Moe, S. Jannicke; Schmidt-Kloiber, Astrid; Dudley, Bernard J.; Hering, Daniel. 2013 The WISER way of organising ecological data from European rivers, lakes, transitional and coastal waters. *Hydrobiologia*, 704 (1). 11-28. 10.1007/s10750-012-1337-0

Title: The WISER way of organising ecological data from European rivers, lakes, transitional and coastal waters

Authors: S. Jannicke Moe[1], Astrid Schmidt-Kloiber[2], Bernard J. Dudley[3], Daniel Hering[4]

1) Norwegian Institute for Water Research, Gaustadalléen 21, NO-0349 Oslo, Norway
2) University of Natural Resources & Life Sciences, Department Water, Atmosphere, Environment, Institute of Hydrobiology & Aquatic Ecosystem Management, Max Emanuel-Straße 17, 1180 Vienna, Austria
3) Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom
4) University of Duisburg-Essen, Faculty of Biology, Aquatic Ecology, Universitätsstraße 5, D-45141 Essen, Germany

Corresponding author: Jannicke Moe.
E-mail: jmo@niva.no, telephone: +47 908 98 108, telefax: +47 22 18 52 00

Abstract. The implementation of the Water Framework Directive (WFD) has required intense research in applied aquatic ecology in Europe, and thus created challenges for data management in international research projects. In the project WISER (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery), biological and environmental data from rivers, lakes and coastal waters in 26 European countries were collated. More than one million records of biological observations were stored in the project's central database, representing phytoplankton, macrophytes, macroalgae, angiosperms, phytobenthos, invertebrates and fish. The central database includes new data from the WISER field campaign in lakes and coastal waters during 2009-2010 (more than 6,000 biological samples from 58 waterbodies in 14 countries). The purpose of this paper is to provide an overview of the data collated within WISER, in order to facilitate future re-use of these data by other scientists. More specifically, the objectives are: (1) to describe the data management in WISER, (2) to describe the structure and content of the WISER central database, and (3) to share experiences and give recommendations for data management in large ecological research projects.

Key words: EU Water Framework Directive (WFD); data management; databases; taxonomy; intellectual property rights; metadata

Introduction

The implementation of the Water Framework Directive (WFD; European Commission, 2000) has required intense research in applied aquatic ecology in Europe, and thus created challenges for data management in international research projects. As specified by the WFD, the assessment of the quality of rivers, lakes and coastal ecosystems must be based on biological indicators. Previously, the monitoring of aquatic ecosystems was focussed on abiotic water quality and mainly limited to physico-chemical variables. Therefore, many EU member states have recently developed national classification systems for assessment of ecological status of water bodies based on phytoplankton, macrophytes, invertebrates and fish (Birk et al., 2012a). Moreover, the WFD has required intercalibration of national classification systems among countries in geographical regions with similar waterbody types (Birk et al., 2012b).

A major consequence of the WFD has been the acquisition of large amounts of biological information on the status of European surface waters, information that may improve our knowledge of the structure of the communities inhabiting these ecosystems. The need for development, validation and intercalibration of biological classification systems compliant with the WFD has triggered large-scale European research projects, such as REBECCA (http://www.environment.fi/syke/rebecca) and WISER (http://www.wiser.eu). Other EU projects have focussed on the challenges of implementing the WFD under climate change (e.g. Euro-Limpacs and REFRESH; http://www.refresh.ucl.ac.uk). Within such large international research projects, extensive amounts of ecological data have often been generated or collated from various sources including previous project data, on-going national monitoring programmes, and new field surveys. These data have been stored in large databases comprising information on hundreds or thousands of water bodies, including the AQEM/STAR taxa database for river biota (Furse et al., 2006), the REBECCA lake phytoplankton and macrophyte databases (Moe et al., 2008), and Baltic sea data in the CHARM project (http://www2.dmu.dk/1_viden/2_miljoe-tilstand/3_vand/4_charm/charm_main.htm). Some databases resulting from these projects are maintained and used actively after the termination of the original project, e.g. the Taxa and Autecology Database for Freshwater Organisms (http://www.freshwaterecology.info; Schmidt-Kloiber & Hering, 2012).

Potentially, data from previous research projects could contribute significantly to other objectives in addition to those of the WFD, e.g. for monitoring the effects of emerging stressors, for improving our knowledge of species distributions and species invasions, for understanding broad-scale drivers shaping community assemblages, and for Habitats Directive/Natura 2000 species inventories and biodiversity records. There is considerable interest in using such data beyond the lifetime of the individual project. Nevertheless, many of these ecological databases have a very limited afterlife. As pointed out by Beniston et al. (2012), much effort is often expended in the initial phases of each new project to collate existing data generated in previous projects, which are often difficult to access, buried in the grey literature or lost on inaccessible databases. At the same time, difficulties in obtaining data represent barriers to policy-relevant research, on topics such as climate change impacts, water quality or biodiversity protection (Beniston et al., 2012).

A key barrier to the use of previously generated data is that scientists who produce the data may be unwilling to share them, due to strong traditions, competition for funding and other circumstances (Beier et al., 2007; Costello, 2009). However, there may be several more practical and technical reasons. Institutional barriers can be a major obstacle to data access, where data are not centralised but are stored in various formats with little compatibility (Beniston et al., 2012). Even if data are accessible, lack of proper data documentation and dissemination after the termination of the project impedes re-use of the data (Refsgaard et al. 2007). Other restrictions on data access include: (i) improper data organisation (e.g., in poorly linked spreadsheets rather than in a relational database) may inhibit efficient data extraction; (ii) there is no contact person responsible for answering requests, (iii) there is no service for extraction of data from the database (such as a user interface or a person to handle data extraction upon request); (iv) there is insufficient documentation specifying analytical methods, sources, taxonomy etc.; (v) there is uncertainty regarding the intellectual property rights of the data for further use.

In the recently completed project WISER (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery), attempts were made to build upon experiences from former projects and improve the usability of the project's data. A data service team was therefore established, aimed at both facilitating the data flow within the project, and providing information about the availability of these data for other scientists. Consequently, a publicly available metadatabase (http://www.wiser.eu/results/meta-database) was developed to provide information on data sources used in WISER, as well as many other relevant datasets hosted by the project partners. The metadatabase holds key information about each data source, such as intellectual property rights (IPR) and contact information for data owners (for more information see Schmidt-Kloiber et al., this issue).

The purpose of this paper is to provide an overview of the data collated within WISER, in order to facilitate future re-use of these data by other scientists. More specifically, the objectives are:

1) to describe data management in WISER,
2) to describe the structure and content of the WISER central database, and
3) to share experiences and give recommendations for data management in large ecological research projects.

Data management in WISER

The principles of data flow in the WISER project broadly reflected the structure of the project organisation (see Hering et al., this issue), as illustrated in Figure 1. Two overarching data categories were defined: 'background data' were from previous research projects, national monitoring programmes etc., and 'foreground data' were collected in the field during the project. The base unit of the data flow is termed 'dataset'. A dataset typically corresponded to a single data file (e.g. an MS Access database or an MS Excel workbook) from one data provider. A list of more than 100 background datasets were identified as available to WISER before the project was initiated (see Schmidt-Kloiber et al., this issue).

Two groups of work packages (WPs) were defined according to their role in the data flow structure (Figure 1). Group 1 comprised WPs that collected foreground and/or background data for their own use, from lakes (WPs 3.1-3.4), transitional/coastal waters (WPs 4.1-4.4) or rivers (WP 5.1; background data only). The lakes and transitional/coastal WPs each worked on a single biological quality element (BQE), while WP5.1 included all river BQEs (see Table 1 for more details). The foreground data comprised one dataset per work package. Group 2 consisted of WPs working with the integration of multiple BQEs, for example the comparison of responses of different BQEs to pressure gradients (see (see Hering et al. (2012) for more details), and which had data needs overlapping with WPs in Group 1. Potential data sources for Group 2 included the new foreground datasets collected by Group 1, as well as the large number of registered background datasets. A major task for the data service team was therefore to facilitate data flow from Group 1 to Group 2, in order to minimise duplication of work on compilation, harmonisation and processing of the same datasets within different WPs.

In each Group 1 WP, a WP data manager was responsible for compiling the relevant foreground and/or background datasets (i.e., for a single biological quality element) into a WP database (Figure 1). The WP data manager was also responsible for quality checking and extraction of data for users within the WP. Examples of scientific results from the use of each WP database can be found in the references in Table 1, as well as in the synthesis papers for lakes (Solheim, 2012) and for transitional/coastal waters (Borja et al., 2012). All WP databases were delivered to the data service team, which subsequently compiled these into the central database (CDB). Group 2 WPs that needed data from the central database received the requested data as an MS Access database or extracted into another preferred data format. In order to facilitate the data flow, a common WISER database structure was developed (see below), which all WPs were encouraged to use. All templates, tools and guidelines for data management were therefore based on the common database structure. Nevertheless, more pragmatic solutions for data flow were sometimes required due timing mismatches between data delivery from Group 1 and data needs by Group 2.

