Article (refereed) - postprint

Schmidt-Kloiber, Astrid; Moe, S. Jannicke; **Dudley, Bernard**; Strackbein, Jörg; Vogl, Robert. 2013 The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters. *Hydrobiologia*, 704 (1). 29-38. 10.1007/s10750-012-1295-6

Astrid Schmidt-Kloiber[1], S. Jannicke Moe[2], Bernard Dudley[3], Jörg Strackbein[4], Robert Vogl[1]

The WISER metadatabase: the key to more than 100 ecological datasets from European rivers, lakes and coastal waters

[1] BOKU - University of Natural Resources & Life Sciences, Department Water, Atmosphere, Environment, Institute of Hydrobiology & Aquatic Ecosystem Management
Max Emanuel-Straße 17, 1180 Vienna, Austria

[2] NIVA, Norwegian Institute for Water Research, Gaustadalléen 21, NO-0349 Oslo, Norway

[3] CEH, Centre for Ecology & Hydrology
Bush Estate, Penicuik, Midlothian, EH26 0QB, United Kingdom

[4] UDE, University of Duisburg-Essen, Faculty of Biology, Applied Zoology/Hydrobiology
Universitätsstraße 5, D-45141 Essen, Germany

Corresponding author:
astrid.schmidt-kloiber@boku.ac.at
phone: +43/1/47654-5225
fax: +43/1/47654-5217

Abstract

In ecological sciences the role of metadata (i.e. key information about a dataset) to make existing datasets visible and discoverable has become increasingly important. Within the EU funded WISER project (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) we designed a metadatabase to allow scientists to find the optimal data for their analyses. An online questionnaire helped to collect metadata from the data providers and an online query tool (http://www.wiser.eu/results/meta-database/) facilitated data evaluation. The WISER metadatabase currently holds information on 114 datasets (22 river, 71 lake, 1 general freshwater and 20 coastal/transitional datasets), which also can be accessed by external scientists. We evaluate if generally used metadata standards (e.g. Darwin Core, ISO 19115, CSDGM, EML) are suitable for such specific purposes as WISER and suggest at least the linkage with standard metadata fields. Further, we discuss whether the simple metadata documentation is enough to reuse a dataset by others and why there is still reluctance to publish both metadata and primary research data (i.e. time and financial constraints, misuse of data, abandoning intellectual property rights). We emphasise that metadata publication has major advantages as it makes datasets detectable for other scientists and generally makes a scientist's work more visible.

Keywords

aquatic metadata, ecological databases, online query tool, data accessibility, intellectual property rights, Water Framework Directive

Introduction

Metadata is loosely defined as “data about data”. More specifically metadata should document and describe all aspects of the data (i.e. the who, why, what, when and where) that would allow understanding of the physical format, content and context of the data, as well as how to acquire, use and cite the data (Michener et al., 1997, Hey & Trefethen, 2003a; Michener, 2006; Michener & Jones, 2012). In that sense a metadatabase should gather information on datasets in order to allow data visibility and assessment. For the data producer/provider metadata are meant to document data for possible own future needs and to inform prospective users of their characteristics, while for the data consumer/user metadata are used to both discover data and assess their appropriateness for particular needs – their so-called “fitness for purpose”.

Basic and applied ecological research asks for the availability of “high-quality” data, the definition of which varies and depends on the specific purpose of a study. Scientists frequently reuse their own old data, use data created by others and/or share data within large work groups (e.g. within EU funded projects). But often these research datasets are not made publicly available nor deposited in permanent archives and therefore risk being lost over time (Shorish, 2010). Michener et al. (1997) suggest that metadata ideally should comprise all information that is necessary and sufficient to enable long-term secondary use (reuse) of the dataset by the original investigator(s), as well as use by other scientists who were not directly involved in the original research efforts. The scientific value of being able to reuse data and to utilise data for multiple objectives that may not have been foreseen by the data originator(s) may far exceed the perceived value associated with publications resulting from the original study (Michener et al., 1997). Thus, to document primary research datasets in metadata collections allowing people to discover and understand these data is an important step forward, even if the data are not stored in a public repository.

