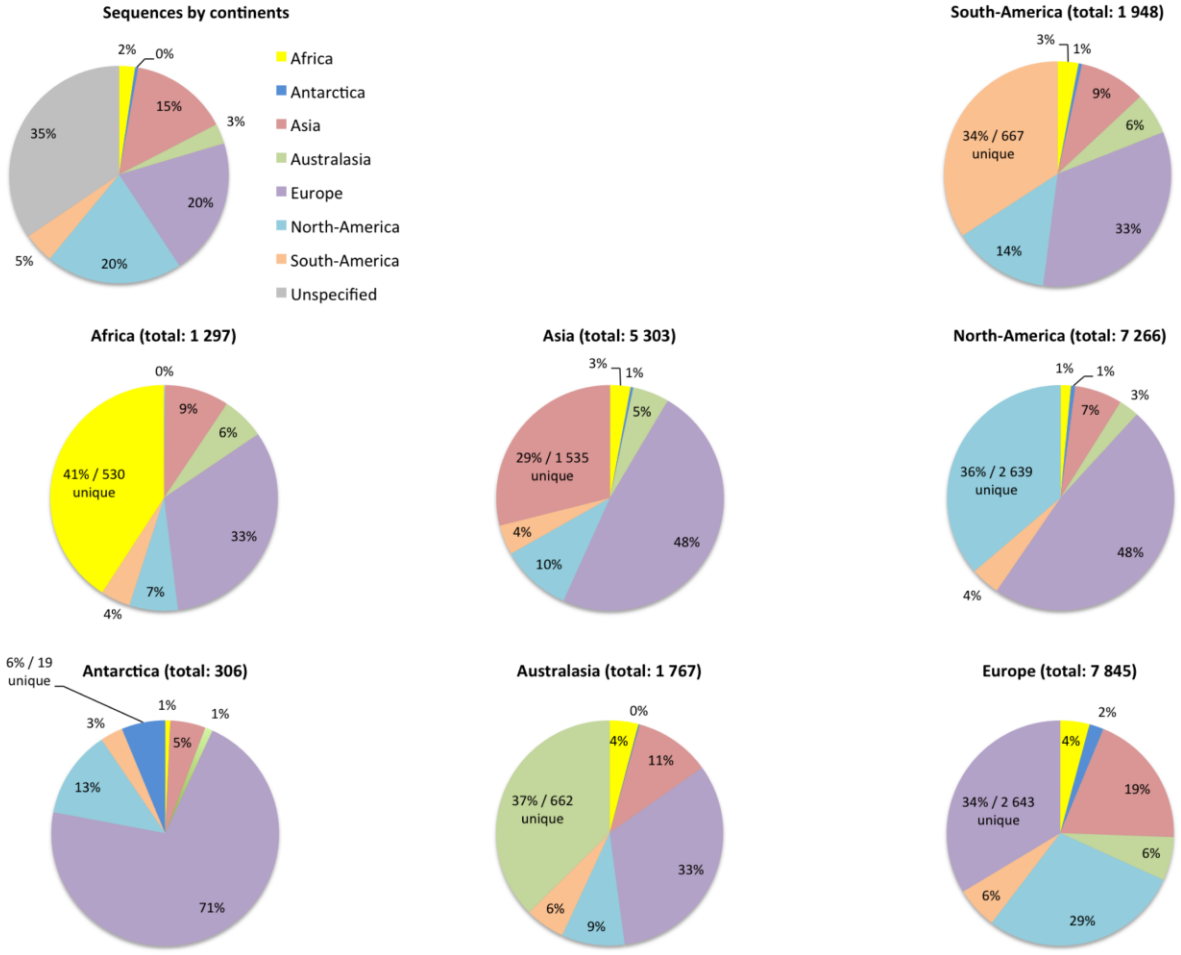


Fig. 42: Unique Species Hypotheses and SH-s shared between continents.

Table 16. Taxonomic coverage – phylum level

Phylum	Index Fungorum	INSD	UNITE SH-s
Ascomycota	78330	13559	24226
Basidiomycota	45727	8573	22517
Blastocladiomycota	0	1	2
Chytridiomycota	1205	172	476
Glomeromycota	195	127	1751
Incertae sedis	7	20	17
Microsporidia	0	92	15
Neocallimastigomycota	0	1	0
Zygomycota	1291	408	944
unidentified	0	14	4364
Total	126755	22967	54312

[1] Only current names on species level and below are included

[2] Only full species names on species level or below are included

[3] 98.5% UNITE SH-s

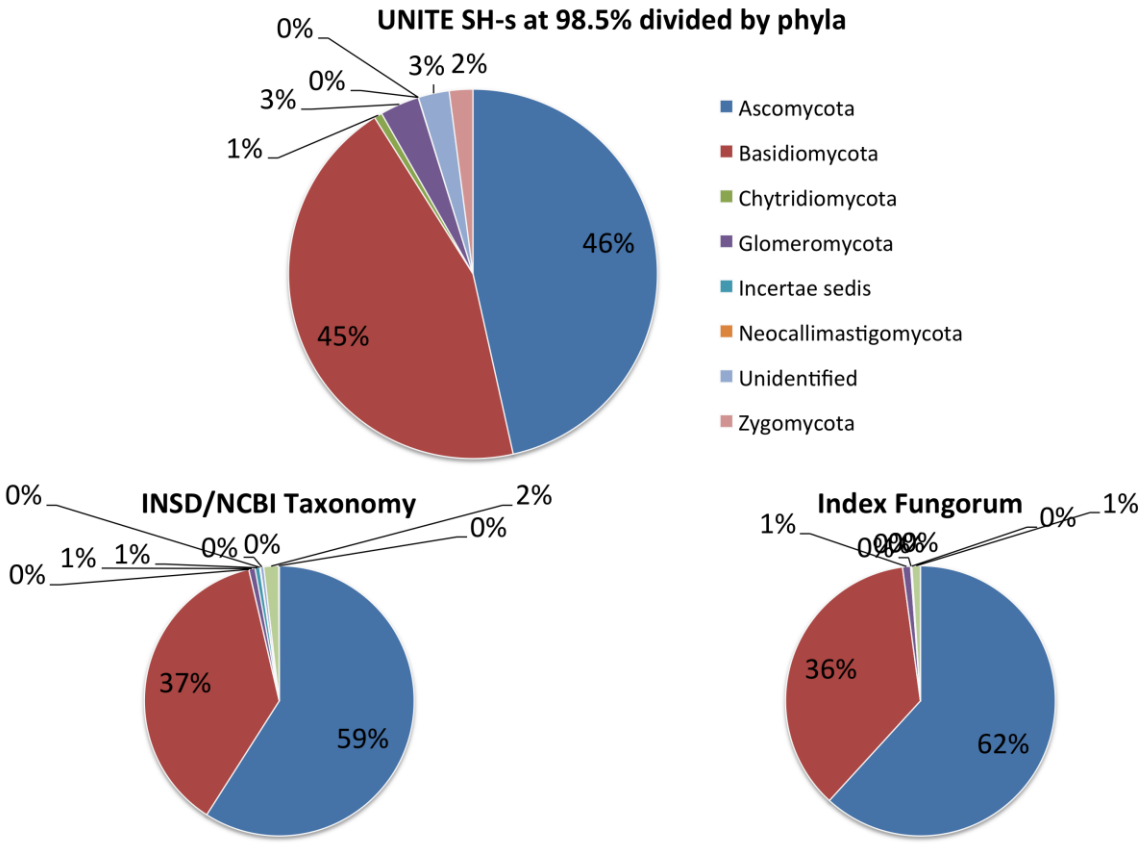


Fig. 43 illustrates that although there are more sequences identified as Ascomycota (43% of total sequences compared to 23% in Basidiomycota), and there are more taxa described in Ascomycota in both Index Fungorum and INSD, the number of taxa resulted in sequence similarity analysis is comparable in these two phyla.

4.10.3 Outline of gaps, biases and data quality (spatial, taxonomic, temporal gaps)

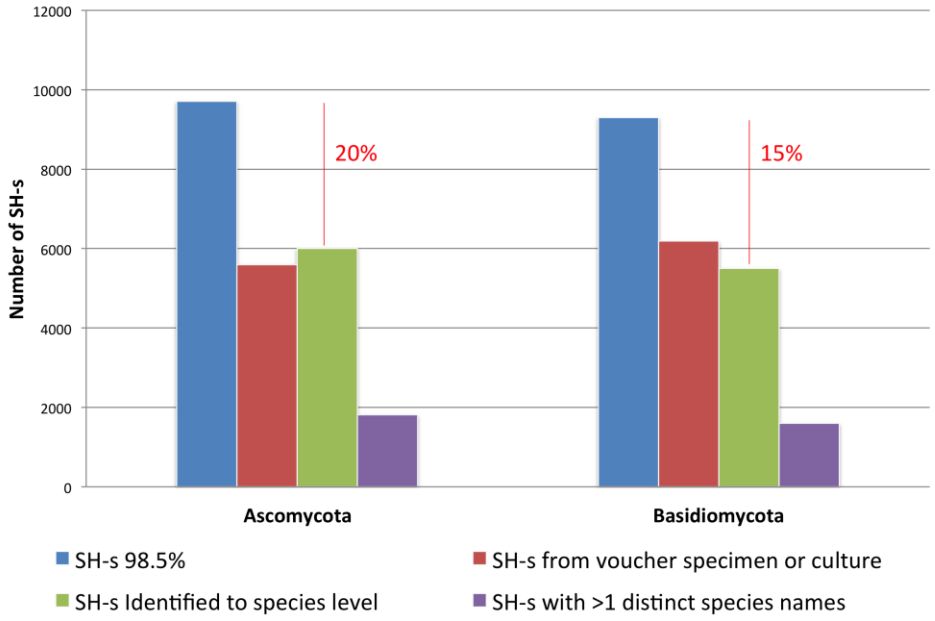
The main gaps in nucleotide sequence data deposited in INSD are related to sequence quality, missing- and misidentifications, lack of metadata, non-use of metadata standards, and sampling bias.

It is suggested that up to 20% of fungal sequences in public repositories can be incorrectly identified (Nilsson et al., 2006), low quality or chimeric. This affects the species identification process in ecological and meta-barcoding studies by compromising the results, and can cause the propagation of incorrect data. There have been a number of research articles published recently (Nilsson et al., 2012, Lindahl et al., 2013) with guidelines on how to collect, analyze, and deposit sequence data to overcome the problems mentioned above.

Fig. 44 illustrates the gap of missing identifications in public sequence repositories reflecting our current knowledge on fungal taxa and how it relates to known, formally described taxa – approximately 35% of species known from DNA cannot be assigned to any known species for which full species name in Linnaean classification is available today. It is important to note here that approximately 20% of formally described fungal species are represented with DNA barcode sequence in INSD, so it might be that DNA-based taxa we cannot assign full

species name to might be described and be present in herbaria, but have never been sequenced.

Fig. 44: Taxonomic coverage – proportion of DNA-based species for which no latin binomial is currently available.



In this document we have shown, and it has also been previously noted (Tedersoo et al., 2011), that the lack of metadata and non-use of standardized vocabulary for recording metadata is common when depositing sequence data into public repositories. With the introduction and propagation of Environment Ontology (ENVO, http://environmentontology.org/) and MixS standard together with environmental packages by Genomic Standards Consortium (GSC, http://gensc.org/) this already shows signs of improvement and acceptance by the community.

There have been a number of third-party annotation efforts for the public nucleotide sequence datasets where subsets of INSD data have been downloaded into inhouse databases, quality filtered, annotated, and made publicly available for the research community. Few of the examples include RefSeq (Schoch et al., 2014), UNITE, and RDP (http://rdp.cme.msu.edu/). The main idea behind these databases is to improve the tools and reference datasets for sequence-based identification.

We also illustrated the sampling bias where majority of sequences originate from Europe and North-America. With the introduction and wider use of next-generation sequencing techniques in recent years that can be applied to a variety of biological samples with global coverage this gap is also expected to disappear in the next five years time.

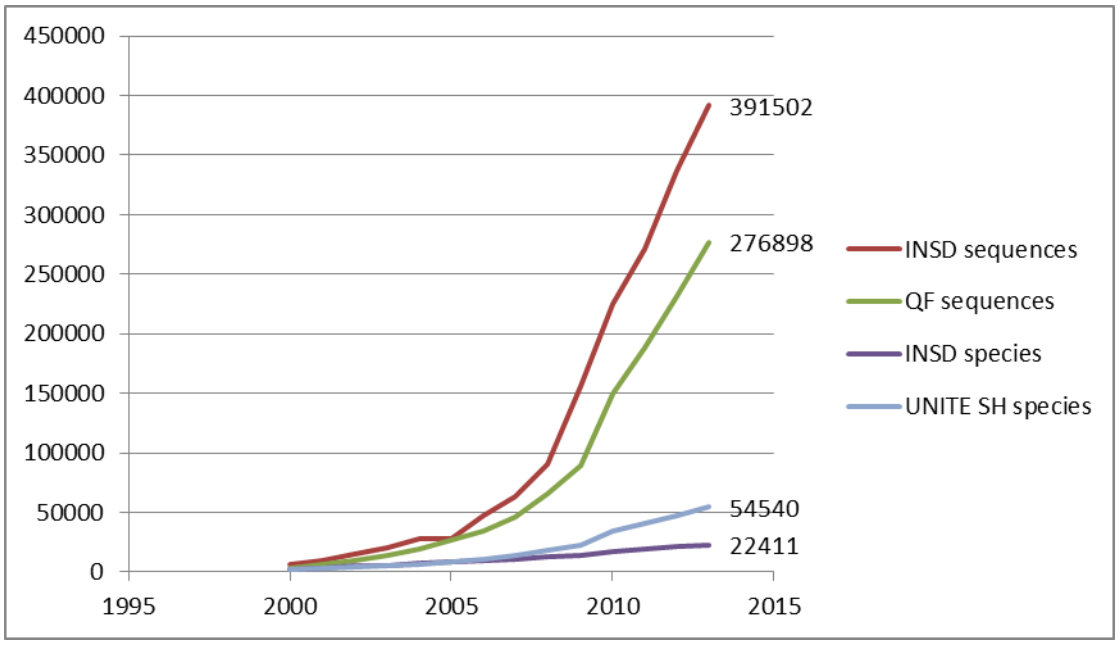
4.10.4 Data accessibility

Data deposited in INSD is publicly accessible and freely usable. There have been developed extensive tools to search and download the data using both web browser and application programming interface.