Structure of the WISER central database

The structure of the central database was designed to meet the needs of several different research problems within the project (see Hering et al., this issue): (i) combination of data across different BQEs, (ii) combination of biological with environmental data, (iii) combination of data from different water categories (i.e. rivers, lakes and transitional/coastal waters), (iv) usability of data for hierarchical uncertainty analysis, (v) combination of data from the WISER field campaign (foreground) with other data (background), and (vi) linkage of data to information in the metadatabase.

The CDB had a hierarchical structure with tables corresponding to the hierarchical levels of the WISER field campaign, each related in a one-to-many relationship to the next: dataset, waterbody, station, sample, and value. A full description is given in the WISER Data Dictionary (http://www.wiser.eu/results/central-database). Each dataset (as defined under *Data management in WISER*) was assigned a unique identifier (DatasetID) and was represented by a unique record in the metadatabase. The DatasetID was thus critical for linking the data to key information about the data, such as data owners and intellectual property rights (IPR). Some of the large international databases available to WISER comprised several sections (e.g. countries) with different data owners and thus different IPR. In order to facilitate the storage and tracking of IPR information in such cases, one single object (such as a database) could be defined as multiple datasets in the WISER metadatabase (Schmidt-Kloiber et al., this issue).

The waterbody table was based on the waterbody concept underlying the WFD, which breaks the network of rivers, lakes and coastal waters down into base units of waterbodies that should be monitored, classified and (if necessary) restored. In principle all waterbodies that are reported to EU under the WFD have a unique national code, but in practice waterbodies are often recorded with different codes in different datasets. In the WISER CDB all waterbodies were stored with the waterbody code originally given in the source dataset. For some waterbodies there was much environmental information available, especially related to the WFD waterbody typology (e.g. levels of altitude, lake surface area or river catchment area), on which ecological status classification systems often depend. All environmental information that was not associated with a sampling date was stored in a separate table related to the waterbody table, in order to limit the size of the more fundamental waterbody table.

Harmonization of the waterbody coding, i.e. to identify common waterbodies from different data sources, was a major challenge. Such harmonization was required because some of the data analysis in Group 2 was based on integrated data from several BQEs (i.e. from different WP databases). For example, the analysis of cross-taxon responses to stress gradients in streams and lakes (Angeler et al., 2012) required raw data from 3-4 BQEs from the same waterbody. The foreground data from the field campaign contained only a limited number of waterbodies (Table 1B), which could easily be harmonised. In the background data, however, a waterbody could appear in several different datasets with different coding. Moreover, geographic coordinates were sometimes missing, which rendered reliable identification of the waterbodies impossible. Consequently, waterbody coding was harmonised only for a subset of the background data, i.e. for lakes in countries from where 3 or 4 BQEs were reported (Belgium, Estonia, Finland, Germany, Ireland, Lithuania, Latvia, Netherlands, Norway, Poland, Romania, Sweden and UK) as well as for all river stations.

The station table contained only the most basic information regarding the sampling location, such as station code, station name, and geographical coordinates. A station was regarded as the spot location where the sample was taken and could be characterised by coordinates. A station always belonged to a single waterbody, whereas a waterbody could contain more than one station.

Biological and environmental samples were stored in separate tables, as relationships between biological and environmental samples within a dataset could be complex (both one-to-many and many-to-one). To find a consistent way of defining a 'sample' for all biological groups, in terms of unique combinations of other sampling information, was a critical task. Analysis of data for several BQEs combined typically involved calculation of a biological index value for each sample, therefore a common definition of the 'sample' level was necessary for data analysis in Group 2 (Angeler et al., 2012). Moreover, an unambiguous definition of biological samples was also needed for a consistent uncertainty analysis of index values for the different BQEs (e.g. (Balsby et al., 2012; Carvalho et al., 2012; Dromph et al., 2012; Dudley et al., 2012b; Mascaró et al., 2012) which in turn was required for assessing confidence in classification results (Clarke, 2012). However, the sampling methods varied substantially among BQEs: for example, phytoplankton was sampled in bottles, macrophytes in transects, fish in net campaigns lasting several days and in different locations. Therefore the database structure was developed in close communication with all WP data managers, in order to ensure that the sample table contained all fields required for a unique definition of 'sample' for all BQEs (i.e., a set of records containing no more than one observation per taxon). As a result, the definition of unique biological samples across all BQEs was a unique combination of the following fields: station, sample date, upper and lower sample depth, sample location (e.g. habitat), sample method, sample type (i.e. BQE), and replicate number. For definition of unique environmental samples, the same set of fields was used except sample method. Any information about methodology used for collection and analysis of the individual samples that were included in the WP databases, were stored in the sample tables. Additional methodological information for the original datasets may be found in the metadatabase, under "Sample specification".

Code lists were developed for the most important fields in the database and distributed to all WPs, to allow for standardisation of the content of the CDB. All code lists are included in the WISER data dictionary (http://www.wiser.eu/results/central-database). Taxonomic code lists were developed by each WP and combined in the CDB. The complete taxonomic code list also provides a link to the taxonomic codes of freshwaterecology.info.

Content of the WISER central database

The WISER central database was composed of WP databases compiled by WPs 3.1-5.1 (Figure 1). The content was therefore determined by the data sources that these WPs selected for their own objectives. Some WP databases contained additional information that was eventually not included in the CDB, e.g. data on climate, land use and other environmental pressures. More information about the content of individual WP databases can be found in the references listed in Table 1.

The CDB contains data from 26 countries (Figure 2), with over one million records of biological values, most of which are species observations with an abundance measure. Summary statistics (numbers of countries, stations, samples etc.) for each biological quality element are given in Table 1. The background data (Table 1A) included 49 datasets, mostly from lakes and rivers as well as fish in transitional/coastal waters. The background data comprised approximately 100,000 biological samples from over 6000 waterbodies in 26 countries, and 70,000 environmental samples from these waterbodies (including chlorophyll-a). The foreground data (Table 1B) included all the data collected in lakes, transitional and coastal waters during the WISER field campaign and delivered to the data service team by the end of 2011. The field campaign resulted in almost 30,000 biological records from over 6,000 samples from 58 waterbodies in 14 countries. In addition, the foreground data contained almost 10,000 samples of environmental data.

The number of biological samples and records was unevenly distributed among countries, water categories and BQEs (Figure 3). The number of samples may not be directly comparable across different BQEs, since very different sampling methods are used for e.g., phytoplankton versus fish. The number of records represents the total number of taxa (usually species level) in all samples combined. Data from rivers were relatively balanced for the different BQEs, but dominated by central Europe (Figure 3a). Data from lakes were dominated by phytoplankton and fish in northern and central-European countries (Figure 3b), while coastal/transitional data were dominated by macroalgae/angiosperms and fish from central- and southern-European countries (Figure 3c).

The total number of taxa per country was typically 2-3 orders of magnitude higher than the average number of taxa per sample from the respective country (Figure 4). The 'taxon' here represents the highest taxonomic resolution available for each record, which was usually species. Within each BQE, the average number of taxa per sample was relatively stable across countries, although the numbers may vary with an order of magnitude in some cases. The total number of taxa per country was more variable across countries. However, it should be noted that the total number of taxa for a country tends to increase with the total number of samples. Therefore, a high number of recorded taxa for a country do not only reflect species richness, but also the amount of data delivered from this country. In general, the highest taxon richness per country was found for river macroinvertebrates (>1000 taxa for a few central-European countries) and for lake phytoplankton (300-1000 taxa for many countries). Other more conspicuous peaks in taxon richness probably reflect that certain countries provided a large number of samples from a particular BQE to the database (for example, data on transitional/coastal macroinvertebrates from Spain and fish from France).

The environmental data were also unevenly distributed among WPs (Table 1), and were strongly dominated by water samples taken for coastal phytoplankton. However, since environmental data collected for one BQE could also be used for analysis of other BQEs in the same waterbody (for the set of waterbodies where coding was harmonised across BQEs), the availability of environmental data for each BQE was somewhat less skewed than what appears from Table 1. For rivers, the following environmental parameters have the highest number of records in the CDB (in descending order): orthophosphate, conductivity, nitrite, pH, water temperature, oxygen saturation (all >5000 records), followed by nitrate, ammonia, total phosphorus, chloride, alkalinity, oxygen and BOD5 (all >2000 records). For lakes, the most common the environmental parameters in the CDB were (in descending order): total phosphorus, chlorophyll a, water temperature, total nitrogen, pH, conductivity, Secchi depth and oxygen (all >40,000 records), followed by water colour, alkalinity, dissolved inorganic phosphorus, chemical oxygen demand, turbidity, dissolved inorganic and organic nitrogen, and orthophosphate (all >20,000 records). For transitional/coastal waters, the most common environmental parameters were salinity, water temperature, oxygen, oxygen saturation, conductivity (all >2000 records); pH, chlorophyll a, total nitrogen, total phosphorus, nitrate, ammonia and orthophosphate (all >100 records).