In ecological sciences the role of metadata in facilitating the scientist’s work has been recognised since the 1980s and several suggestions about ecological metadata and their management have been presented (Michener et al., 1987; Kirchner et al., 1995; Michener et al., 1997; Fegraus et al., 2005; Michener, 2006). Collecting metadata in dedicated databases is becoming more and more common, especially for biodiversity related datasets the importance of broad data compilations is being increasingly recognised. The Global Biodiversity Information

Facility (GBIF[1]), for example, collates and centralises not only primary biodiversity data but also offers standards and tools for (meta-) data collection (e.g. the GBIF Integrated Publishing Toolkit (IPT) or the specific GBIF metadata profile). Metadata standards such as FGDC Content Standard for Digital Geospatial Metadata (CSDGM)[2], ISO 19115[3], Darwin Core[4], Ecological Metadata Language (EML) Specification [5], Natural Collections Descriptions (NCD)[6] or Directory Interchange Format (DIF) of the Global Change Master Directory (GCMD)[7] are already in place or being further developed.

The many EU-funded projects over the last decade have produced a large bulk of data. Both legislation, including the Aarhus Convention[8], the OECD Principles and Guidelines for Access to Research Data from Public Funding[9] or the INSPIRE Infrastructure for Spatial Information in the European Community[10], and the rules of funding agencies of environmental studies request that these data be publically accessible. An important issue in the reuse of research datasets is the recognition of intellectual property rights of the data owners. This can include attaining permission from the data owner, acknowledgement of data creators and compilers, or even offers of authorship. Whitlock (2011) provides an useful overview on best practices in data archiving in ecology and evolution. He invites data collectors, data users as well as journals, editors and publishers, to embrace their important role in facilitating the transition of the ecology and evolutionary biology fields into one that gives more respect to the legacy of researchers' data. A well-elaborated and focussed metadatabase with properly documented intellectual property rights will certainly facilitate these aims.

The European Union project WISER[11] (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) was a large, multi-agency, collaborative exercise that targeted at supporting the implementation of the Water Framework Directive (WFD)[12]. As European surface waters are impacted by multiple environmental stressors and water uses, the WFD requires to improve the ecological status of all surface waters

---

[1] http://www.gbif.org/
[2] Federal Geographic Data Committee: http://www.fgdc.gov/metadata
[3] ISO 19115 Geographic information – Metadata: www.iso.org
[4] http://www.tdwg.org/activities/darwincore/
[5] http://knb.ecoinformatics.org/software/eml/
[6] http://www.tdwg.org/standards/312/
[7] http://gcmd.nasa.gov/User/difguide/difman.html
[8] http://www.unece.org/env/pp/welcome.html
[9] http://www.oecd.org/dataoecd/9/61/38500813.pdf
[10] http://inspire.jrc.ec.europa.eu/
[11] 7th Framework Programme, Theme 6 (Environment including Climate Change), contract No. 226273, www.wiser.eu
[12] http://ec.europa.eu/environment/water/water-framework/index_en.html

that do not meet the “good ecological status”. Targeting three main ecosystem types (rivers, lakes, coastal/transitional waters) and four organism groups (fish, macro invertebrates, macrophytes, phytoplankton) scientists in the WISER project developed assessment methods to identify the degree and causes of degradation, as well as methods to design recovery schemes and to predict their success.