4.10.5 Trends in accumulation of occurrence data / integration of historical data

Fig. 45 shows the accumulation of fungal rDNA ITS sequences and full species names in INSD and UNITE system. While the number of sequences shows growth pattern similar to that of exponential, the number of full species names shows little increase in recent years. The number of *Species Hypotheses*, where species are based on sequence similarity, include molecular data from all biological samples, and exact taxonomic identification is often unavailable, grows linearly.

Fig. 45: Accumulation of sequences and full species names in INSD and UNITE system.



With the application of next generation sequencing techniques in recent years there has been an explosion of molecular data coming from various biological samples in ecological studies. These data (both sequence reads and associated analysis) are currently being deposited in Sequence Read Archive (SRA, http://www.ebi.ac.uk/ena/).

4.10.6 Recommendations for closing the gaps

The main gap of the genetic data in INSD databases is a lack of important metadata like locality. Not only country name but also georeferences should become mandatory fields for the uploading sequence data into INSD databases. This will enhance the quality and usability of the genetic data very much.

4.10.7 Literature

Schoch CL, Seifert KA, Huhndorf S et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences USA*, 109, 6241–6246.

Kõljalg U, Nilsson H, Abarenkov K et al. (2013). Towards a unified paradigm for sequence-based identification of Fungi. *Molecular Ecology*, 22(21), 5271 - 5277.

Nilsson RH, Ryberg M, Kristiansson E et al. (2006). Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS ONE*, 1(1), e59

Nilsson RH, Tedersoo L, Abarenkov K et al. (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys*, 4, 37-63. doi: 10.3897/mycokeys.4.3606

Lindahl BD, Nilsson RH, Tedersoo L et al. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, 199(1), 288 - 299.

Tedersoo L, Abarenkov K, Nilsson RH et al. (2011). Tidying up International Nucleotide Sequence Databases: ecological, geographical and sequence quality annotation of ITS sequences of mycorrhizal fungi. *PLoS ONE*, e24940, 1 - 7.

Schoch CL, Robbertse B, Robert V et al. (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. *Database (Oxford)*, 1-21. doi: 10.1093/database/bau061

4.11 GENERAL REVIEW OF GAPS IN EUROPEAN TAXONOMIC DATABASES: FAUNA (DATABASE FAUNA EUROPAEA)

4.11.1 Introduction - Short overview of the datasource

One important source of information regarding biodiversity is the information on the scientific names of species, subspecies and higher taxa. Only by having specific valid and unique identifiers like names, information of biodiversity can be assigned correctlyand names could be used as metadata to index biodiversity-related information (Patterson et al., 2010), e.g. occurence records with date, location and time to species. Here we outline the gaps of two taxonomic databases, Fauna Europaea (FaEu) for terrestrial and freshwater species and Euro+Med for plant species that list the formally described species of mainly European organisms.

Fauna Europaea (FaEu) is Europe's main zoological taxonomic index, making the scientific names of all European land and freshwater animals integrally available in one authoritative database via the Fauna Europaea web portal. Fauna Europaea covers about 240 000 taxon names, including 145 000 accepted (sub)species, brought together by a network of more than 400 specialists. Fauna Europaea is a unique (standard) reference on a European scale, serving as a scientific baseline for many users in science, government, industry, nature conservation, and education. Fauna Europaea is also part of PESI, the Pan-European Species directories Infrastructure which provide a robust infrastructure for the nomenclatural needs of European users for the different realms (freshwater, marine and terrestrial). As part of PESI, Fauna Europaea is selected as an INSPIRE directive (i.e. a formal standard) for the European taxonomic names. To ensure the collation of high quality data, more than 400 specialists, including 65 Group Coordinators are contracted. Advanced on-line and off-line tools for data import and data management were developed, and procedures for data validating applied, including a review process on the inclusiveness and quality of the data sets, taken care about a network of national Focal Points and other thematic partners, fully supported by the digital infrastructure. The Fauna Europaea index runs as a gateway serving the integration and sharing of biodiversity data, supporting major biodiversity informatics initiatives, like LifeWatch, EU BON, and GBIF.

4.11.2 Coverage of the dataset

Fauna Europaea is a project providing a web-based information infrastructure for the taxonomy of all European land and freshwater animals. The project started in March 2000 as a European Commission FP5 funded project, is producing an index of scientific names (including important synonyms) of all living European land and freshwater animals, their geographical distribution at country level (up to Ural, excluding Caucasus region), and some additional optional information. The coverage extends from the Azores in the West to parts of Russia in the East, and from Franz Josef Land in the North to Madeira in the South.

4.11.3 Outline of gaps and biases

Fragmentation of data sources

On a European scale, there is, through PESI, a common taxonomy existing for the freshwater, marine and terrestrial species. However, there are different approaches on a global perspective from national lists, to lists focusing on specific taxa. There is a need to maintain a unitary taxonomy (Godfray, 2002) that is up to date to avoid gaps and fragmentation in taxonomic knowledge.

Gap in taxonomy: Newly described species and cryptic species

Over the years, there is still a significant number of valid names given to new or existing species each year. As the figure shows, there is still a high number of species being detected or valid names are given to a taxon among the different phyla. The Fig. 46 shows, for European species, the number of species described per decade respectively number of species where a new valid name was assigned. There is particularly for the Arthropods a high number of species described consistently since the 1940ies (Fig. 46, please pay attention to the logarithmic scale of the y-axis). Looking at the aggregated numbers per decade for other species, there are also high rates of new species for Mollusc species, particularly with an increase in the period from 2001-2010. Contrary to that, there was a slight decrease for Annelidae. Some decline in described species could be also assigned to the problems taxonomy has since a couple of years, namely the lack of (human)resources and restricted funding opportunities.

Some basic numbers of Fauna Europaea:

- Number of species: 132’077
- Number subspecies: 14’191
- Number of synonyms (species): 41’556
- Number of synonyms (subspecies): 5’630
- References: 5’997

As the figure shows, there is a constant work needed in the realm of taxonomy, where many new species are, even in Europe, described every year. For the European species that are part of the FaEu database, the first descriptions started over 250 years ago (in the year 1557 by Carl Alexander Clerck) and stayed on a quite high rate also within the last 70 years as the figure shows. By this quite high number of new species described and detected it could be infered that there is still a high number of undescribed taxa among the different phyla in Europe. Thus, on a European scale, the extant biodiversity is still not adequately assessed due to unknown species or cryptic diversity (Bickford et al., 2007) nor described which also leads to an underestimation of biodiversity in general. Furthermore, not only the valid names are important for research purposes but also (undiscovered) synonyms.

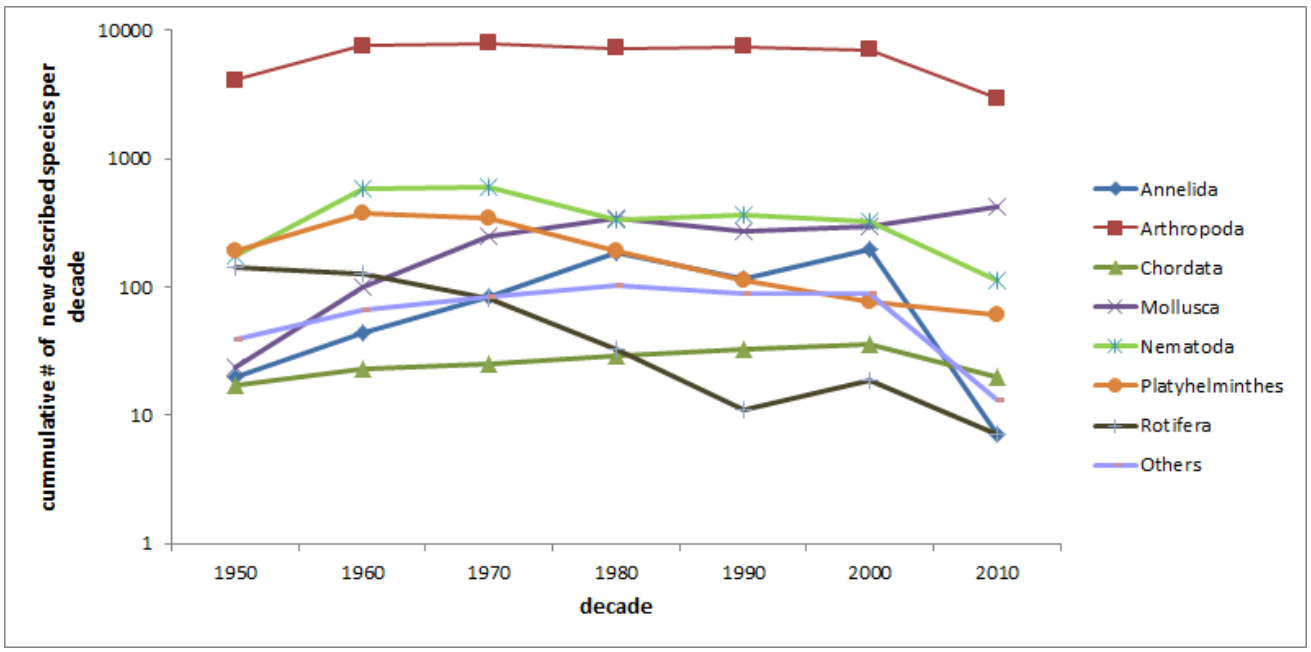


Fig. 46: Number of new taxon names (subspecies?) per decade from 1940 - 2010 (1950 includes the years from 1941-1950) according to different phyla. The number also includes newly described species or species with a new valid name.

For some species and groups, taxonomic data in Europe is quite complete, e.g. for the classes birds, mammals and reptiles the the number of new described species was quite low, ranging from 0-8 new described or named species in 20 years (1991-2010). However, also such relatively small numbers could mean a quite substantial effort in terms of taxonomy, as e.g. the 6 new species for amphibians represent 8% of Europe's amphibian species (see also Fig. 47). There are also orders on other taxonomic groups were still a high percentage of new valid species names were assigned. As the table shows (see Table 17), there is quite a number of orders where the new species within ~ 30 years (1981-2010) account for 30-44% of the species in general.

Table 17: the ten orders with the highest percentage of newly described species in the time period from 1981-2010. Only orders were counted with a overall species number of at least 50 species.

order	class	phylum	New species 1990-2010	All species until 2010	New Species / existing
Protura	Entognatha	Arthropoda	78	179	44%
Opisthopora	Oligochaeta	Annelida	173	472	37%
Chaetonotida	Chaetonotoidea	Gastrotricha	78	221	35%
Dorylaimida	Adenophorea	Nematoda	261	769	34%
Dictyoptera	Insecta	Arthropoda	68	202	34%
Ephemeroptera	Insecta	Arthropoda	113	339	33%
Neotaenioglossa	Gastropoda	Mollusca	303	915	33%
Pseudoscorpiones	Arachnida	Arthropoda	288	901	32%
Arhynchobdellida	Hirudinea	Annelida	14	47	30%
Mononchida	Adenophorea	Nematoda	32	108	30%

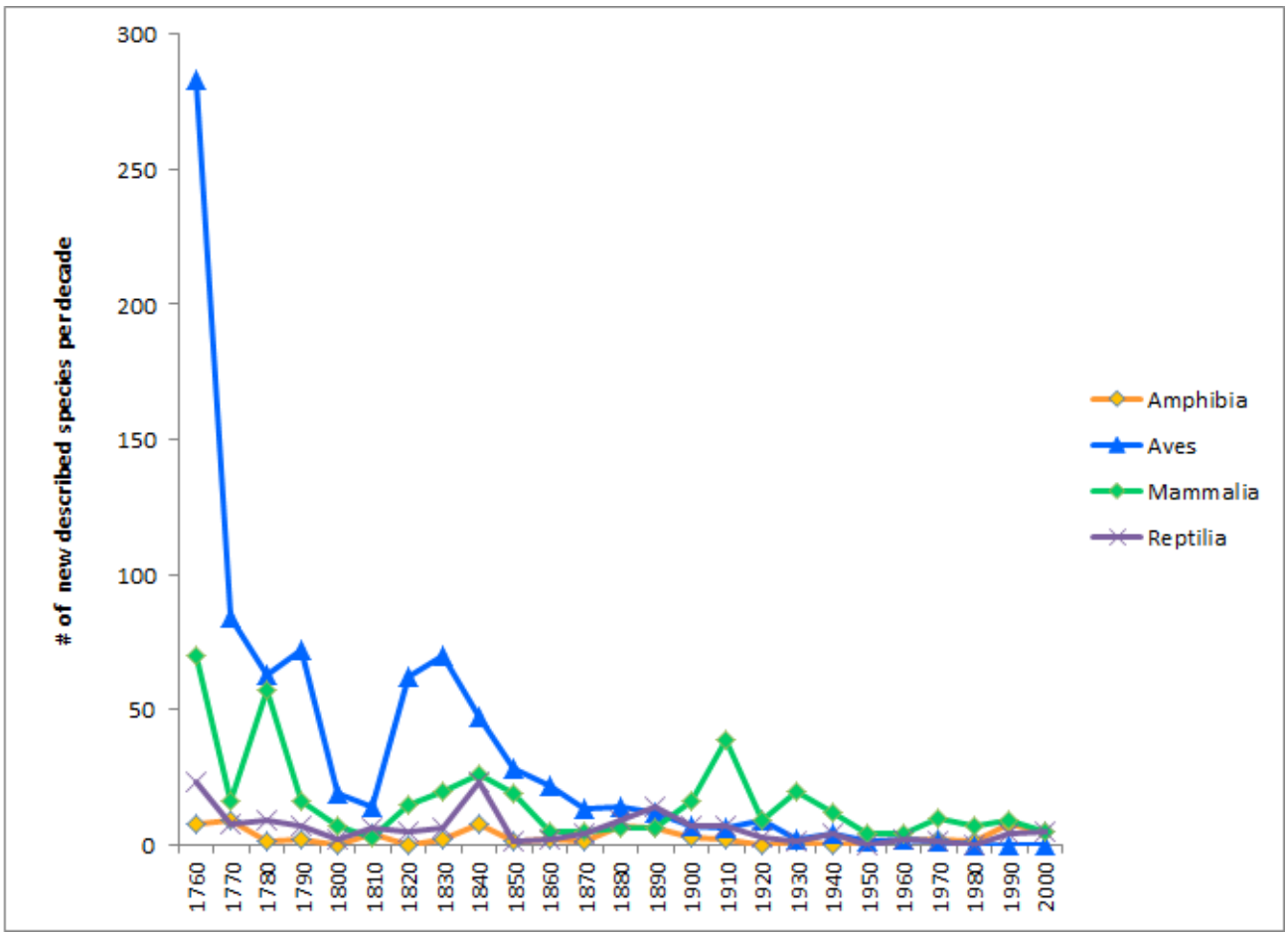


Fig. 47: Number of new taxon names per decade from 1758 - 2010 for four different classes of vertebrates.