Future use of WISER data

The data service team intended to make the WISER central database publicly available as far as possible via a web-based tool, i.e. all datasets where the IPRs would allow unconditional downloading and further use. However, the majority of the datasets that were finally stored in the CDB have IPRs that are too strict to allow unmonitored distribution; e.g. the data owners have requested to be offered co-authorship in publications or to be informed about further use of the data. It would therefore have been irresponsible to make these data publicly available, as it would have been infeasible to follow up each data download and check that the IPRs are respected in each case. Further, using data collected for a different purpose may require more knowledge of the individual datasets than the information currently available in the metadatabase. It is therefore recommended that further use of the WISER data involve collaboration with scientists from the WISER consortium.

For scientists who are interested in using the WISER data, the following approach is recommended. Scientists who are interested in a substantial part of the total WISER CDB (e.g., all lakes data) should contact the WISER data service team (authors of this paper). Scientists who are interested in all data for a single biological quality element (e.g., phytoplankton in lakes) should contact the respective WP data manager (see Table 1; further contact information is given in the data dictionary). Scientists who are interested in a single dataset (e.g., phytoplankton in Norwegian lakes) should contact the respective contact person listed in the metadata query output (http://www.wiser.eu/results/meta-database). Note that the metadatabase contains information and contact details for twice as many datasets as are stored in the WISER central database (see Schmidt-Kloiber et al., this issue for more details).

Concluding remarks and recommendations

The main purpose of this paper was to inform about the structure and content of the WISER central database, in order to facilitate further scientific use of this very comprehensive data resource. A second purpose was to share experiences of the WISER data service team, which might be relevant for other research projects involving compilation and multiple uses of ecological data. Some technical recommendations for data compilation are provided by Moe et al. (2008). In the following, the experiences of the data service team and highlighted and a series of recommendations are offered for other environmental projects.

1) Overview of data sources. The initial mapping of available data sources from the very beginning of the project's preparatory phase was an important first step. A preliminary overview of the available datasets allowed all partners to indicate their data needs in a consistent way in the project proposal. This information in turn enabled the data service team to map the overlapping data needs, and on this basis elaborate a data management plan which was presented and discussed at the first project meeting.

2) Information on intellectual property rights (IPR). Based on lessons learned from previous projects, information on IPR and contact information of data providers were requested for each dataset from the very beginning of the project and stored in the metadatabase. As mentioned, the IPR rules imposed by data owners were ultimately rather strict, usually requiring that data owners must be contacted for each new use of the data. Although all IPR information was available from the publicly available metadatabase throughout the project, project partners often struggled with finding and following the specific IPR criteria for the datasets they had used. In hindsight, we would recommend that only datasets for which there is no requirement to contact data owners should be distributed within the project. (We would nevertheless recommend users to contact the data owners, who will often be able to contribute with additional documentation and knowledge).

3) Centralised vs. decentralised data management. Originally a more centralised data management was suggested for the project, but early feedback from project partners revealed that Group 1 WPs preferred to manage their own data independently of a central database. The two-step data compilation procedure with both WP databases and a central database was planned accordingly (Figure 1). This decentralised data management was more flexible and efficient for Group 1, and contributed to the high scientific productivity (see examples in Table 1). On the other hand, this solution may have compromised and delayed the data delivery to Group 2 relative to the original plan. The best solution for other projects will depend on the aims and resources of the project; for more discussion of cost and benefits of project databases see Moe et al. (2008) and Refsgaard et al. (2007).

4) 'High tech' vs. 'low tech' solutions. Assistance with data compilation and data extraction was generally offered in two ways: a 'high tech' option where the data service team developed database tools with user interfaces (e.g. the WISER data extraction tool; (e.g. the WISER data extraction tool; Dudley et al., 2012a), and a 'low tech' alternative where data extraction etc. was performed upon demand and in dialogue with the data users. The 'low tech' solutions were often preferred by the data users; therefore the development of more advanced tool-based solutions was given lower priority in this project (even though more advanced solutions could have facilitated future use of the data). Hence, before investing resources in developing tools and user interfaces for data users, establishing whether partners are interested and willing to use such tools can be worthwhile.

5) Coordination of data management. Large research projects will normally experience delays and other deviations from the work plan, and the data management strategy may need to be adapted in the light of the progress. The WISER project had rather ambitious plans for data flow given the tight time schedule, and therefore required close communication with all data managers and frequent update of data management plans. For example, it was discovered that information for waterbody classification based on the new WISER field data (Group 1) would not be available in time for a planned analysis of integrated classification using all four BQEs (Group 2); therefore alternative data sources were used instead (Caroni & van den Bund, this issue). The central coordination of data management was clearly beneficial for WISER and is recommended for other research projects with shared use of data.

In conclusion, to ensure adaptive data management in research projects with composite and overlapping data needs like WISER, some degree of central coordination of the data flow is recommended, including a proper metadatabase. The time required for this task can easily be underestimated; therefore sufficient resources should be allocated from the beginning of the project. A data management period of e.g. three months after the official termination of the project can be useful, allowing time for harmonising data, completing metadata and placing the data on a public data repository. This investment will facilitate future re-use of the project's data by project partners as well as other scientists. Finally, we support the recommendation of Beniston et al. (2012) for easier access to data and information in water- and climate-related sciences: the establishment of a general well-defined and easily accessible 'clearinghouse' of relevant and structured data and metadata, which explicitly includes data produced by EU-funded and related projects.

Acknowledgments

We thank Roar Brænden, Jan Karud, Jörg Strackbein and Robert Vogl for constructive discussions and technical advice. All WP data managers (Table 1) are acknowledged for collaboration and delivery of data to the central database. The WISER project was funded by the European Union under the 7th Framework Programme, Theme 6 (Environment including Climate Change), contract no. 226273. We thank Christian Feld, Laurence Carvalho and two anonymous reviewers for helpful comments to the manuscript.

References

Alvarez, M. C., A. Franco, R. Pérez-Domínguez & M. Elliott, 2012. Sensitivity analysis to explore responsiveness and dynamic range of multimetric fish-based indices for assessing the ecological status of estuaries and lagoons. Hydrobiologia (this issue).

Angeler, D. G., R. K. Johnson, S. J. Moe & D. Hering, 2012. Cross-taxon responses to stress gradients in streams and lakes. Report and manuscript on the use of BQEs, habitats and ecosystems for detecting human-induced change WISER deliverable vol 6.3-2, 39.

Argillier, C., S. Caussé, M. Gevrey, S. Pédron, J. D. Bortoli, S. Brucet, M. Emmrich, E. Jeppesen, T. Lauridsen, T. Mehner, M. Olin, M. Rask, P. Volta, I. J. Winfield, F. Kelly, T. Krause, A. Palm & K. Holmgren, 2012. Development of a fish-based index to assess the eutrophication status of European lakes. Hydrobiologia (this issue).

Balsby, T., J. Carstensen & D. Krause-Jensen, 2012. Sources of uncertainty in estimation of eelgrass depth limits. Hydrobiologia (this issue).

Beier, U., E. Degerman, A. Melcher, C. Rogers & H. Wirlöf, 2007. Processes of collating a European fisheries database to meet the objectives of the European Union Water Framework Directive. Fisheries Management and Ecology 14:407-416.

Beniston, M., M. Stoffel, R. Harding, M. Kernan, R. Ludwig, E. Moors, P. Samuels & K. Tockner, 2012. Obstacles to data access for research related to climate and water: Implications for science and EU policy-making. Environmental Science & Policy 17:41-48.

Birk, S., W. Bonne, A. Borja, S. Brucet, A. Courrat, S. Poikane, A. Solimini, W. van de Bund, N. Zampoukas & D. Hering, 2012a. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18:31-41.

Birk, S., N. Willby, A. Borja, M. Kelly, W. Bonne, W. v. d. Bund, S. Poikane, U. Schmedtje & N. Zampoukas, 2012b. Intercalibrating ecological status classification: Europe’s quest for common management objectives of aquatic ecosystems. Hydrobiologia (this issue).

Borja, A., E. Barbone, A. Basset, G. Borgersen, M. Brkljacic, M. Elliott, J. M. Garmendia, J. C. Marques, K. Mazik, I. Muxika, J. M. Neto, K. Norling, J. G. Rodríguez, I. Rosati, B. Rygg, H. Teixeira & A. Trayanova, 2011. Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct European coastal and transitional ecosystems. Marine Pollution Bulletin 62:499-513.

Borja, A., M. Elliott, P. Henriksen & N. Marbà, 2012. Transitional and coastal waters ecological status assessment: advances and challenges resulting from implementing the European Water Framework Directive. Hydrobiologia (this issue).