The primary aim of the WISER metadatabase was to summarise available data from research in European surface waters, which could be provided by the project partners and to improve the WISER scientists’ options for discovery of data relevant to the project. This included the compilation of information about pre-existing datasets, new datasets derived from the WISER field exercises (Borja et al., 2012; Hering et al., 2012; Solheim et al., 2012), and datasets from ongoing monitoring programmes. More specifically, the metadatabase aimed to give an overview on (1) the general data availability per water category (i.e. rivers, lakes, coastal and transitional waters), (2) biological quality elements (i.e. fish, macro invertebrates, macrophytes, phytoplankton and phytobenthos), (3) geographical intercalibration groups (GIGs, i.e. geographical/political regions of Europe), (4) the availability of environmental data, (5) data comparability and data resolution (e.g. regarding the taxonomic level) and (6) the usability and accessibility of data (i.e. intellectual property rights of the datasets). The secondary aim of the metadatabase was to make the metadata for all of these data publicly available. The metadatabase is now a long-term resource for aquatic ecological researchers, which will facilitate appropriate data acquisition.

In this paper we illustrate the design and creation of the WISER metadatabase as well as its use for external scientists. We describe our experiences with metadata compilation and discuss the content of the database as well as the use of metadata standards. Finally we provide indications how to overcome the reluctance in publishing both metadata and data.

Building the WISER metadatabase

Design and development of the metadatabase

The construction of the WISER metadatabase started with the compilation of a standard set of information fields generally describing a dataset. The metadata fields were extensively reviewed, extended and complemented by WISER project partners to guarantee the availability of exactly that information that might be needed for their work. This resulted in ten main information blocks containing a variety of data characterising fields as illustrated in Table 1.

Populating the metadatabase

In order to populate the metadatabase we developed a questionnaire that was made available online. The user interface of this questionnaire was developed according to standard web technologies (using PHP as server-side and Javascript as client-side scripting languages). Entries into the questionnaire were saved into a MySQL database. To facilitate data entry for data providers most of the fields were designed as check boxes, radio buttons or selection lists. For additional information several comments-fields were made available. Further, a handbook providing supplementary information was compiled to give help while entering data.

The basis for the WISER metadatabase was a list of about 90 datasets, which WISER partners had announced as available for the project in the preparatory phase. Project partners were originally required to complete one questionnaire response for each dataset that they held or contributed. More than 20 datasets were additionally provided by the project partners and registered in the metadatabase in the course of the project, primarily lake data from the WFD intercalibration exercises carried out by the geographical intercalibration groups (GIGs). Finally, 8 datasets resulting from the WISER field campaign were included in the metadatabase.

The metadatabase query tool

The metadata query tool was made available via a web interface programmed in PHP using AJAX as communication technology between client and server. The tool accesses the MySQL questionnaire database. In its functions the tool was designed to help WISER scientists to find appropriate datasets for their analyses and methods development. Further it should serve to gain information about these datasets, especially about the intellectual property rights and the accessibility of the data through the WISER central database.

The query tool is available through the WISER website (http://www.wiser.eu/results/meta-database/) and can be used by all scientists and the interested public. It consists of ten main query blocks for specifying the search: (1) water categories (ecosystem types), (2) GIGs, (3) typological criteria, (4) ecoregions, (5) countries, (6) stressor types, (7) restored sites per stressor type, (8) biological quality elements, (9) sample types, and (10) seasons. After choosing a set of criteria four different result tables are immediately displayed: (1) a summary of the selected parameters, (2) a summary of available numbers of sites per selected criterion, (3) a summary about additional available information within the queried datasets and (4) a list of appropriate datasets. This latter list includes the dataset name and identification number as well as the technical and scientific contact persons and their contact information. From each dataset name there is a link to an overview presenting the entire questionnaire with all available information about this dataset. Additionally intellectual property right issues and the availability of datasets in the WISER central database are indicated with traffic light systems ranging from green lights for available and accessible datasets to red lights for restricted data.