In terms of number of species there was particularly a high number of species in arthropods and molluscs described (see Fig. 47). The figure reflects cummulative yearly increase of species since the year 2004, with an particular strong increase for arthropods throughout the years. For the mollusc species there was a particular strong increase in the years 2006-2009. As the figures shows exemplarily, there are still a large number of undetected or cryptic species even in Europe. Also an overview of new described species within 20 years (see Fig. 49) among the orders with the highest increase shows that there are still large gaps in our knowledge regarding taxonomy. Again, the orders with the highest numbers of new described species are part of the phyla of the molluscs and arthropods. The findings for European species are also found in other groups, for example in Raphidiidae, a family of snakeflies, where taxonomic efforts strongly increased since the 1970ies and still a lot of new species are still described. Today, 202 species of Raphidiidae are already described and it is estimated that there are possibly 50-60 species still to be discovered (Aspöck and Aspöck 2014).

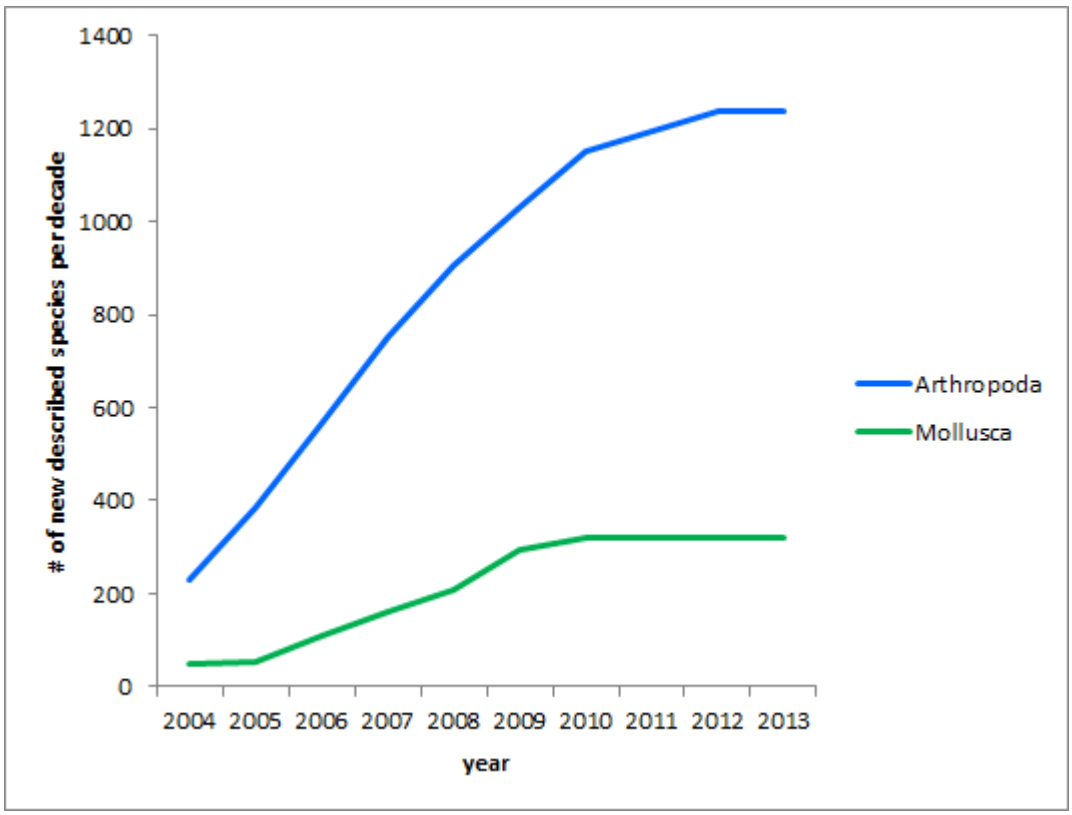


Fig. 48: Cummulativ number of new desribed species for European athropods and molluscs (2004-2013).

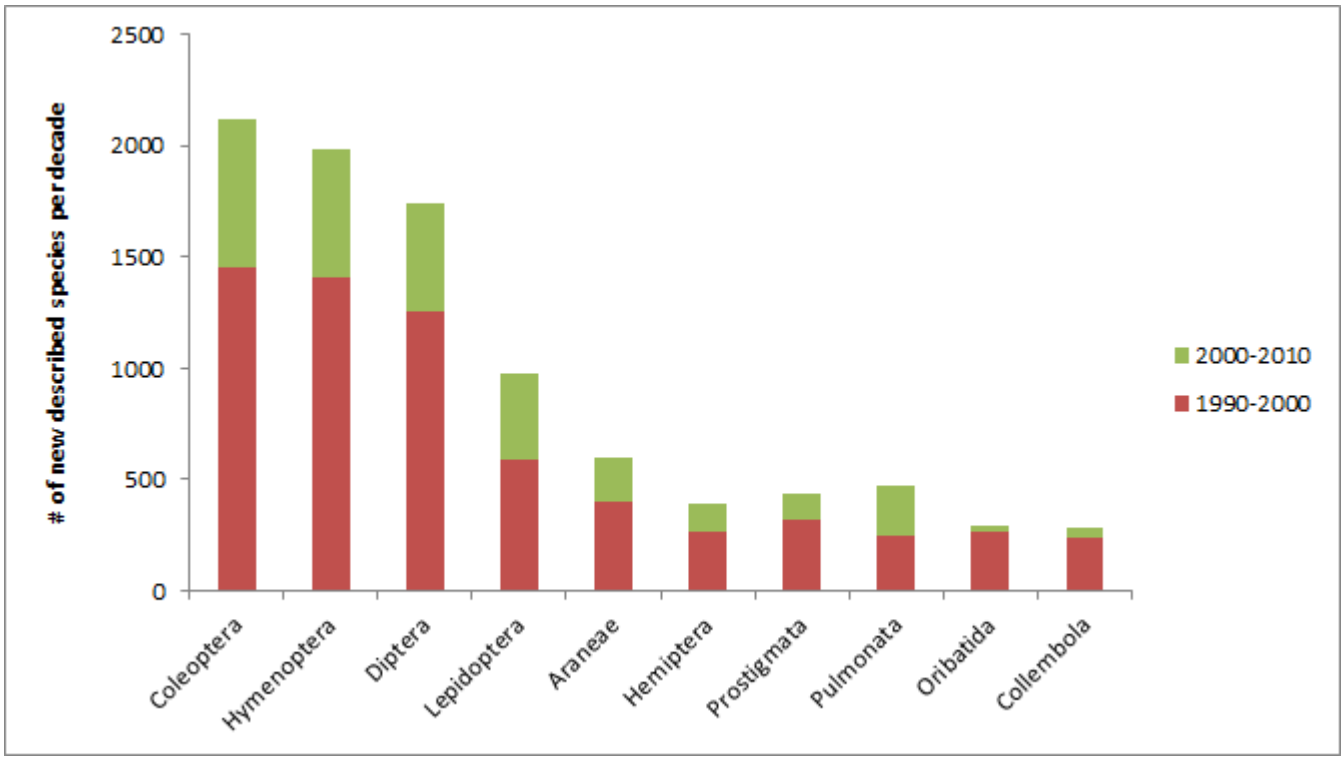


Fig. 49: Number of newly described species for different orders, illustrating the strong increase for the time period 1991-2000 and 2001-2010.

Gap in delay of published names

As Fig. 50 shows, there is also a delay between the publishing date of a valid taxon name and the date when the taxon name was ultimately published in an online database of taxonomic names. The delay can be quite substantial, for example only for 6% of the species the names were published within the same year. For 17% of the species name there was a delay of one year, 60% for a delay of 2-5 years and 17% for a delay of at least five years. As there is still a quite large delay in the publication of species names, ways should be found to close the existing and obvious gaps. A way to speed up the process to detect new species could be a combined approach of a routine to identify 'dark species' (for example with the help of barcoding projects, e.g. iBOL, the International Barcode of Life) and the possibility for a quick description of new species, for instance using the Pensoft tools.

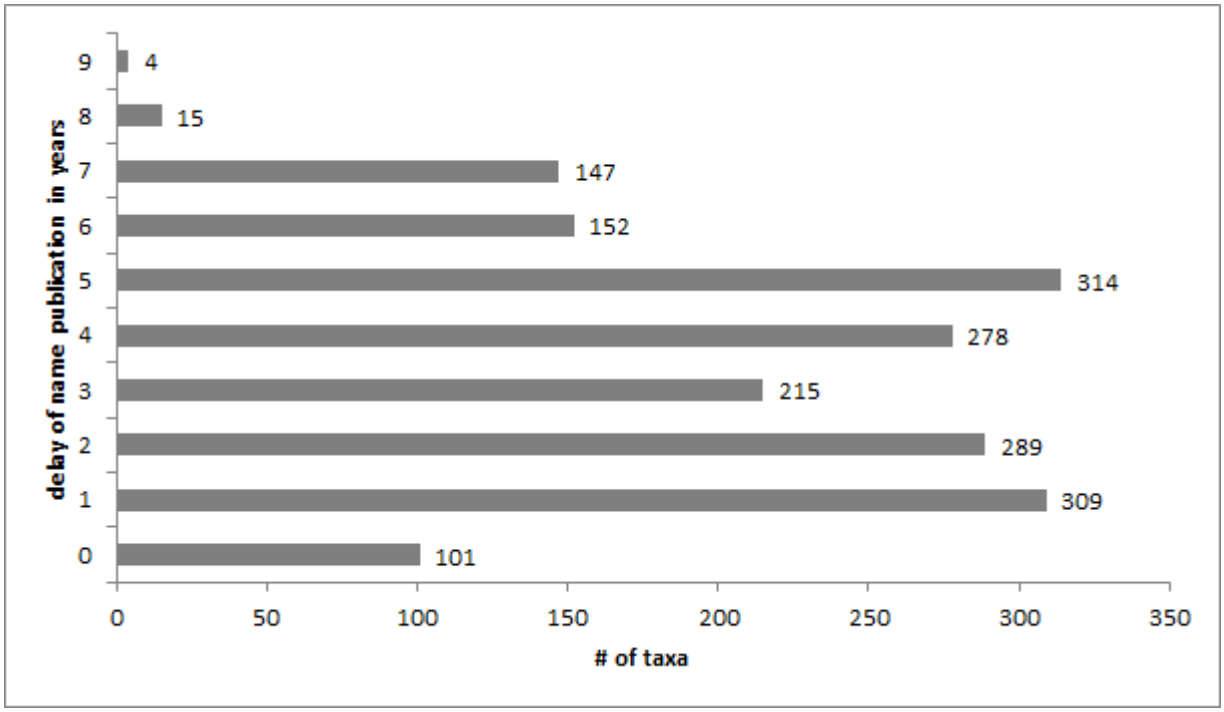


Fig. 50: Delay (in years) between publishing date of the taxon name and online publishing of the name in a database (here: in FaEu) for species published between 2004-2013.

4.11.4 Data accessibility

Data could be accessed and queried via the Homepage. Extracts of the taxonomic content of the database could be obtained via request for scientific use. It is planned to apply a CC-BY-SA license to the dataset.

4.11.5 Recommendations

- On a European scale, there is, through PESI, a common taxonomy existing for the freshwater, marine and terrestrial species and plants. There is a need to maintain a unitary taxonomy that is up to date to avoid gaps and fragmentation in taxonomic knowledge. Thus, extending and linking existing projects on species taxonomy has further to be enabled.
- Most databases for taxonomy have free access -however, some databases still allow only restricted use. For taxonomic data, all databases and projects would need to give free access to their data to update and homogenize current approaches.
- Also the human resource-side of taxonomy has to be tackled: Few funding opportunities for taxonomic work and overaging of taxonomists limit the ability
- Developing a routine to identify 'dark species' (in collaboration with barcoding projects) and the possibility for a quick description of new species, for instance using the Pensoft tools.
- Set up of a real innovative data management environment for expert networks. Resolving gaps should be an integrated part of the data management efforts, the results of validation routines proposed in a most user-friendly way.

4.11.6 Literature

Aspöck, H., Aspöck, U., 2014. Die Autoren der Taxa der rezenten Raphidiopteren (Insecta: Endopterygota). Entomologica Austriaca 21, 9-152.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends in ecology & evolution 22, 148-155.

Godfray, H.C.J., 2002. Challenges for taxonomy. Nature 417, 17-19.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L., Remsen, D.P., 2010. Names are key to the big new biology. Trends in ecology & evolution 25, 686-691.

4.12 GENERAL REVIEW OF GAPS IN EUROPEAN TAXONOMIC DATABASES: FLORA - VASCULAR PLANT SPECIES (EURO+MED)

4.12.1 Introduction - Short overview of the datasource

Euro+Med PlantBase provides an on-line database and information system for the vascular plants of Europe and the Mediterranean. It is a rich resource of information on the plant diversity of the Euro-Mediterranean region, including the Caucasus countries, which will be of use to a wide variety of users including professional biologists, agronomists, foresters, horticulturalists, conservationists, environmental planners and national and international organisations. The project has received initial backing from the European Commission for three years and additional funding within the likewise EU-funded project PESI (a Pan-European Species Directories Infrastructure).