Caroni, R., W. van de Bund, R.T. Clarke & R.K. Johnson, 2012. Combination of multiple biological quality elements into waterbody assessment of surface waters. Hydrobiologia (this issue).

Carvalho, L., S. Poikane, A. L. Solheim, G. Phillips, G. Borics, J. Catalan, C. d. Hoyos, S. Drakare, B. Dudley, M. Jarvinen, C. Laplace-Treyture, K. Maileht, C. McDonald, U. Mischke, J. Moe, G. Morabito, P. Nõges, T. Nõges, I. Ott, A. Pasztaleniec, B. Skjelbred & S. Thackeray, 2012. Strength and uncertainty of lake phytoplankton metrics for assessing eutrophication impacts in lakes. Hydrobiologia (this issue).

Clarke, R., 2012. Estimating confidence of European WFD ecological status class and WISER Bioassessment Uncertainty Guidance Software (WISERBUGS). Hydrobiologia (this issue).

Costello, M. J., 2009. Motivating Online Publication of Data. BioScience 59:418-427.

Dahm, V., D. Hering, D. Nemitz, W. Graf, A. Schmidt-Kloiber, M. Seebacher, A. Melcher & C. K. Feld, 2012. Effects of physico-chemistry, land use and hydromorphology on three riverine organism groups: A comparative analysis with monitoring data from Germany and Austria. Hydrobiologia (this issue).

Dromph, K., S. Agusti, A. Basset, J. Franco, P. Henriksen, J. Icely, S. Lehtinen, S. Moncheva, M. Revilla, K. Sørensen & L. Roselli, 2012. Sources of uncertainty in assessment of marine phytoplankton communities. Hydrobiologia (this issue).

Dudley, B., J. Moe, A. Schmidt-Kloiber & L. Carvalho, 2012a. Extraction of data from WISER databases. Paper presented at the Current questions in water management Book of abstracts to the WISER final conference, Tallinn, Estonia, 25-26 January 2012.

Dudley, B. J., M. Dunbar, E. Penning, A. Kolada, S. Hellsten, A. Oggioni, V. Bertrin, F. Ecke & M. Søndergaard, 2012b. Measurements of uncertainty in macrophyte metrics used to assess European lake water quality. Hydrobiologia (this issue).

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy Official Journal of the European Communities L 327. vol Official Journal of the European Communities, 72.

Feld, C. K., V. Dahm, A. Lorenz, M. Logez, A. Marzin, V. Archaimbault, J. Belliard, C. Chauvin, F. Delmas, D. Pont, A. Melcher, H. Kremser, S. Muhar, S. Schmutz, T. Schulze, M. Seebacher, A. Zitek, C. Wolter, P. F. M. Verdonschot, H. Michels & H. E. Keizer-Vlek, 2012. Driver-Pressure-Impact and Response-Recovery chains in European rivers: observed and predicted effects on BQEs WISER Deliverable. vol D5.1-2, 227.

Furse, M., D. Hering, O. Moog, P. Verdonschot, R. Johnson, K. Brabec, K. Gritzalis, A. Buffagni, P. Pinto, N. Friberg, J. Murray-Bligh, J. Kokes, R. Alber, P. Usseglio-Polatera, P. Haase, R. Sweeting, B. Bis, K. Szoszkiewicz, H. Soszka, G. Springe, F. Sporka & I. Krno, 2006. The STAR project: context, objectives and approaches. Hydrobiologia 566:3-32.

Haase, P., D. Hering, S. C. Jähnig, A. W. Lorenz & A. Sundermann, 2012. The impact of hydromorphological restoration on river ecological status: a comparison of fish, benthic invertebrates, and macrophytes. Hydrobiologia (this issue).

Hering, D., H. Bennion, S. Birk, A. Borja, A. Basset, J. Carstensen, L. Carvalho, R. Clarke, H. Duel, M. Dunbar, M. Elliott, A.-S. Heiskanen, S. Hellsten, P. Hendriksen, K. Irvine, E. Jeppesen, R. Johnson, A. Kolada, A. L. Solheim, O. Malve, N. Marba, J. C. Marques, J. Moe, S. Moncheva, G. Morabito, T. Noges, D. Pont, M. Pusch, A. Schmidt-Kloiber, S. Schmutz, A. Solimini, W. v. d. Bund, P. F. M. Verdonschot & C. K. Feld, 2012. Assessment and recovery of European water bodies: key messages from the WISER project. Hydrobiologia (this issue).

Järvinen, M., S. Drakare, G. Free, A. Lyche-Solheim, G. Phillips, B. Skjelbred & U. Mischke, 2012. Phytoplankton indicator taxa for reference conditions in lowland Northern and Central European lakes. Hydrobiologia (this issue).

Karus, K. & T. Feldmann, 2012. Factors influencing macrophyte metrics in Estonian coastal lakes in the light of ecological status assessment. Hydrobiologia (this issue).

Maileht, K., T. Nõges, P. Nõges, O. Ingmar, U. Mischke, L. Carvalho & B. Dudley, 2012. Water colour, phosphorus and alkalinity are the major determinants of the dominant phytoplankton species in European lakes. Hydrobiologia (this issue).

Marbà, N., D. Krause-Jensen, T. Alcoverro, S. Birk, A. Pedersen, J. M. Neto, S. Orfanidis, J. M. Garmendia, I. Muxika, A. Borja, K. Dencheva & C. M. Duarte, 2012. Diversity of European seagrass indicators: Patterns within and across regions. Hydrobiologia (this issue).

Mascaró, O., T. Alcoverro, K. Dencheva, D. Krause-Jensen, N. Marbà, J. Neto, V. Nikolic, S. Orfanidis & A. Pedersen, 2012. Exploring the robustness of the main macrophyte-based indices developed to assess the ecological status of coastal ecosystems under the Water Framework Directive: An uncertainty analysis. Hydrobiologia (this issue).

Mjelde, M., S. Hellsten & F. Ecke, 2012. Developing av water level regulation index for aquatic macrophytes in Nordic lakes. Hydrobiologia (this issue).

Moe, S. J., B. Dudley & R. Ptacnik, 2008. REBECCA databases: experiences from compilation and analyses of monitoring data from 5,000 lakes in 20 European countries. Aquatic Ecology 42:183-201.

Orfanidis, S., K. Dencheva, K. Nakou, S. Tsioli & I. Rosati, 2012. Benthic macrophyte community changes across an anthropogenic pressure gradient in Mediterranean and Black Sea water systems. Hydrobiologia (this issue).

Refsgaard, J. C., L. F. Jørgensen & A. L. Højberg, 2007. Data availiability and accessibility. State of the art on existing data required for modelling for research purposes and for the implementation of the Water Framework Directive. Geological Survey of Denmark and Greenland.

Schmidt-Kloiber, A. & D. Hering, 2012. www.freshwaterecology.info - the taxa and autecology database for freshwater organisms, version 5.0. In. http://www.freshwaterecology.info/ Accessed 04.04.2012.

Schmidt-Kloiber, A., S. J. Moe, B. J. Dudley, J. Strackbein & R. Vogl, 2012. The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters. Hydrobiologia (this issue).

Solheim, A. L., 2012. Comparison of different biological quality elements for assessment of ecological status in lakes. Hydrobiologia (this issue).

Figure captions

Fig. 1 Conceptual diagram of data flow in WISER. For more information about the different work packages (WPs), see Hering et al. (2012)

Fig. 2 Geographic coverage of the WISER central database for rivers (a), lakes (b) and transitional/coastal waters (c). Countries labelled with 2-letter iso country code are represented in the database

Fig. 3 Number of samples and records, respectively, for each biological quality element from rivers (a), lakes (b) and transitional/coastal waters (c), in the WISER central database. The number of records equals the number of taxa per sample summarised for all samples. Each bar displays the number of both foreground and background samples (or records, respectively), from each country. Note the logarithmic scale of the y-axis

Fig. 4 Number of taxa per country and average number of taxa per sample, respectively, for each biological quality element from rivers (a), lakes (b) and transitional/coastal waters (c), in the WISER central database. Vertical lines show ± 1 standard deviation. Note the logarithmic scale of the y-axis