Content of the WISER metadatabase

The WISER metadatabase currently describes 22 river, 71 lake and 20 coastal/transitional datasets; one dataset holds general information regarding all freshwater types (i.e. www.freshwaterecology.info), making a total of 114 datasets (Annex 1). Eight datasets were compiled based on the WISER field exercises, the remainder originated from previous EU-funded and national projects, as well as finalised and ongoing monitoring initiatives, hosted by both WISER partners and other collaborating institutions. In total, data providers registered 12,271 waterbodies and 39,865 sites for rivers, 15,612 waterbodies and 24,822 sites for lakes and 422 waterbodies and 7933 sites for coastal/transitional waters (Figure 1, Annex 1). Eight river, one lake and three coastal/transitional datasets did not provide numbers of waterbodies. Numbers of unique waterbodies and sites are actually slightly lower due to overlapping of some of the datasets. The WISER central database contains 56 of the datasets listed in the metadatabase. More information on availability of and accessibility to these data is given in Moe et al. (2012).

Experiences from developing and managing the WISER metadatabase

Are existing metadata standards suitable for such specific purposes as WISER?

Collecting metadata has already become a major task in environmental sciences (Costello, 2009; Whitlock, 2011), both for storing information on datasets, databases and data repositories, as well as for detecting appropriate data for scientific purposes.

The WISER metadatabase was primarily meant to facilitate the scientific work within the project. Together with other tools developed by the WISER data service team (Dudley et al., 2012; Moe et al., 2012) it should help project partners with their analyses. Therefore the fields chosen for the metadatabase questionnaire development were designed for finding appropriate and relevant data for WFD related work. After consultation with several existing metadata standards used in biology, ecology and related realms (e.g. Darwin Core, ISO 11915, CSDGM, EML etc.), it was clear that – due to the differences in purposes – the information necessary for WISER and EU water management differed from, and by far exceeded, the information recorded by these standards. The WISER metadatabase therefore used a combination of specifically created and adapted as well as general standard metadata fields. A future aim of the WISER metadatabase might be to link the main metadata fields to a standard set of information fields, such as proposed through the standards mentioned above. Dedicated metadata harvesting, search and retrieval tools – which bring together information from different metadatabases – could then help the scientists to detect suitable data. The Mercury system, for example, works across a range of metadata specifications and standards. It collects metadata and key data from contributing project servers distributed around the world and builds a centralised index (Devarakonda et al., 2009).

In our experience it is often necessary to store more information than standard metadata fields propose. Therefore we recommend, if metadata are collected in individual, more specific fields, at least the basic set of describing variables should be linked to metadata standards using, for example, EML to make a dataset comparable to others, detectable through metadata search tools and therefore useable for other scientists.

Which datasets are registered in the WISER metadatabase and how can they be used by other scientists?

In the WISER metadatabase (Annex 1) there is a clear difference in the number of available datasets per water category. There are fewer datasets available for rivers than for lakes, but the rivers datasets contain a greater number of sites. While for the marine domain meta information on many datasets is already compiled in the EurOBIS data portal (Vandepitte et al., 2011), to our knowledge this is the first time that datasets of previous large European projects dealing with freshwaters and transitional/coastal waters were registered in a common metadatabase and valuable information for further usage of these datasets was recorded. These included datasets from the AQEM[13], Star[14] and EFI+[15] projects for rivers and the REBECCA[16] project for lakes. There were

[13] www.aqem.de, FP5: contract number EVK1-CT1999-00027
[14] www.eu-star.at, FP5: contract number EVK1-CT 2001-00089

differences in completeness of the individual metadata, especially the number of waterbodies was not available for several datasets. On the other hand the metadata compilation included a large amount of information on site typology, the environment, stressors and/or physico-chemical parameters, which proved useful for WISER partners who used the metadatabase to search for datasets relevant for their tasks. The metadatabase was especially useful for those scientists dealing with analyses across different organism groups or different water categories, for example, the integrated assessment of multiple biological quality elements within waterbodies (see Angeler et al., 2012; Verdonschot & Keizer-Vlek, 2012).