Euro+Med Plantbase integrates and critically evaluates information from Flora Europaea, Med-Checklist, the Flora of Macaronesia, and from dozens of regional and national floras and checklists from the area, as well as from additional taxonomic and floristic literature. This is complemented by the European taxa of several families taken from the World Checklist of Selected Plant Families and of the Leguminosae from the International Legume Database and Information Service ILDIS. By April 2014 it provides access to 187 plant families, corresponding to ca. 92 % of the European flora of vascular plants.

4.12.2 Coverage of the dataset

Taxonomic groups covered by Euro+Med Plantbase are vascular plants, including ferns and fern allies of Europe, Transcaucasia, the circummediterranean countries and the Macaronesian Islands except Cabo Verde (see Fig. 51). In future, inclusion of other taxonomic groups such as lichens and bryophytes in Euro+Med Plantbase is envisaged.

The following data are included in the Euro+Med taxonomic core:

- The scientific name of each taxon;
- The standardized author citation;
- The place and date of publication;
- Common names in different languages, together with the literature reference;
- The basionym;
- Selected homotypic and heterotypic synonyms that have been used in the standard literature;
- Distribution according to published sources, together with the literature reference;
- Status of occurrence in the different E+M areas, together with the literature reference;
- Endemic to region/country/territory.

As to the status of occurrence, the following are included:

- indigenous (= native) species;
- naturalized aliens;
- frequently occurring casuals;
- introduced plants with unclear status of naturalization;
- presumably (regionally) extinct plants;
- plants that are conspicuously cultivated outdoors (including crops planted on a field-scale and forestry, street and roadside trees, but not commonly grown park and garden plants).

Fig.51: Map of the area covered by the Euro+Med PlantBase project

4.12.3 Outline of gaps and biases

For the gap analysis in data of vascular plants in the Euro+Mediterranean region, the Euro+Med Plantbase is continuously being evaluated and gaps are identified. The data, however, are not of the same quality for all taxonomic groups. We have to consider:

- full treatments with up-to-date taxonomy and full geographic coverage;
- treatments with gaps in up-dating the taxonomy (new names published in the last 15 years not fully integrated, new generic or other taxonomic concepts not considered);
- treatments with gaps in geographic coverage (literature from certain areas, e.g. Caucasus and Near East, not fully integrated);
- treatments with gaps in up-dating floristic and occurrence data (new, recently published datasources such as floras, checklists or other not fully integrated)
- treatments not yet edited by specialists (i.e., only raw data in the database, not yet available on-line)
- treatments imported from external sources (WCPS, World Checklist of Selected Plant Families e.g. for Labiatae; ILDIS, International Legume Database Information System). The inherent gaps within these datasources (e.g., no breakdown for Transcaucasian countries at country level as in E+M) will not be filled externally. Must be replaced with our own datasets as soon as feasible

4.12.4 Results Taxonomic gaps (Table 18):

	families	genera	species	subspecies (incl. nominate subspecies)	taxa ([species+subspecies] - nominate subspecies)	percent
Online accessible	187	2635	30174	12081	**39171**	**91,86**
Still to be edited and published	45	262	2676	1143	**3471**	**8,14**
Total	222	2897	32850	13222	**42642**	**100**

Concrete taxonomic gaps in Euro+Med Plantbase (Table 19): 45 families remain to be edited and published online. In order of size, these are (families with less than 10 taxa not listed):

Ranunculaceae 737 taxa

1. *Polygonaceae* 414 taxa
2. *Dipsacaceae* 364 taxa
3. *Cistaceae* 302 taxa
4. *Violaceae* 242 taxa
5. *Convolvulaceae* 152 taxa
6. *Onagraceae* 151 taxa
7. *Valerianaceae* 145 taxa
8. *Linaceae* 134 taxa
9. *Polygalaceae* 106 taxa
10. *Rutaceae* 71 taxa
11. *Amaranthaceae* 65 taxa
12. *Rhamnaceae* 63 taxa
13. *Caprifoliaceae* 63 taxa
14. *Tamaricaceae* 55 taxa
15. *Aceraceae* 46 taxa
16. *Cucurbitaceae* 44 taxa
17. *Berberidaceae* 37 taxa
18. *Aizoaceae* 31 taxa
19. *Callitrichaceae* 27 taxa
20. *Tiliaceae* 24 taxa
21. *Oxalidaceae* 23 taxa
22. *Cactaceae* 21 taxa
23. *Anacardiaceae* 19 taxa
24. *Polemoniaceae* 15 taxa
25. *Celastraceae* 15 taxa
26. *Pyrolaceae* 15 taxa
27. *Vitaceae* 14 taxa
28. *Frankeniaceae* 13 taxa

4.12.5 Geographical gaps:

	families	genera	species	subspecies (incl.nominate subspecies)	taxa ([species+subspecies] - nominate subspecies)	percent
Data complete	148	2078	23609	9698	30942	**72,56**
Data still missing (for transcaucasian countries and makaronesian islands)	74	819	9241	3524	11700	**27,44**

4.12.6 Data accessibility

The database is on-line and can be queried from the homepage, http://ww2.bgbm.org/EuroPlusMed/. Extracts of the taxonomic database content and subsets are under certain conditions available upon request for scientific use. The E+M Plantbase is currently licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported license (CC-By-SA-3.0 Unported). Search modalities will be improved when the migration from the current system to the EDIT Platform for Cybertaxonomy will have been completed. In particular, the new Euro+Med information system will be equipped with a capable service layer for machine to machine communication (http://cybertaxonomy.eu/cdmlib/rest-api.html). The RESTful service layer has granular methods giving access to all objects of the underlying data model as well as a set of streamlined service tailored for seamless integration in scientific workflows and portals.

4.12.7 Recommendations

Some of the identified gaps are currently being filled by specialist editors, who are responsible for certain groups of taxa. This process is sometimes very long and, given the limited time most researchers can devote to this task, unpredictable. Most responsible editors are not being payed to fulfil this task, nor is this task considered part of their obligations in the institutions where they work. While smaller gaps can be filled by work at the Euro+Med Plantbase secretariat, the editing and evaluation of larger groups such as Legumes (currently provided externally by ILDIS) or Labiatae (currently provided externally by WCPS) is a task for which additional funding should be looked for, as this will take several full person-years.

Cooperation with similar databases, which do have taxonomic and geographic overlap with Euro+Med Plantbase, is being sought to identify discrepancies and to match the taxonomy of those sources (e.g. Tela Botanica in France, African Plants Database in Genève, etc.).

4.13 GENERAL REVIEW OF GAPS IN EUROPEAN ENVIRONMENTAL TEST SITE DATA: LTER DATA

4.13.1 Introduction - Short overview of the datasource

The LTER-Europe network (Long-Term Ecosystem Research) is the European branch of the International LTER network (ILTER), a global network of research sites located in a wide array of ecosystems. With their long-term series of environmental observations, LTER sites can help with understanding environmental change across the globe. ILTER's focus is on site-based ecological and socio-economic research and monitoring (known as LTER and LTSER).

Traditional LTER sites cover an area of typically about 1-10 km², comprising mainly one habitat type and form of land use. Activities concentrate on small-scale ecosystem processes and structures (biogeochemistry, selected taxonomic groups, primary production, disturbances etc.). LTER-Europe distinguishes three classes of LTER sites with respect to infrastructure, comprehensiveness of ecosystem approach and age.

(1) Master sites are highly instrumented and permanently operating sites, featuring an ecosystem approach in terms of combining regular sampling (weekly as standard), permanent measurements and inventories at appropriate intervals across drivers and ecosystem compartments. Experimental approaches shall be existent or possible. All year access and power supply must be secured in order to enable, for example, measurement of climate data according to international standards. Other networks and related projects have been using this category of site (e.g. EMEP, CarboEurope, UNECE ICPs, national monitoring networks) and operation should be ongoing for at least 10 years.

(2) Regular sites should in principle comply with the description of Master sites, but differ in volume of instrumentation as well as multiple use and availability of long-term data across all ecosystem compartments and disciplines.

(3) Emerging & extensive sites are those having been recently established (3-5 years of observation) and currently being developed towards a higher category OR sites with a narrow, specific long-term monitoring and scientific focus, and therefore not following the full ecosystem approach (e.g. for reasons of limited considered spatial scale).

LTSER platforms typically consist of several LTER sites and additional partners. A minimum of five partners including non-scientific client groups and stakeholders (local decision makers, provincial administration, regional developers) have to agree on a common sociological, economic and natural-scientific research agenda supporting transdisciplinarity and participatory approaches. Platforms feature three functional layers: first, the physical infrastructure comprising *in situ* research sites, technical infrastructure, laboratories, monitoring networks, collections, museums, visitor centres and databases; second, a pro-active involvement of the research community on the regional, national and international level; and third, an integrative management serving as an interface between all above elements that should implement effective trans-disciplinary communication and participatory approaches.

Currently (6/2014), the LTER-Europe network consists of 24 national LTER networks comprising 438 LTER and LTSER sites, including a suite of long-term observations of many different

environmental variables, such as genetic data, species occurrence data, climate, habitat condition, ecological function and services, as well as socio-economic data.

With this site-based approach, LTER data typically shows a high heterogeneity across Europe. Methods are not fully harmonised and sampling is not executed according to a preconceived strategic plan. Data currently mostly remains with the individual sites and platforms, but as an important step to exchange information and data, a new web based tool to collect and manage metadata not only for sites, but also for persons and dataset was developed: the LTER Europe DRUPAL Ecological Information System (DEIMS; http://data.lter-europe.net/deims/). DEIMS provides a view on metadata within the LTER network for the whole community and partly holds actual datasets or provides links to data repository systems of the individual sites.

4.13.2 Coverage of the dataset

The distribution of LT(S)ER sites covers the entire area of Europe (Fig. 52). However, sites are most densely represented in central Europe and the United Kingdom. Freshwater (lakes and rivers), forests (mixed, deciduous and evergreen) and alpine areas are the most well represented biomes in the LT(S)ER network (Fig. 53). Looking from a biogeographic region perspective, continental, Atlantic, Mediterranean, and alpine regions are the best represented regions (Fig. 54), while nemoral, boreal and northern alpine regions, compared to their spatial extent, are underrepresented in the network.

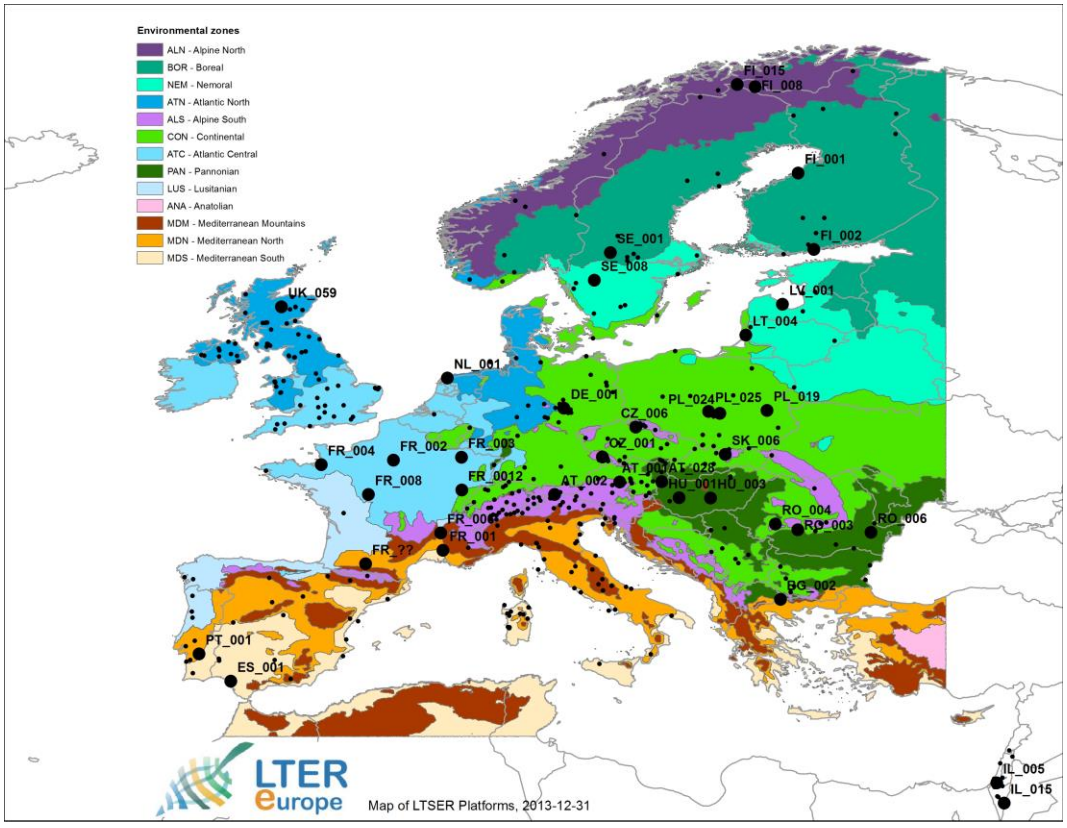


Fig. 52: Map showing location of Long-Term Ecosystem Research (LTER, small points) and Long-Term Socio-ecological Research (LTSER, large points) sites across Europe.