Figure
Click here to download Figure: CDB_ms_figures_20120831.doc

Figure 1

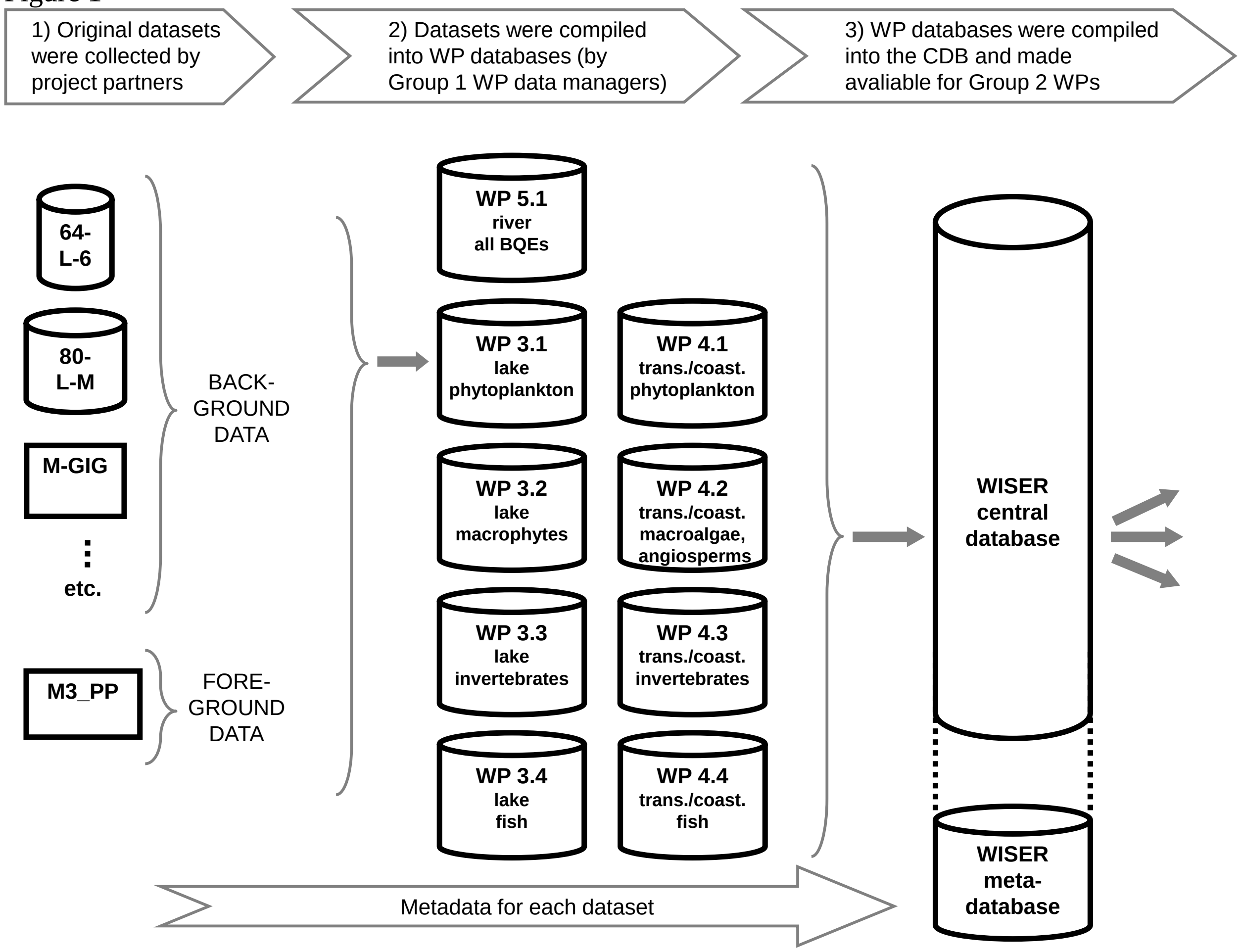
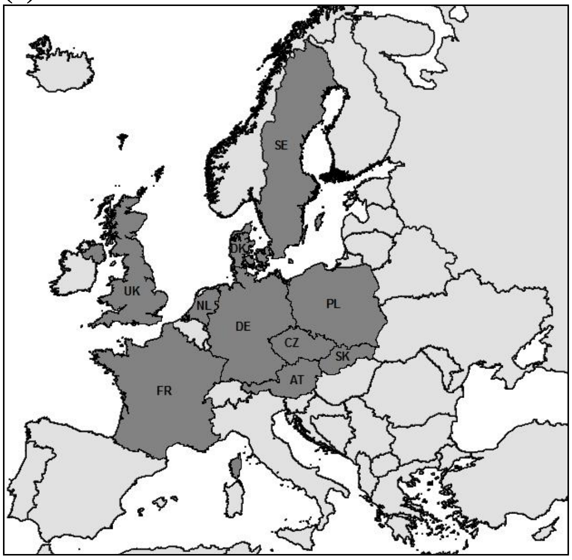
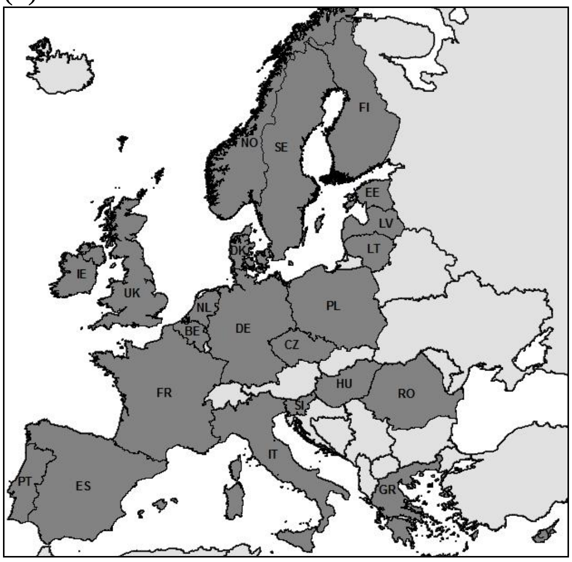


Figure 2

(a) Rivers



(b) Lakes



(c) Transitional/coastal waters

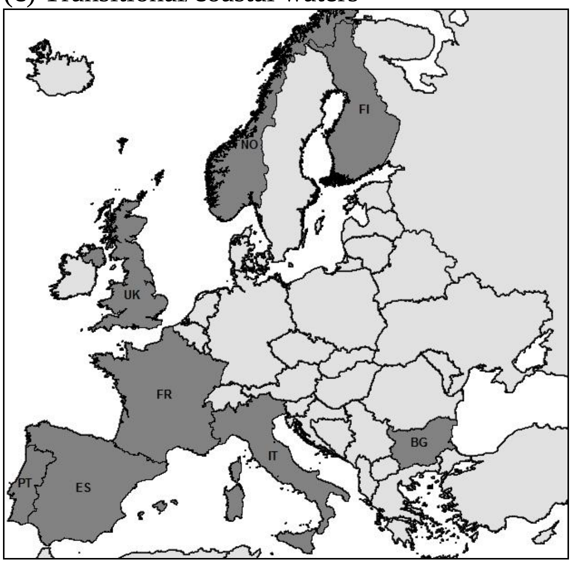


Figure 3 (a)