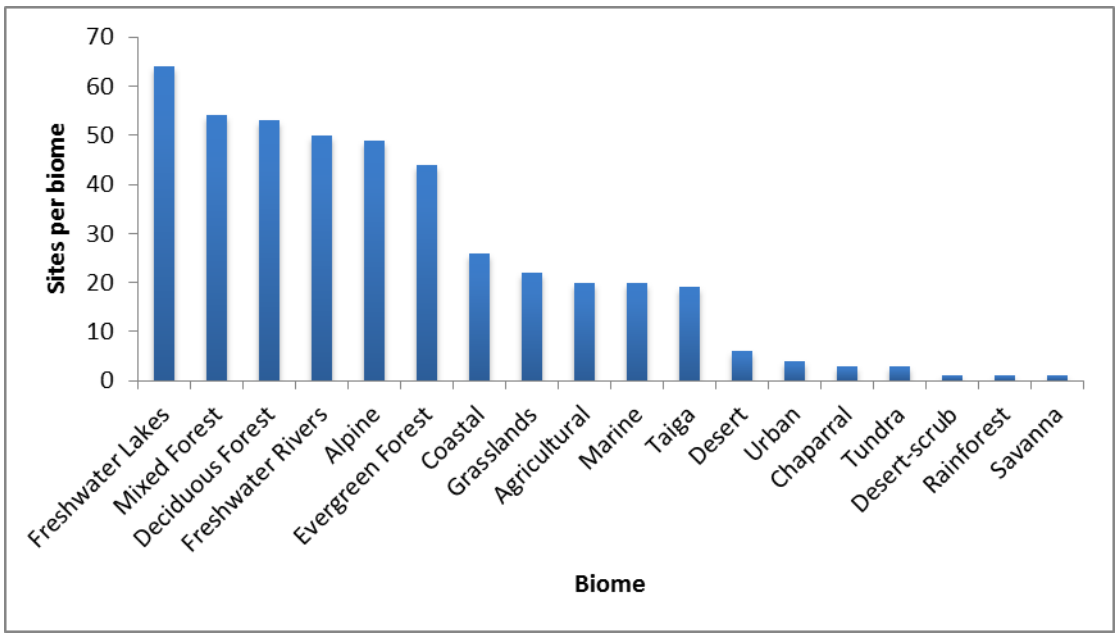


Fig. 53: Number of LT(S)ER sites represented per biome.

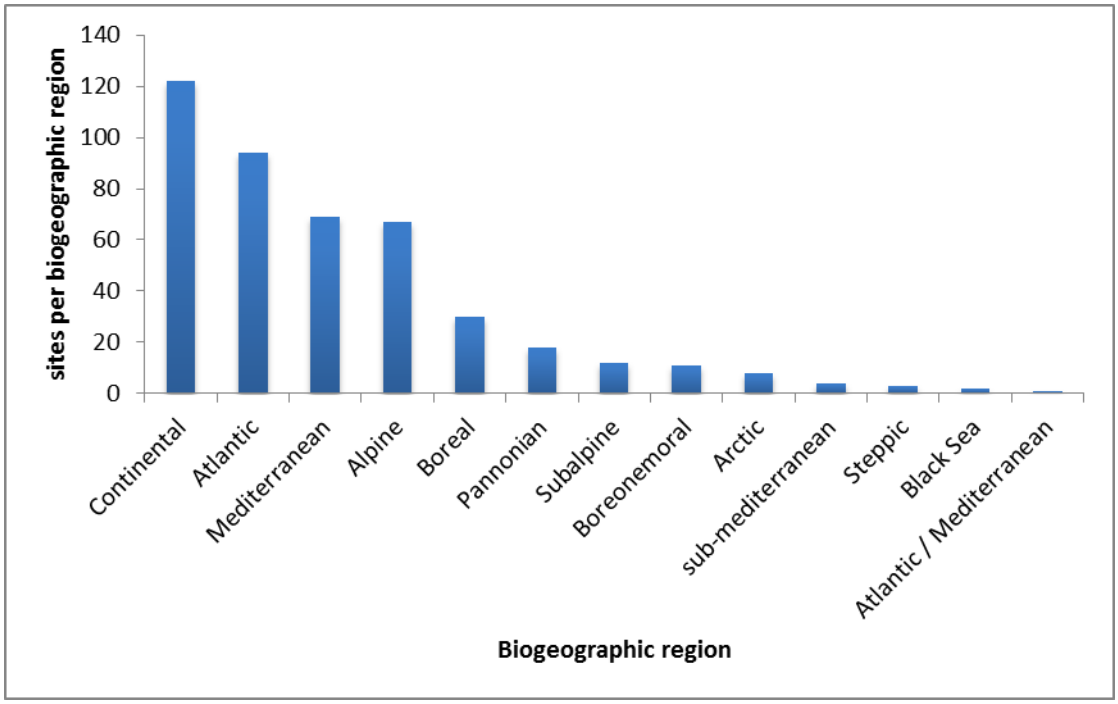


Fig. 54: Number of LT(S)ER sites represented per biogeographic region.

In terms of the different domains, terrestrial environments make up the majority of LT(S)ER sites (Fig. 56).

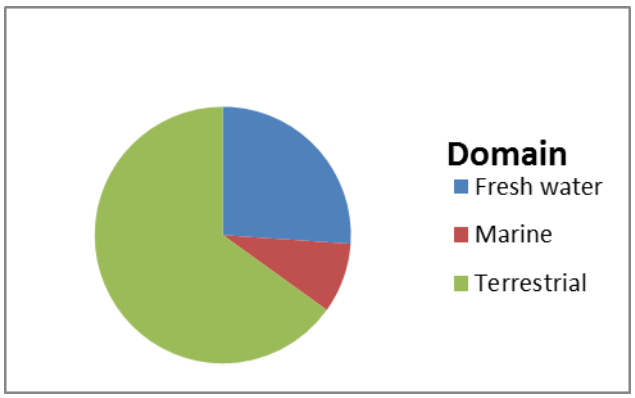


Fig. 55: Relative representation of the three domains, freshwater, marine and terrestrial in the LT(S)ER network.

The United Kingdom and Italy have the largest number of LT(S)ER sites compared to the rest of Europe. In the Netherlands and Belgium, national LTER networks have been started only recently and thus consist of fewest sites (Fig. 56).

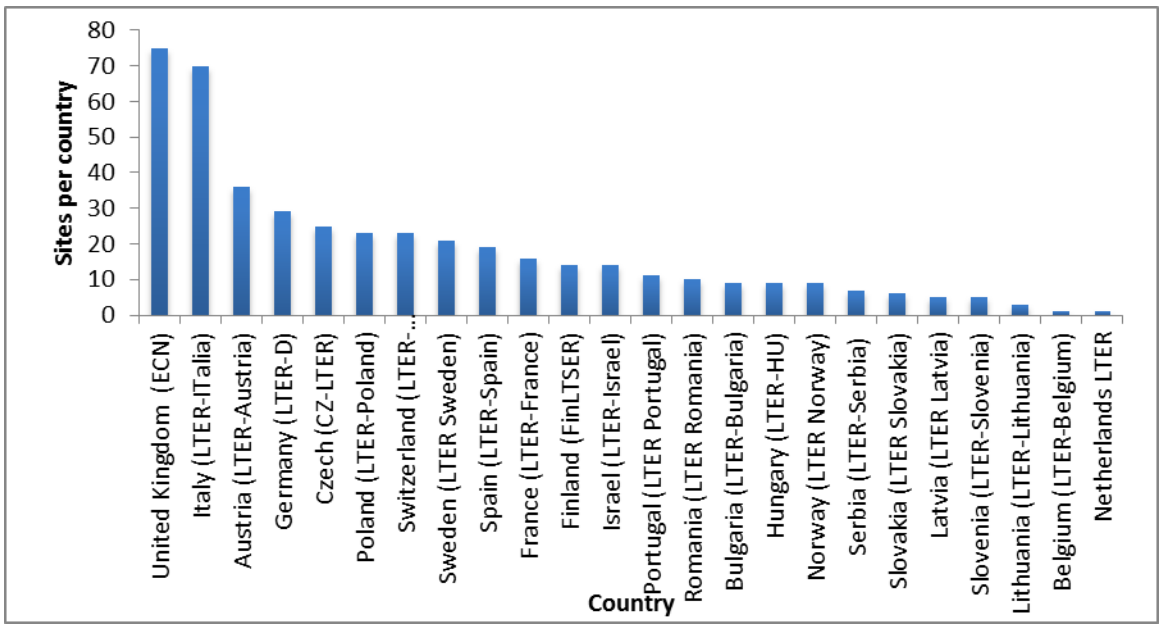


Fig. 56: Number of LTER or LTSER sites present in each country of Europe.

In terms of the length of time that these sites have been operating, the range is expansive (Fig. 57). There are a significant number of sites that have been operating over a long period. In fact, 48 sites have been operating for 50 or more years and 12 for over 100 years (Fig. 57). Of these older sites, the mean number of parameters assessed was 9.8 ± 0.9 (S.E.) and the number of research topics 20.9 ± 2.1. The oldest of these are the Czech glacial lakes (143 years), the Dutch Wadden Sea Area LTSER (142 years), and the Sonnblick Observatory in Austria (128 years).

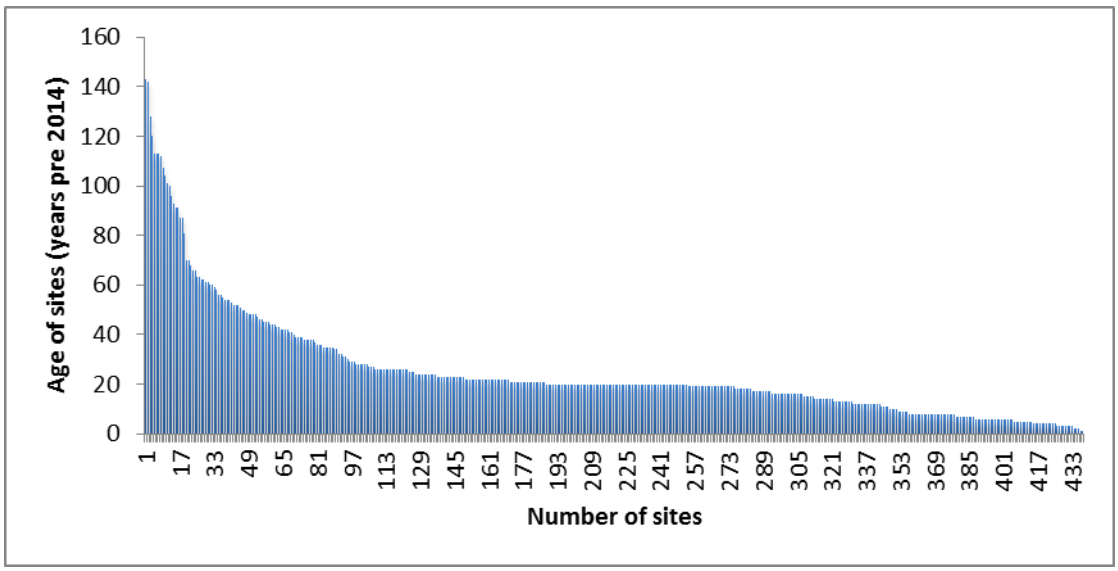


Fig. 57: Frequency plot of LT(S)ER site age ranked in order of oldest to youngest. Site age is represented by the number of years pre-2014 in which a site was initiated.

A large range of research topics/subjects are carried out at these LT(S)ER sites (Fig. 58). The most common of which are vascular plants, climate change, species composition, habitat and ecosystem structure, pollution effects, biogeochemical cycles, and hydrology (Fig. 58). Research topic density at sites, which might be regarded as an indicator for site complexity, is relatively well distributed across Europe (Fig. 59).

In terms of the parameters measured, similar patterns are observed to the research topic (Fig. 60). Biodiversity of plants is the most commonly measured parameter across all sites, followed by meteorology and climate, habitat and ecosystem structure, and hydrology (Fig. 60). Some evidence suggests that density of parameters measured tends to be higher in general in the central European sites, but the pattern is highly variable (Fig. 61).

Reassessing these trends solely for sites that have been operating for 50 years or more highlighted a similar pattern, with ecosystem structure, climate change, hydrology and vascular plants among the most common (Fig. 62). Likewise, sites 50 years or older tend to assess similar parameters to the younger sites in general, although phenology became relatively more commonly measured in these older datasets (Fig. 63). Furthermore, there does not appear to be any trend in terms of the size of plots across Europe (Fig. 64).

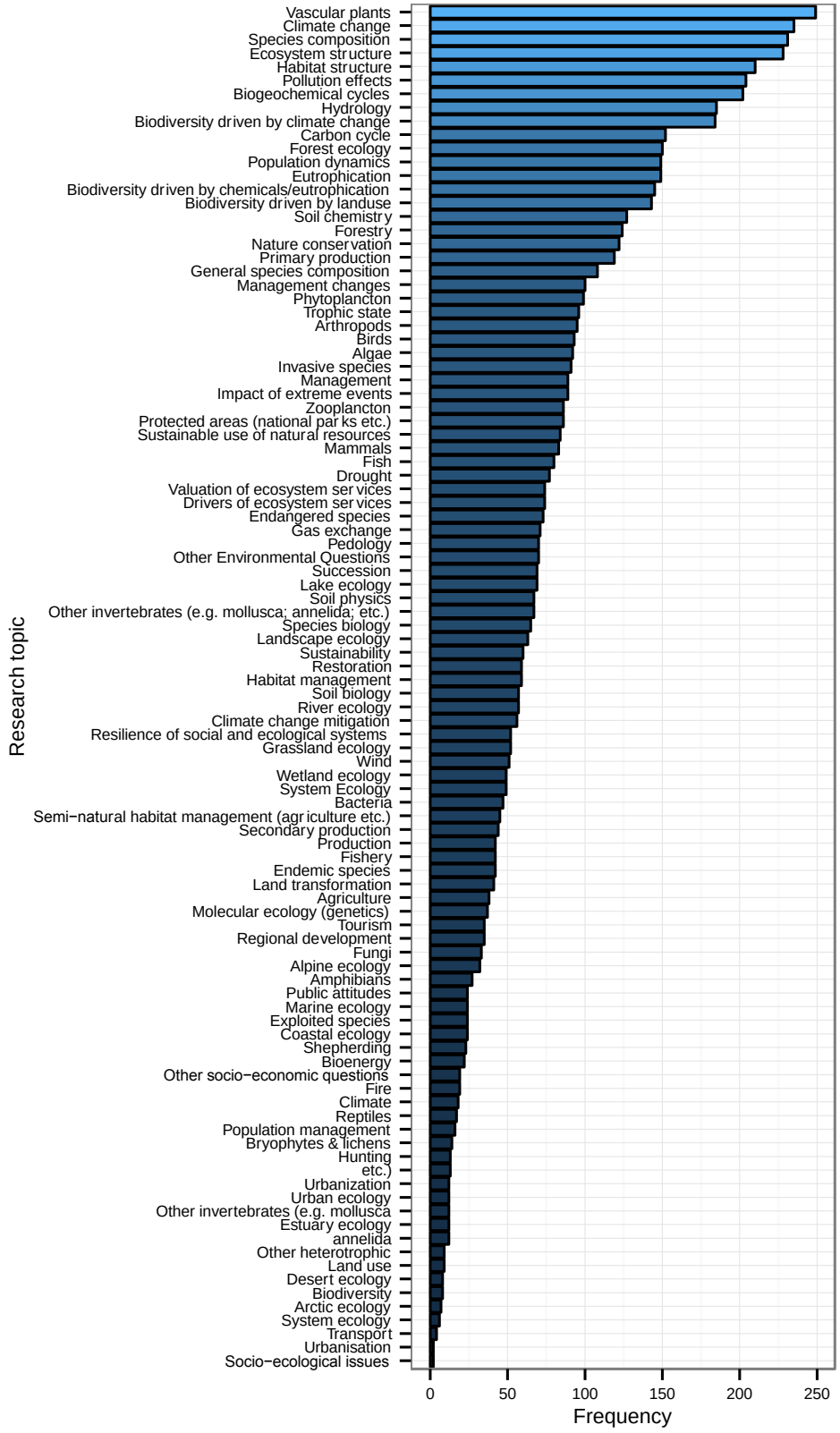


Fig. 58: Frequency of LT(S)ER sites including each of the various research topics.