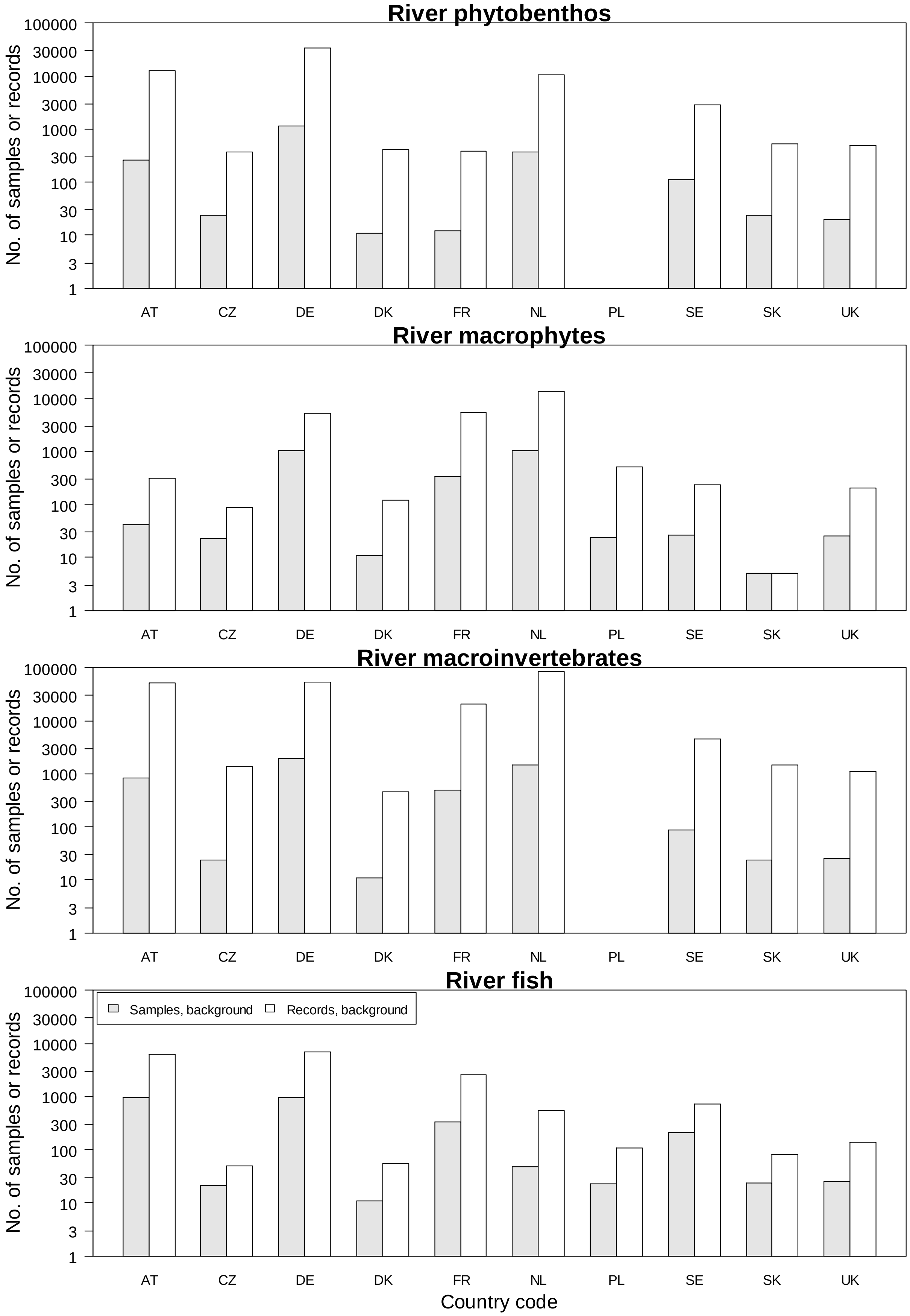

River phytobenthos
River macrophytes
River macroinvertebrates
River fish
No. of samples or records
100000
30000
10000
3000
1000
300
100
30
10
3
1
AT
CZ
DE
DK
FR
NL
PL
SE
SK
UK
Samples, background
Records, background
Country code

Figure 3 (b)

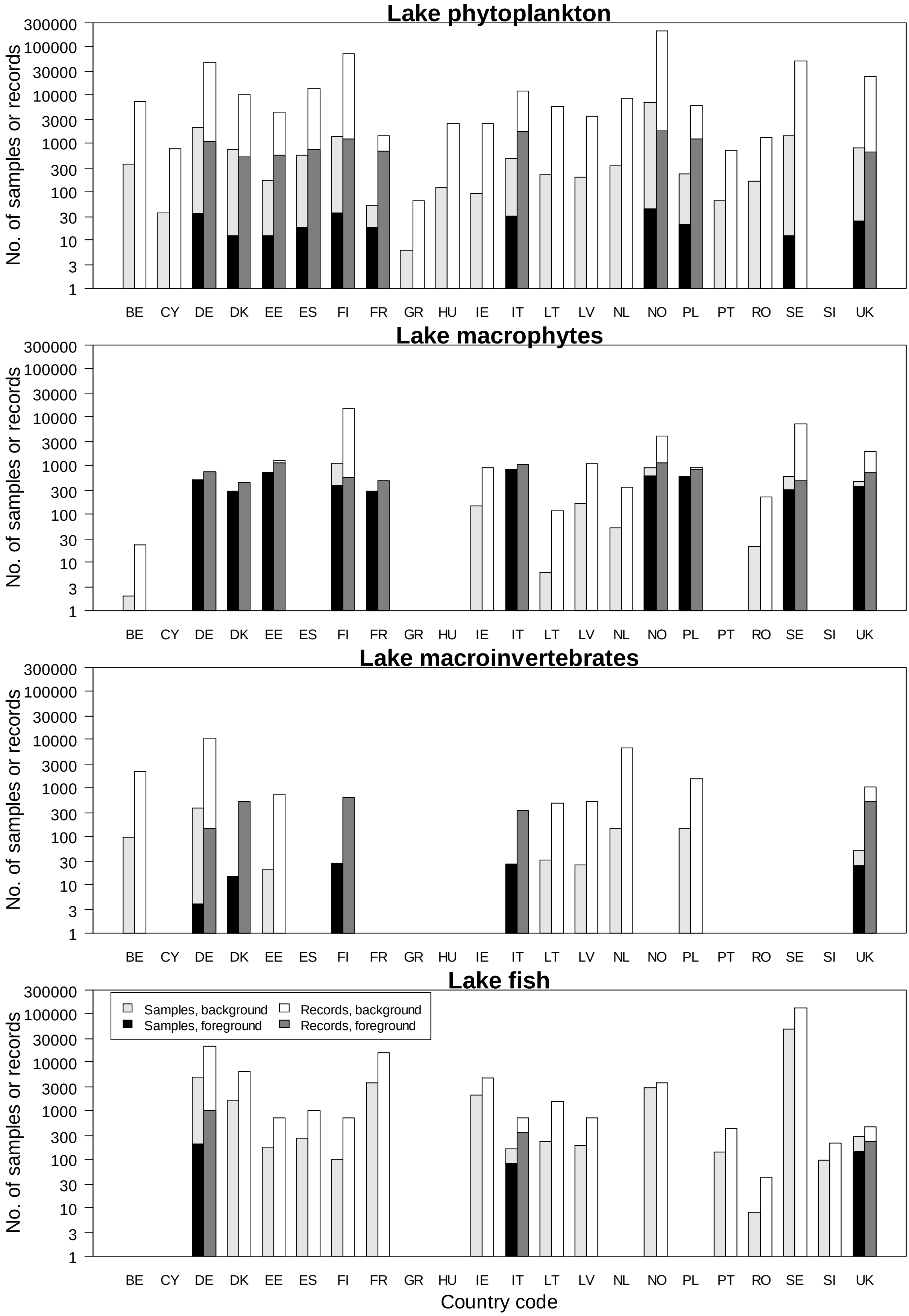

Lake phytoplankton
Lake macrophytes
Lake macroinvertebrates
Lake fish
No. of samples or records
300000
100000
30000
10000
3000
1000
300
100
30
10
3
1
BE CY DE DK EE ES FI FR GR HU IE IT LT LV NL NO PL PT RO SE SI UK
Samples, background
Samples, foreground
Records, background
Records, foreground
Country code

Figure 3 (c)

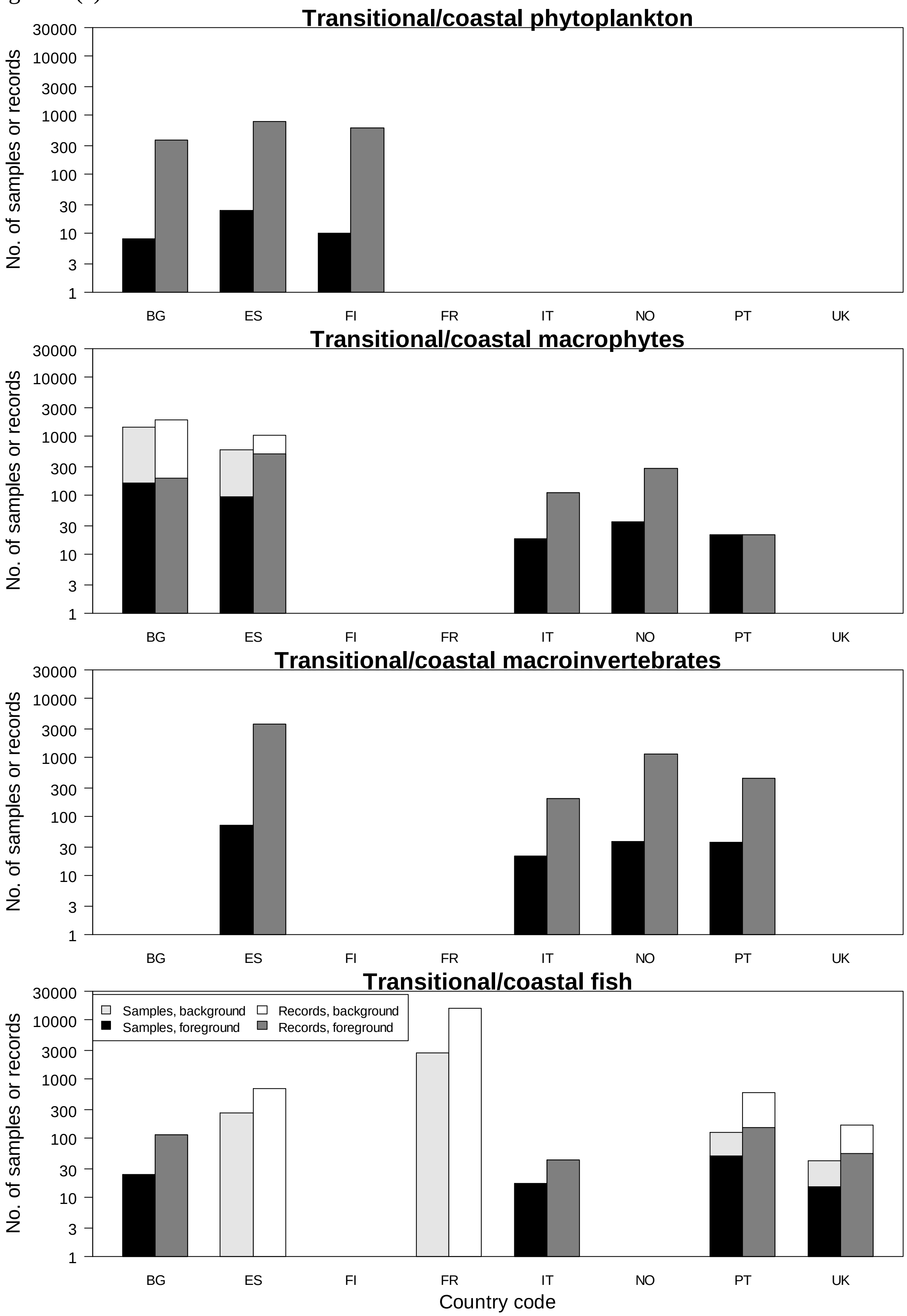

Transitional/coastal phytoplankton
No. of samples or records
30000
10000
3000
1000
300
100
30
10
3
1
BG
ES
FI
FR
IT
NO
PT
UK
Transitional/coastal macrophytes
Transitional/coastal macroinvertebrates
Transitional/coastal fish
Samples, background
Records, background
Samples, foreground
Records, foreground
Country code