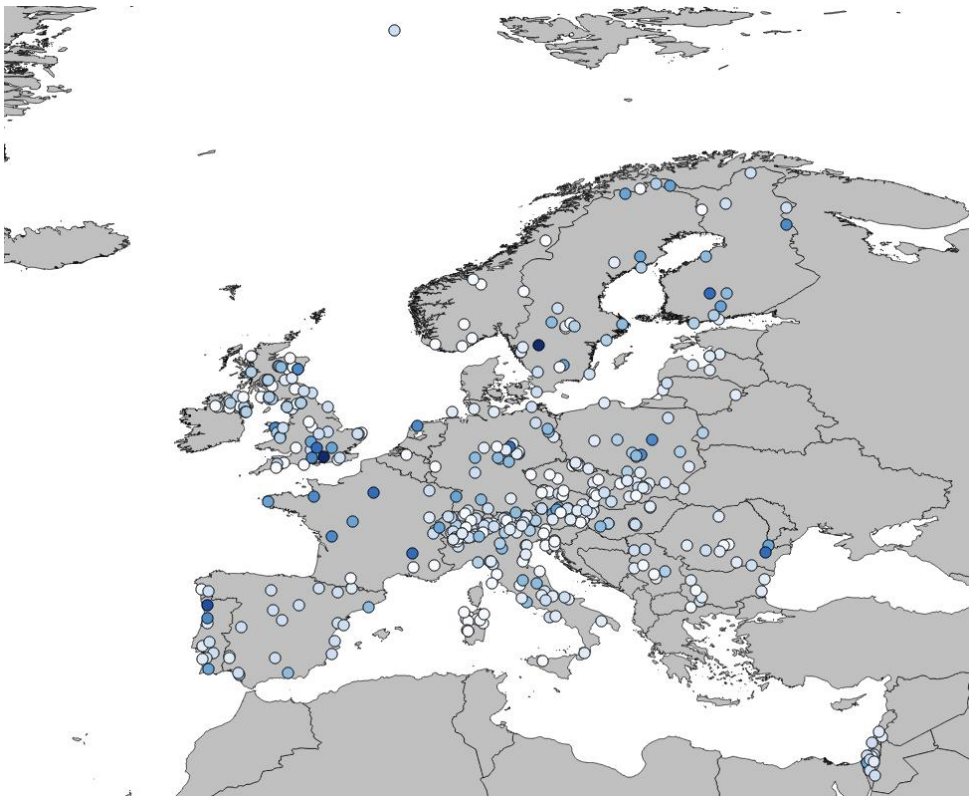


Fig. 59: Map showing the spatial coverage of the number of project objectives / research topics included at each LT(S)ER site. Darker color represents more project objectives (range = 0-72).

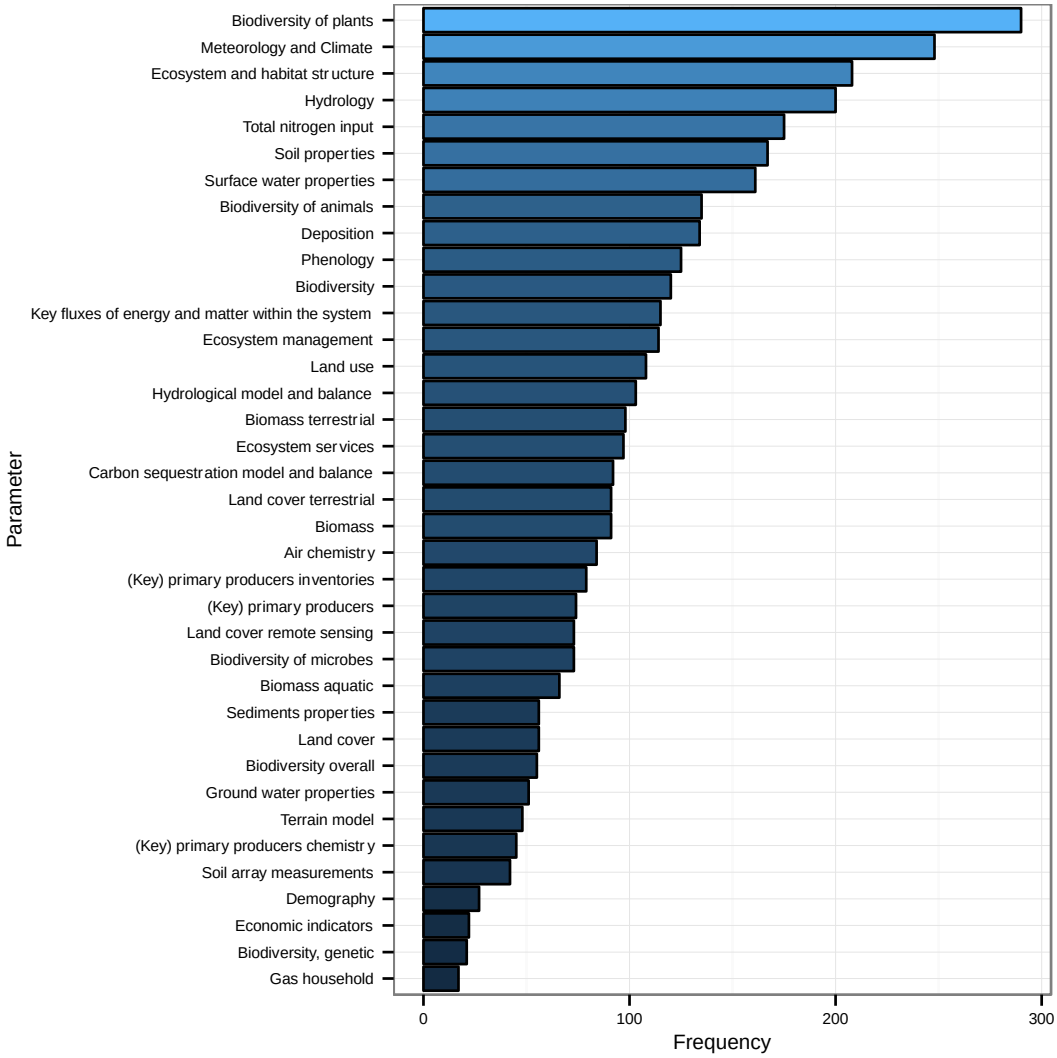


Fig. 60: Frequency of LT(S)ER sites including each of the various measured parameters.

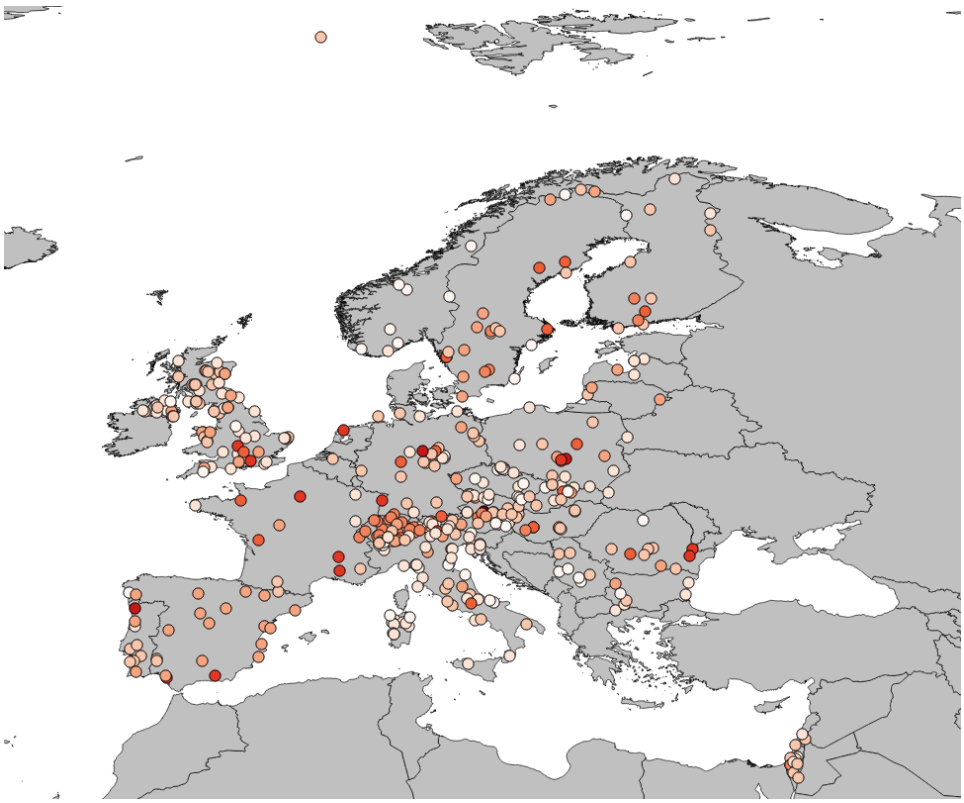


Fig. 61: Map showing the spatial coverage of the number of parameters measured at each LT(S)ER site. Darker color represents more parameters measured (range = 0-33).

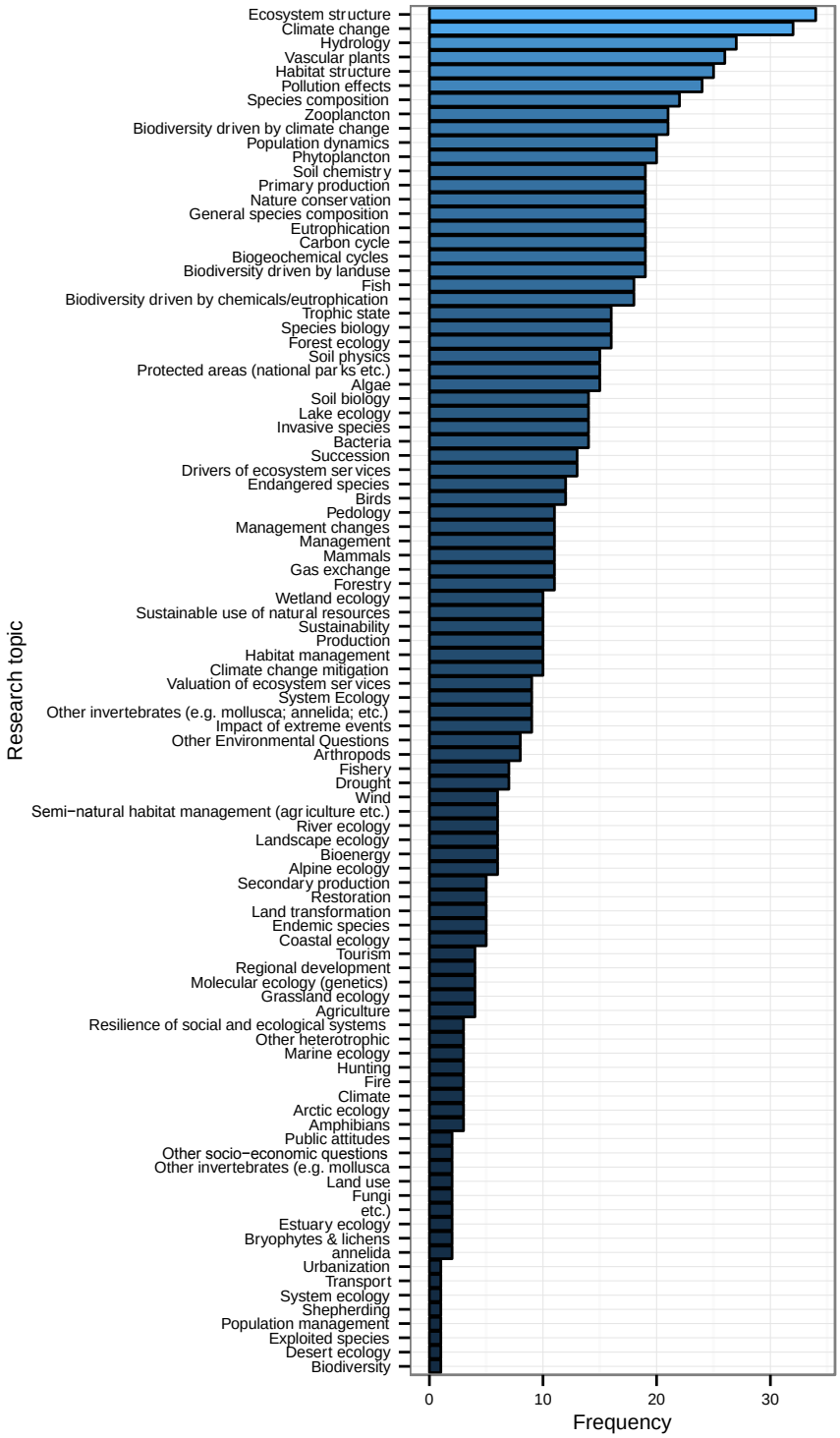


Fig. 62: Frequency of LT(S)ER sites including each of the various research topics for sites that have been operating for at least 50 years.