Figure 4 (a)

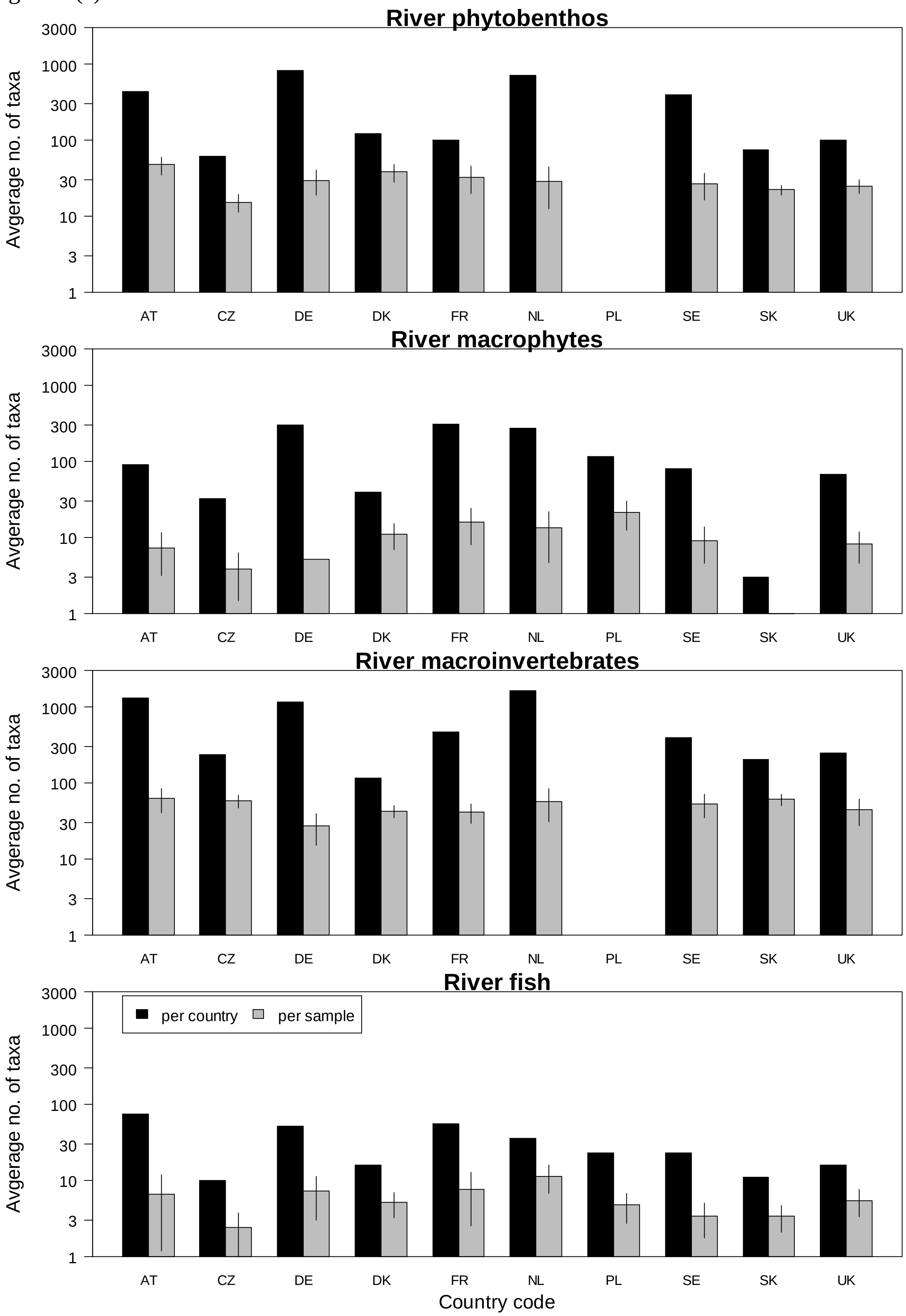

River phytobenthos
River macrophytes
River macroinvertebrates
River fish
per country
per sample
Avgerage no. of taxa
3000
1000
300
100
30
10
3
1
AT
CZ
DE
DK
FR
NL
PL
SE
SK
UK
Country code

Figure 4 (b)

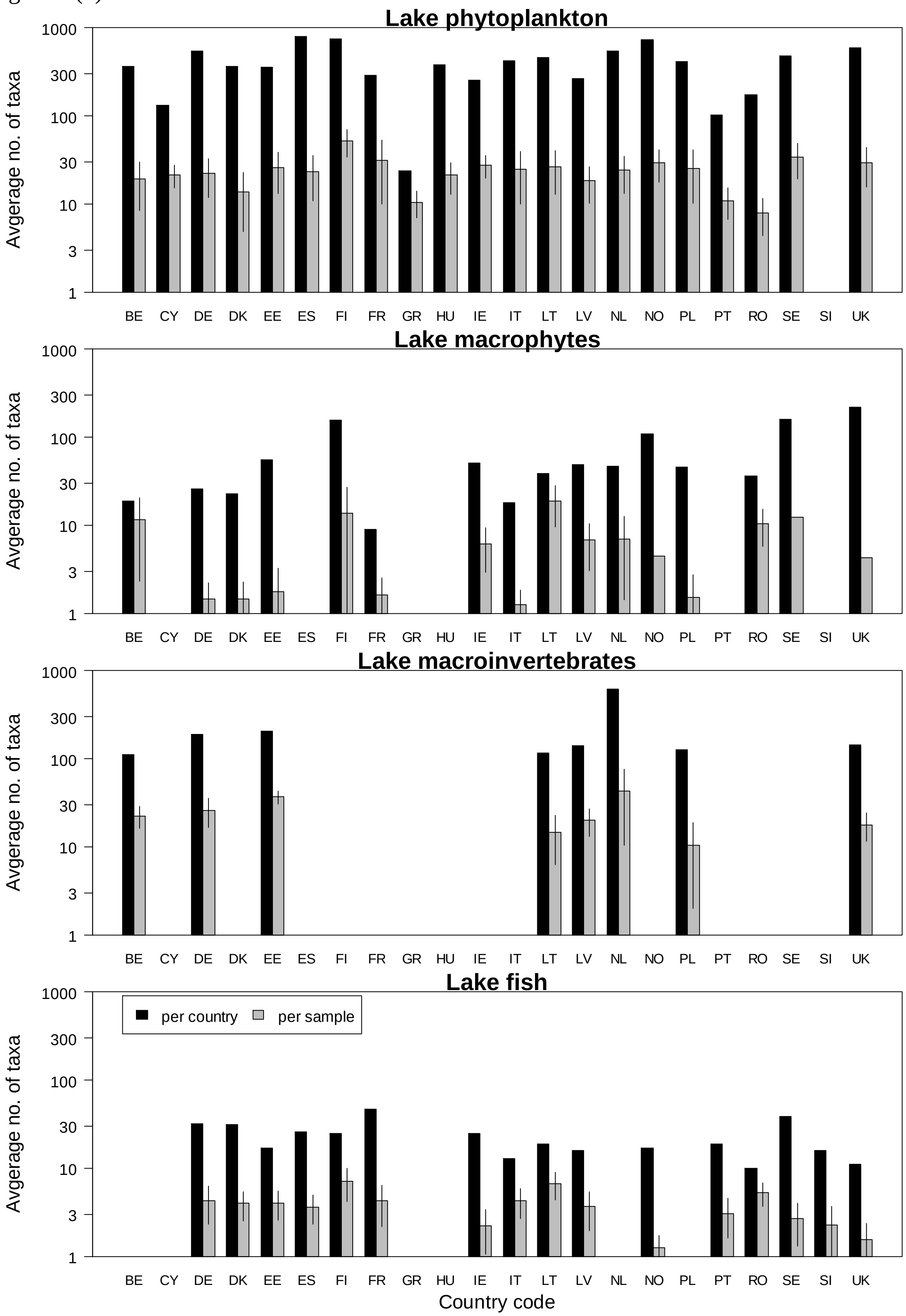

Lake phytoplankton
Avgerage no. of taxa
1000
300
100
30
10
3
1
BE CY DE DK EE ES FI FR GR HU IE IT LT LV NL NO PL PT RO SE SI UK
Lake macrophytes
Avgerage no. of taxa
Lake macroinvertebrates
Avgerage no. of taxa
Lake fish
per country
per sample
Avgerage no. of taxa
Country code