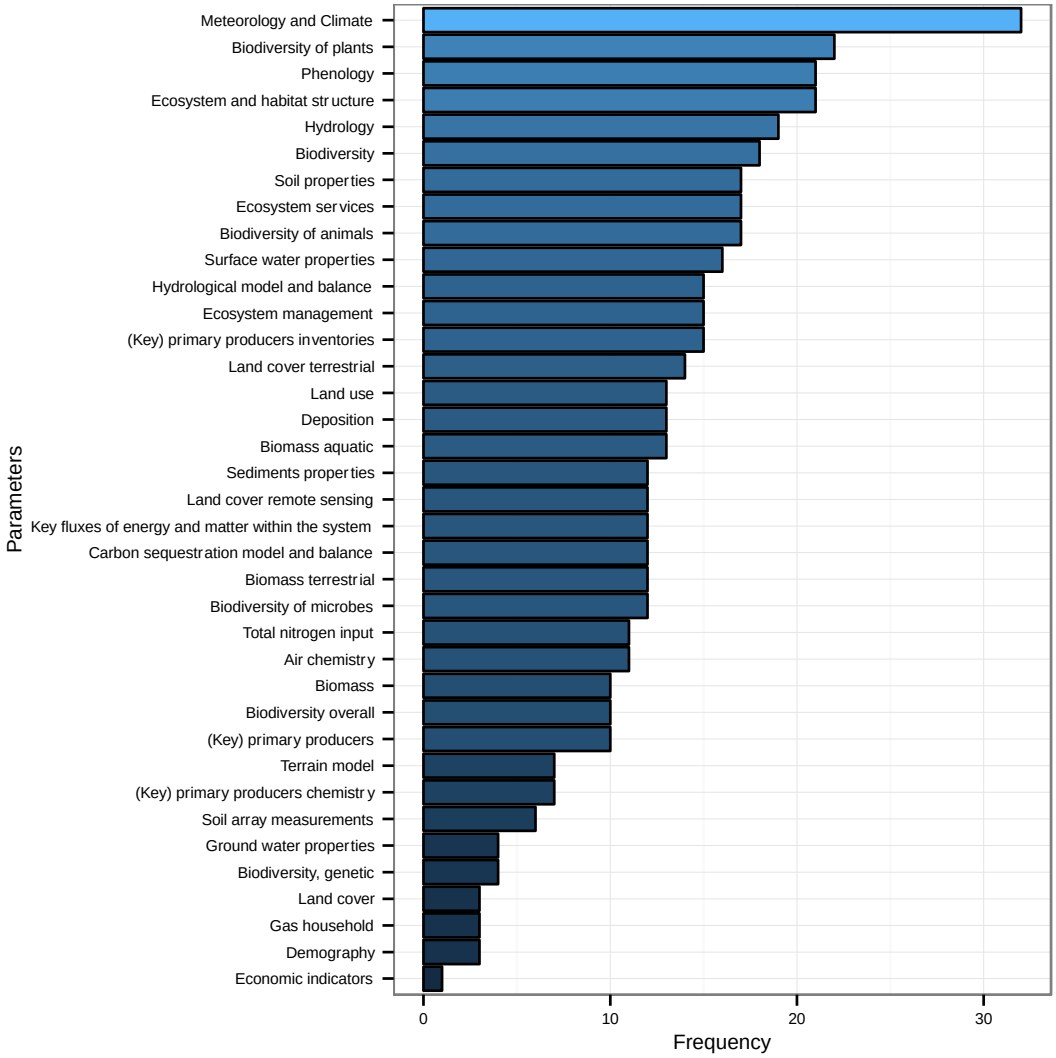


Fig. 63: Frequency of LT(S)ER sites including each of the various measured parameters for sites that have been operating for at least 50 years.

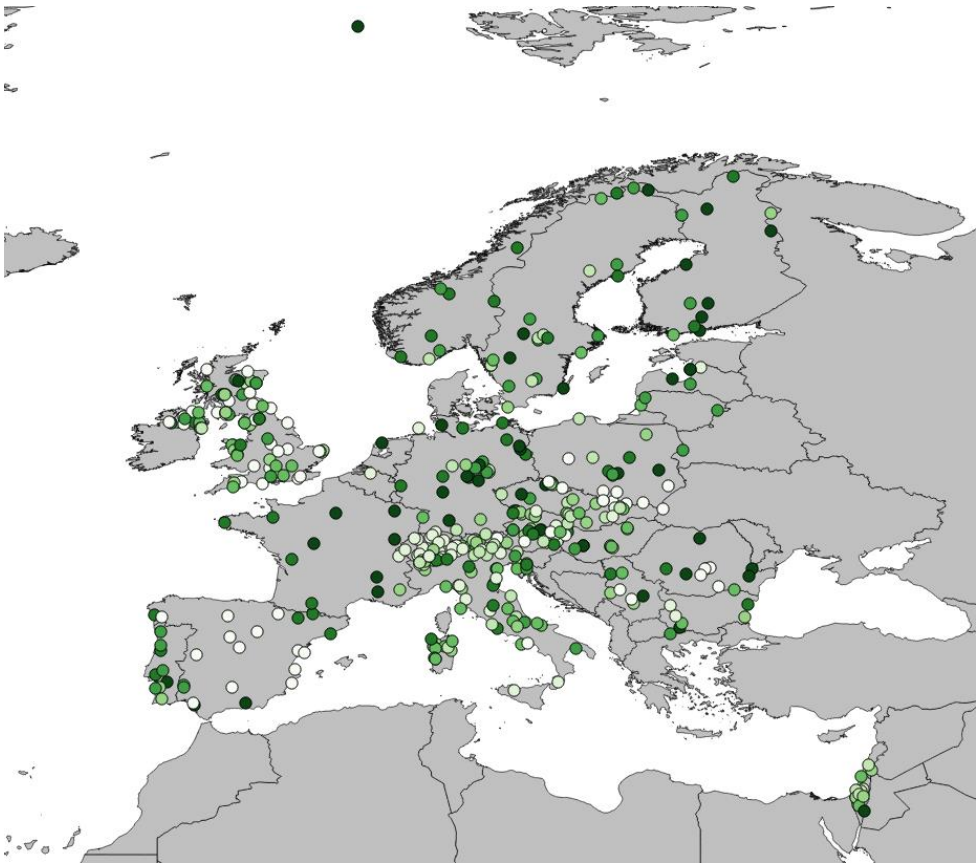


Fig. 64: Map of LT(S)ER size in hectares. Darker color represents larger sites (range = 0-40,000,000 ha).

4.13.3 Outline of gaps and biases (e.g. spatial, taxonomic, temporal) and data quality

- As previously stated, this network consists heavily of terrestrial sites. In its current form, marine environments are the most under-represented domain.
- Arctic, sub-mediterranean, steppic, Black Sea and Atlantic/Mediterranean each make up less than 2% of sites, but this also represents the spatial coverage of these areas across Europe.
- Freshwater (lakes and rivers), and mixed, evergreen and deciduous forest, as well as alpine environments make up a large proportion of the sites.
- Chapparal, desert and desert-scrub, rainforest, savanna, tundra and urban make up a small proportion of the sites. Again this is to be expected for all but urban. Urban areas represent a clear gap in the LTER network and thus should be a focal area in the future.

Focusing on specific groups, we can see, for example, that the measured parameter "biodiversity of plants" is well represented across Europe being included at 290 sites and no clear gap is evident in its spatial coverage (Fig. 65). This is also represented when looking at the research topic of interest "vascular plants", which is included at 249 sites (Fig. 66).

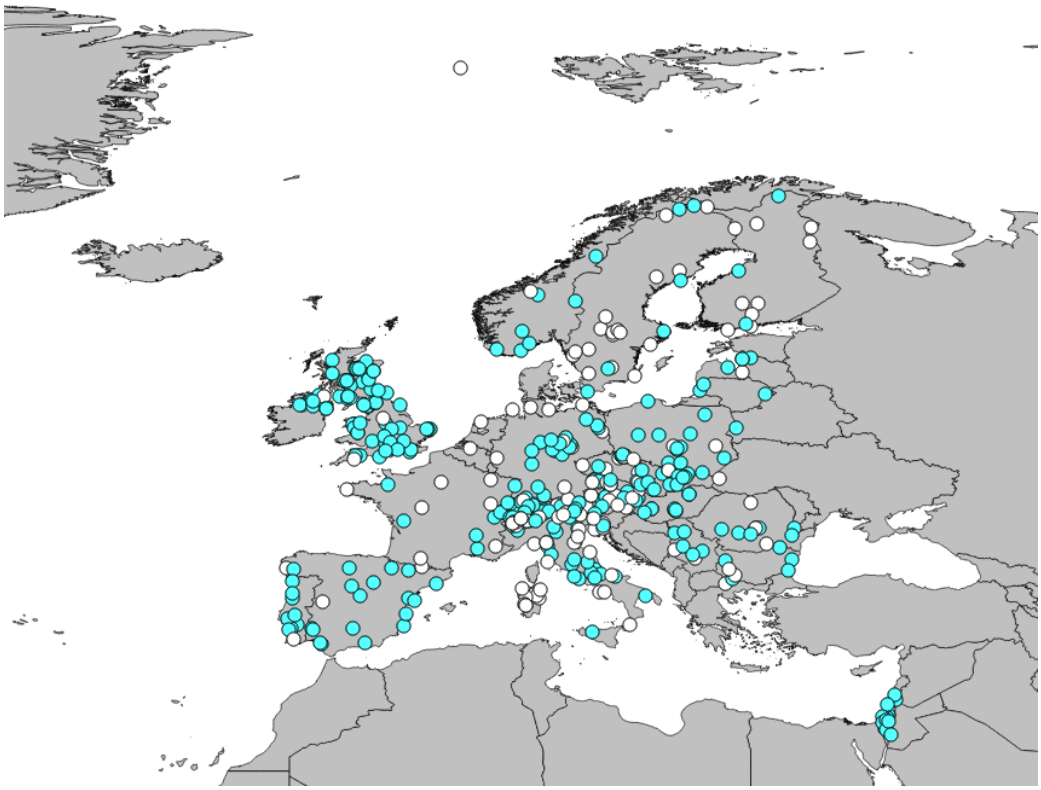


Fig. 65: Map illustrating spatial coverage of LT(S)ER sites that include the measured parameter "biodiversity of plants". 290 sites in total include this parameter. Colored points represent inclusion of parameter, white indicate not measured.

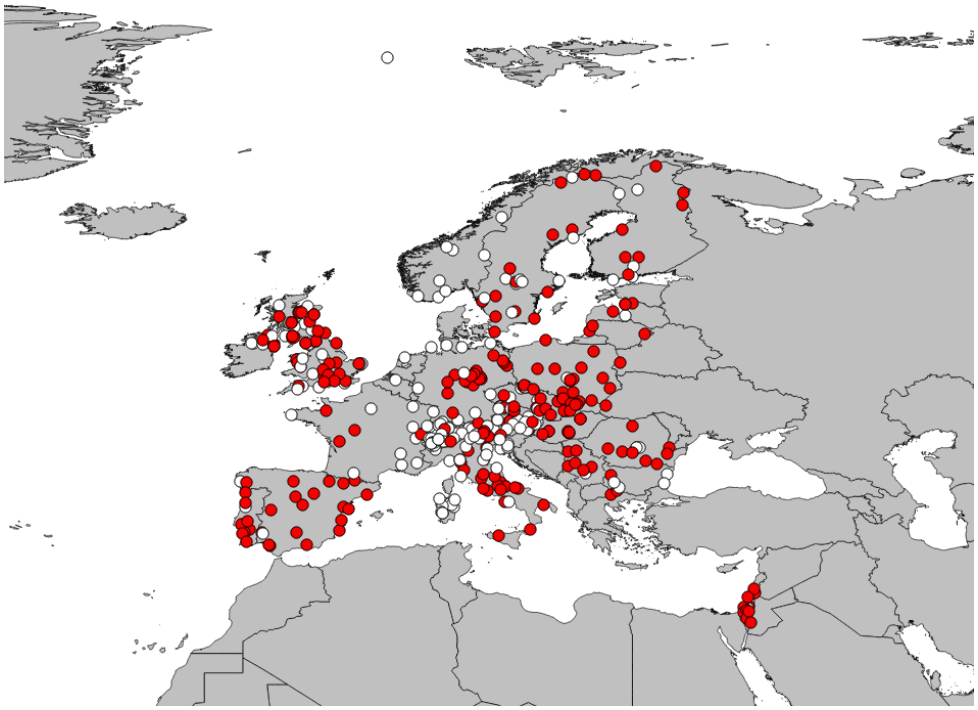


Fig. 66: Map illustrating spatial coverage of LT(S)ER sites that include the research topic “vascular plants”. 249 sites in total include this topic. Colored points represent a focus on this research topic and white indicate not.

Birds are less well represented in terms of the number of sites that include them as a research topic (93 sites; Fig. 68). Nonetheless, a relatively even spatial distribution of sites is present across Europe (Fig. 68).

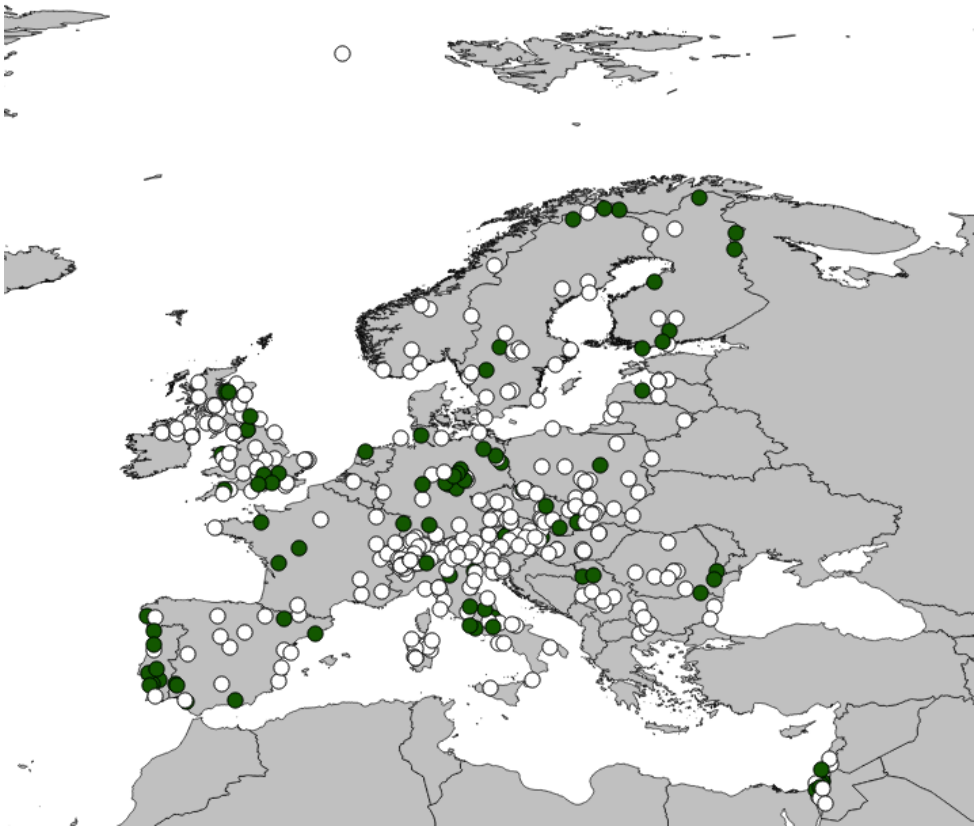


Fig. 67: Map illustrating spatial coverage of LT(S)ER sites that include the research topic "birds". Ninety-three sites in total include this topic. Colored points represent a focus on this research topic and white indicate not.

4.13.4 Data accessibility

Metadata are collected and managed for LTER-Europe via the Drupal Ecological Information Management System (DEIMS; http://data.lter-europe.net/deims/ or http://sp7.irea.cnr.it/wp4/az2/geoportal/public/). Access to metadata is freely availably, however, raw data availability varies with dataset (Table 20). Most are free upon request but some are restricted (Table 20). This also differs between users. Access to data mostly relies on contacting the data owners, which is more often through offline contact than online (Table 21). Almost half the data are stored centrally, but the remainder is distributed either within one or several institutions (Table 22).

Table 20: Accessibility of data for individual datasets. The number of sites with different data access policies are given in the table.

	Data Access for Administration	**Data Access for Public**	**Data Access for Research**
Free	40	26	28
Free upon request	253	253	287
Restricted	139	138	117
Other / Not specified	6	21	6

Table 21: Data request format for the individual datasets.

Access request format	Number
Offline (Mail or Telephone)	280
Online (Reference for access)	156

Table 22: Data storage location for each of the individual datasets.

Storage location	Number
Central	210
Distributed within institution	119
Distributed over multiple institutions	100
Other	9

4.13.5 Trends in accumulation of occurrence data / integration of historical data

Data accumulation in the LTER database is not known (not relevant, respectively). As reference points, many sites have historical data to relate to, but systematic integration in actual time series of observations with consistent methods is difficult.

4.13.6 General recommendations and prioritization for closing the gaps

General recommendations and prioritisations for closing existing gaps is difficult, as data have to be continuously monitored, processed and stored, involving also a strong financial commitment from the individual contributing institution. As no money is distributed through the LTER network, but funding has to be secured from the individual sites, there is only very limited influence of the network to adapt local sampling and data processing and sharing schemes. Nevertheless, identification of gaps in the network is useful, as with this knowledge, additional institutions involved in long-term monitoring and research programs can be contacted and invited to join.

Other recommendations:

- Centralising data or other measures to provide easier access to all data would benefit users seeking to use data.
- Enlarge the LTER network with a specific focus on underrepresented areas and topics
- Offer technical solutions to safe long-term data storage, as many sites are affected by non-permanent/uncertain funding. This harbors the risk that data series disappear if programs are stopped.
- Create incentives for sites to make data available online.
- Encourage data owners / create incentives for free access to data being allowed.
- As many LTER sites generate data that are also part of other monitoring schemes (e.g. ICP), a powerful data format conversion tool that allows exporting data from the LTER database in all important formats that are used in other communities would be a big incentive for data owners to upload their data to DEIMS.
- Especially large sites have their own data repositories. Creation of standard interfaces through which data is retrievable from the data owner through.
- DEIMS would be helpful.

5 ANNEXES

5.1 ANNEX 1: HIGH LEVEL QUESTIONS ON BIODIVERSITY AND, AS A SUBSET, THE TARGET HIGH-LEVEL QUESTIONS FOR THE EU BON GAP ANALYSIS

A List of high level questions

1. **Species and Habitats in Europe**
   - 1.1 What is the current status and trends regarding conservation (abundance/ distribution) for species of the habitats directive/birds directive?
   - 1.2 Is the fragmentation of species populations in Europe declining or increasing – what is the effectiveness of different measures to halt/decrease fragmentation and what are the effects of the fragmentation on species populations?
   - 1.3 What are the positive/negative impacts of subsidies - like EU funded projects, plans and programmes (conservation, land-use, industry) - on European biodiversity?
   - 1.4 Is the genetic diversity of cultivated plants/domesticated animals and their wild relatives sustained and what are the trends?
   - 1.5 Is the biodiversity loss, in particular the extinction of known European threatened species (or rare species/iconic/phylogenetically distinct species), stopped? Do we have sufficient knowledge regarding taxonomic information of biodiversity, i.e. species names and their number?
   - 1.6 How does the protection of EU's priority species (habitat and bird directive) also serve the protection of other species and ecosystem services?
   - 1.7 What novel approaches for the mapping and modeling of biodiversity and ecosystem services can be developed to overcome limitations of current models – do we have sufficient knowledge for developing these models?
   - 1.8 How does the protection of EU's priority species (habitat and bird directive) also serve the protection of other species and ecosystem services?

2. **Ecosystems, biodiversity and their functions**
   - 2.1 What is the relationship between species diversity/ abundance and ecosystem functions and services like provisioning services (food, fresh water etc.) regulating services (climate, water, pollination) or cultural services (recreation, educational, cultural) for different ecosystems?
     - Please specify here for which ecosystem service the data could be used
   - 2.2 How is biodiversity and intact ecosystems linked to human health and how can biodiversity improve human health? For e.g., how are biodiversity and intact ecosystems linked to the evolution and the spread of pathogens (virus, bacteria, priones)?
   - 2.3 Can biodiversity increase resilience of ecosystems regarding drivers of change such as climate change, pollution, overexploitation etc.?
     - Please specify here for which driver / ecosystem service the data could be used

3. Ecosystems and their services

3.1 What is the current status of European ecosystems and their essential services (mapping/assessing of services)?

3.2 What are the trends and scenarios for future ecosystem services and ecosystem functions (provisioning, regulating and cultural services) in Europe?

3.3 How do land degradation and biodiversity loss /loss of ecosystem services and functions interact?

3.4 What are the priorities for ecosystem restoration – where should restoration take place? Where could restoration help in terms of risk-reduction regarding natural disasters (floodings, erosion, avalanches)?

3.5 How could ecosystem resilience be improved through restoration and conservation, what are the ecosystem based adaptation capabilities?

4. Sustainable Land-Use and Use of Freshwater Systems and Oceans

4.1 Is there a measurable improvement in the conservation status of species or their populations and habitats due to agro-environmental measures/sustainable forest management plans/ managed zones of biosphere reserves, on a European / regional / local perspective?

4.2 Is there an increasing fraction of ecosystems used sustainably, particularly in regions where vulnerable species or habitats are located?

5. Protected Areas

5.1 How does the preservation of European protected areas positively affect biodiversity (national parks, biosphere reserves, marine protected areas but also urban nature reserves)?

5.2 What is the state of marine and terrestrial protected areas – are they effectively managed and secured? How is the cost-effectiveness varying among European conservation programs?

5.3 How can European protected areas be designed to increase carbon storage benefits and mitigate climate change impacts?

6. Drivers of change

6.1 How can the most important drivers of change regarding biodiversity be identified and ranked?

6.2 How do global change drivers (climate change, land use change/habitat destruction / overexploitation of resources, pollution, biological invasions and new drivers) affect biodiversity in the future? What are the temporal and spatial distribution of threats and pressures on biodiversity and the cumulative and interactive effects of the different drivers?

6.3 How do specific types of agriculture/forestry/fishery/aquaculture affect biodiversity (like intensive farming types etc.)? How do new trends in agriculture and renewable energy production (biomass for fuel or energy production) affect biodiversity?

6.4 What are the biodiversity impacts of EU consumption patterns, particularly for resources? How do specific products or services positively or negatively affect biodiversity in all phases of the life-cycle?

6.5 How will a changing European demography and economic activities (production of goods and services) affect species in a temporal and spatial perspective?

7. Invasive Species and biodiversity

7.1 What is the current status of alien species and particularly of invasive species in Europe?

7.2 Which species are threatening biodiversity or ecosystem services?

7.3 Are priority invasive species and their pathways identified, controlled and removed?

B: List of Target high level questions

Under each question, we provide suggestions for the types of data required to answer it.

All gap analysis partners are requested to relate their results to these questions

Can we identify status and trends of [European] species? Can we identify status and trends of biodiversity taking interspecific phylogenetic or intraspecific genetic diversity into account? Can we assess the risk of extinction?

- Data on functional traits (ecological, life-history, morphological etc) of species
- Data on phylogeny / genetic diversity of species
- Data on species lists and their phylogenetic/taxonomic relationships
- Occurrence / abundance data over time
- Current red list status of the species (of the species)
- Data on major threats to European species
- Scenarios (and data) on future environmental and climate change

Can we assess the status and trends of [European] ecosystems and ecosystem services?

- Lists of species and the ecosystem services they perform or contribute to (by their functional traits)
- Occurrence data for relevant ecosystem services
- Comparable geo-referenced occurrence (abundance) data over time
- Can we infer ecosystems from occurrence data or do we need independent data on ecosystems and their composition under ideal / natural conditions? Does sufficient taxonomic data exists (regarding number of species / species names, estimation of number of dark taxa etc.)
- Data on major threats to ecosystem functioning in Europe (e.g. on species composition and abundance).

Are we closing the biodiversity knowledge gap (poorly known organisms, ecosystem services, areas)?

- Trends in accumulation of occurrence data (of different quality) over time with respect to taxonomic groups, geographic areas, ecosystem services, genetical information etc
- Lists of species and the ecosystem services they provide when analyzing ecosystem services knowledge gaps
- Improvement of the quality of occurrence data (removal of duplicates, validation etc.
- Improved quality of the taxonomic information (building a global registry of species names, compatibility problems between CoL and GBIF classifications), are we also closing the gaps for less intensive studied/considered groups (e.g. bacterial or viral diversity)

Are we filling the gaps in historical knowledge (in relation to available historical data in collections, literature and non-mobilized digital datasets) so we can evaluate long-term trends?

- Trends in accumulation of historical occurrence data (of different quality) according to different timespans (long-term distribution data at least with the beginning of the 1980ies).
- Estimates of the total amount of available historical data in collections, literature, and non-mobilized digital datasets

Can we identify trends in the spread and effects of alien and invasive species [in Europe]?

- Data on traits (ecological, life-history, morphological etc) of species
- High-resolution occurrence / abundance data over time
- Occurrence / abundance data over time
- Data on major routes and vectors of penetration of alien species in Europe
- Data on most invaded ecosystems
- Data on the ecological and economic impact of alien species to European ecosystems
- Scenarios (and data) on future environmental and climate change

Can we identify drivers behind [European] changes in biodiversity over time?

- Data on biodiversity changes (see above)
- Data on traits (ecological, life-history, morphological etc) of species
- Data on human impact, changes in land use etc
- Data on climate and environment
- Data on vulnerability and adaptability of species regarding drivers
- Certain scientific questions could be answered through systematic reviews with correlative and experimental data (to be extracted)

Can we assess the effect of [European] marine and terrestrial protected areas on the conservation of biological diversity?

- High-resolution occurrence data over time (monitoring data) for protected areas and control areas
- Species lists and their higher phylogeny/classification (testing randomness through the taxonomic distinctness) under a BACI (before-after-control-impact) approach.
- Data on biodiversity changes (see above)

5.2 ANNEX 2: CHAPTERS AND AUTHORS

5.2.1 General part Chapter 2

Authors: *Florian Wetzel, Kessy Abarenkov, Urmas Koljalg, Christoph Häuser*

5.2.2 Specific Analyses Chapter 4

Monitoring trends in GBIF mobilized content to help address gaps
Authors: Tim Robertson, Mélianie Raymond, Andrea Hahn, Donald Hobern

Gap analysis – distribution of vascular plants in Europe.
Authors: Karol Marhold, Matúš Kempa, Alexander Sennikov, Pertti Uotila

Gap Analysis about Marine Species Distribution and Traits.
Authors: Nicolas Bailly, Kathleen Kesner-Reyes and Christos Arvanitidis, Sarah Faulwetter, Eva Chatzinikolaou

Gap analysis of available biodiversity information sources and identifying priorities.
Author: Corinne Martin

Availability of freshwater biodiversity data
Author: Aaike de Wever

Gap analysis European Monitoring schemes - using EuMon for evaluating European monitoring schemes
Authors: Jean-Baptiste Mihoub, Dirk Schmeller, Klaus Henle

Example for a EU-wide monitoring scheme: Atlas of European Breeding Birds (version 1&2) and the Pan European Common Bird Monitoring Scheme
Authors: Lluis Brotons, Sergi Herrando

Gap analysis of pollinator species (Hymenoptera: Apoidea: Anthophila)
Authors: Florian Wetzel, Isabel Calabuig, Lotte Endsleff, Lyubomir Penev, Pavel Stoev, Tim Robertson; Michael Kuhlmann

Gap analysis of Nucleotide Sequence Databases
Authors: Urmas Koljalg, Kessy Abarenkov

Gap analysis of taxonomic of databases for European terrestrial, marine and freshwater animal species
Authors: Florian Wetzel, Günther Korb, Lyubomir Penev, Pavel Stoev, Yde de Jong

Gap analysis of vascular plant species taxonomy
Authors: Anton Güntsch, Eckhard von Raab-Straube

Gap analysis of environmental datasets - LTER Data
Authors: Stefan Stoll, Jonathan Tonkin

5.2.3 Further Contributions

Authors: *Quentin Groom, Alexander Kroupa, Christian Schmid-Egger*