Figure 4 (c)

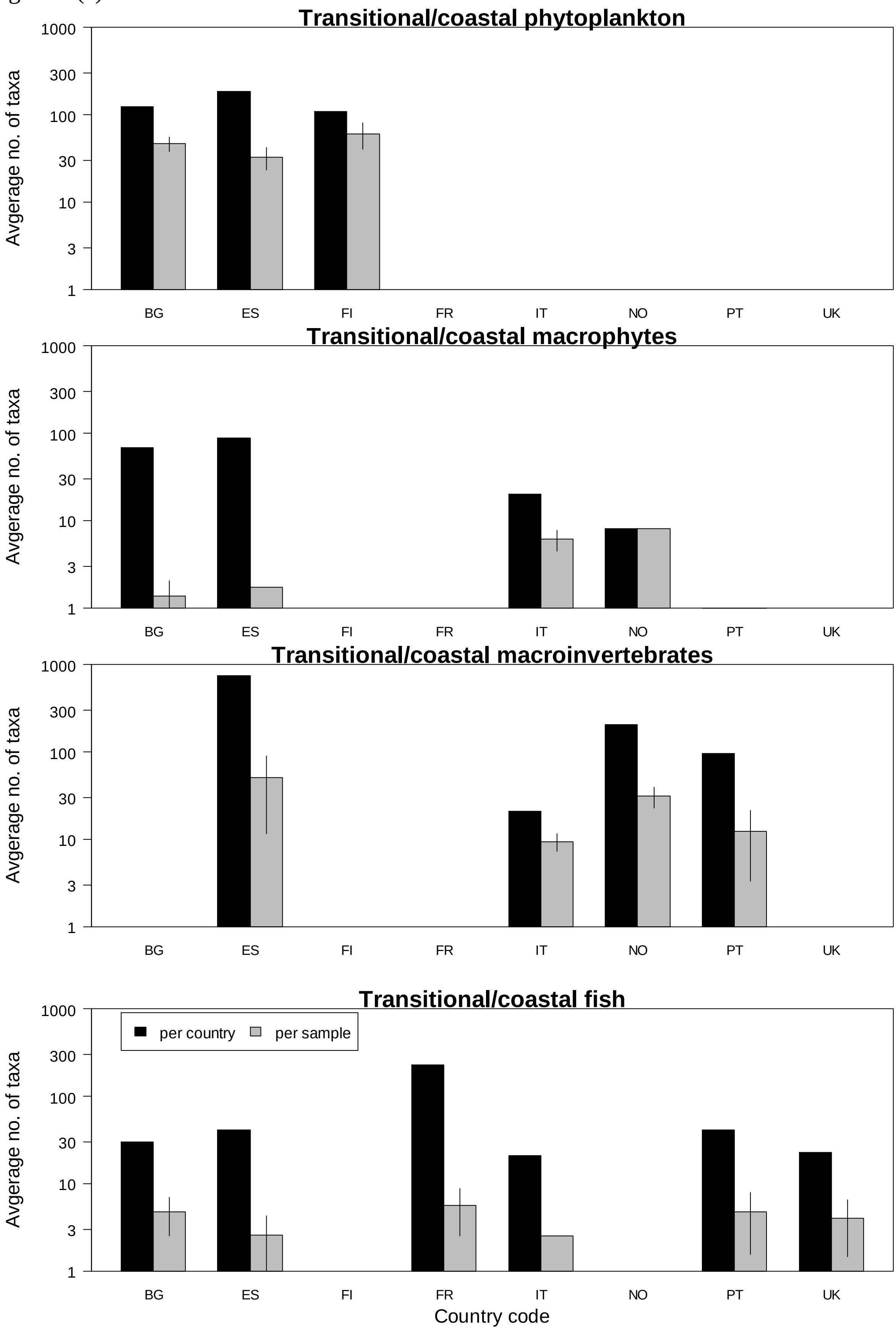

Transitional/coastal phytoplankton
Transitional/coastal macrophytes
Transitional/coastal macroinvertebrates
Transitional/coastal fish
Avgerage no. of taxa
1000
300
100
30
10
3
1
BG
ES
FI
FR
IT
NO
PT
UK
per country
per sample
Country code

Table 1 Summary content of the WISER central database: number of countries, waterbodies, stations, biological samples and records and environmental samples and records for each biological quality element (BQE). The counts of waterbodies and stations include only those containing biological samples (not including chlorophyll-a). The counts of environmental data include only waterbodies that contain biological data in the same WP database. (a) Background data from national monitoring data, previous research projects etc. (b) Foreground data from the WISER field campaign 2009-2010. Cited publications provide examples of the scientific use of each WP database. More information on the individual datasets constituting each WP database can be found at: http://www.wiser.eu/download/WISER_Dataset_IPR_overview.xls.zip

(a)

WP	BQE	Countries	Water-bodies	Stations	Biol. samples	Biol. records	Envir. samples	Envir. records	Data manager	Scientific publications
5.1	River phytobenthos	9	795	1 580	1 963	61 598	6 148	134 332	Andreas Melcher, Martin Seebacher	(Dahm et al., 2012; Feld et al., 2012; Haase et al., 2012)
	River macrophytes	10	683	1 959	2 557	25 927				
	River macroinvertebrates	9	1 380	3 281	4 911	217 501				
	River fish	10	805	2 247	2 617	17 376				
3.1	Lake phytoplankton[1)]	21	2063	2193	16 238	463 837	63 426	383 941	Birger Skjelbred, Geoff Phillips	(Järvinen et al., 2012; Maileht et al., 2012)
3.2	Lake macrophytes	12	1571	1 613	1 724	27 773	0	0	Bernard Dudley	(Mjelde et al., 2012)
3.3	Lake macroinvertebrates	8	179	628	870	23 016	0	0	Jürgen Böhmer	
3.4	Lake fish	16	2005	47 292	64 690	185 343	0	0	Stéphanie Pedron, Simon Causse	(Argillier et al., 2012)
4.2	Transitional/coastal macroalgae and angiosperms	2	32	62	1831	2 306	3	3	Rosa G. Novoa	(Mascaró et al., 2012)
4.4	Transitional/coastal fish	4	57	1 912	2778	17 003	3 022	14 366	Anne Courrat , Mario Lepage	(Alvarez et al., 2012)
Total		**26**	**6748**[2)]	**62 767**	**100 179**	**1 041 680**	**72 599**	**532 642**		

(b)

WP	BQE	Countries	Water-bodies	Stations	Biol. samples	Biol. records	Envir. samples	Envir. records	Data manager	Scientific publications
3.1	Lake phytoplankton[3)]	10	29	94	186	10 047	976	3 107	Birger Skjelbred, Jannicke Moe	(Carvalho et al., 2012)
3.2	Lake macrophytes	10	28	159	4 848	7 497	0	0	Bernard Dudley	(Dudley et al., 2012; Karus & Feldmann, 2012)
3.3	Lake macroinvertebrates	5	12	30	96	2 159	31	31	Oliver Miler, Mario Brauns	
3.4	Lake fish	3	14	310	430	1 587	0	0	Stéphanie Pedron	(Argillier et al., 2012)
4.1	Transitional/coastal phytoplankton	3	5	18	42	1 755	0	0	Karsten Dromph	(Dromph et al., 2012)
4.2	Transitional/coastal macroalgae and angiosperms	5	15	65	328	1 112	8 357	25 521	Rosa G. Novoa	(Marbà et al., 2012; Orfanidis et al., 2012)
4.3	Transitional/coastal macroinvertebrates	4	11	61	165	5 408	56	559	Karl Norling	(Borja et al., 2011)
4.4	Transitional/coastal fish	4	7	71	213	361	213	803	Anne Courrat	(Alvarez et al., 2012)
Total		**14**	**58**[2)]	**808**	**808**	**6 308**	**29 926**	**9 633**		

1) This database also contains background data on chlorophylla a from 6532 waterbodies, 10 090 stations and 72 823 samples.

2) The total number of waterbodies is lower than the sum across all WPs, because some waterbodies were recorded in more than one WP database.

3) This database also contains foreground data on chlorophylla a from 32 waterbodies, 103 stations and 237 samples