The WISER metadata query tool was primarily developed to facilitate data searches for the WISER scientists. However, as this tool is now available online, it provides potential for the WISER metadatabase to become a widely used resource in applied water issues, well beyond the life of the WISER project. This potential is enhanced by the availability of information specifically useful to managers, such as geographical extent and inclusion of many publicly funded databases. The metadatabase and the query tool together also provide information about accessibility and the intellectual property rights, so they allow researchers to determine where to get data and the regulations for using this data. The WISER metadatabase will be maintained online as part of the WISER website, but there are no updates planned regarding additional content due to the end of funding. Nevertheless we plan to link essential freshwater parts of the WISER metadatabase to the BioFresh[17] metadatabase and data portal to guarantee future visibility of the existence of the datasets.

Is well-codified metadata documentation sufficient for the successful reuse of data?

In the WISER project we developed special metadata fields and features (additional to standard metadata fields) that were of particular use for WISER scientists and WFD related work. This approach may be incompatible with harmonised and unrestrained metadata exchange (Heidorn, 2008). Still, we question whether simple written data documentation fulfils the requirements needed for the successful reuse of data. For best success Edwards et al. (2011) suggest not only to consider static metadata products (static, definitive descriptions of data characteristics as e.g. in metadatabases), but to also include metadata processes – i.e. the communications about metadata – that might take place during requests for and conversations about data with the data providers. These

[15] http://efi-plus.boku.ac.at, FP6: contract number 044096
[16] www.rbm-toolbox.net/rebecca/, FP6: contract number SSPI-CT-2003-502158
[17] http://www.freshwaterbiodiversity.eu, FP7: contract number 226874; data portal: http://data.freshwaterbiodiversity.eu/

processes include direct, non-codified, ephemeral, and maybe also incomplete communication as important dynamic resources on the way to gain full knowledge of a dataset. Well-codified metadata products should substitute for direct contact with data producers and increase the precision of knowledge about usability of datasets for purposes for which they were not originally intended, or by people who did not participate in creating them. Edwards et al. (2011) claim that metadata products will only be powerful resources if metadata processes are also available. As metadata often remain incomplete and inaccurate (Wayne, 2005), metadata processes can stand in (Edwards et al., 2011). In projects such as WISER where scientists come together at least once a year these metadata processes are relatively easy to be followed up. However, under different circumstances – if datasets are used outside a consortium – it might become crucial to gain this undocumented information. We therefore recommend to individually contact data holders/providers before data usage for a personal conversation even if the data are well documented.

Why is it still so difficult to collect metadata and how can we improve the situation?

The advantage of metadata collection associated with a dataset is unquestioned and is already entrenched in the (ecological) scientific community (see meta data documentation of large data portals as GBIF, EurOBIS[18] for the marine and BioFresh for the freshwater realm). Despite there seems to be a change of culture, in our experience from WISER as well as other projects (mainly from the freshwater domain), many scientists are still reluctant to provide even metadata. The reasons for this reluctance can include:

(1) Provision of accurate information and sufficient documentation for the reuse of data requires time and therefore financial resources with which scientists are often not provided (Wayne, 2005; Shorish, 2010; Edwards et al., 2011). Many scientists regard this as money taken away from their budget to do “real” research (Hey & Trefethen, 2003b). To overcome this issue we suggest – in accordance with Wayne (2005) and Hey & Trefethen (2003b) – that funding agencies should make the creation of metadata and the placement of data generated by research projects into the public domain a condition of the grant and also provide additional resources for such activities. In that respect Beniston et al. (2012) suggest the implementation of guidelines or even a EU Directive on the good governance of data sharing including general rules on data formats and standards, data storage after a project’s completion or the general terms of access. Alternatively, one could think about incentives that outweigh these perceived costs. In social networking, for example, people provide amounts of information for no

[18] http://www.marbef.org/data/eurobis.php/

apparent personal gain. Still the researchers' main interest is using data, not describing them for the benefit of invisible and unknown future users, to whom they are not accountable and from whom they will receive little or no benefit (Edwards et al., 2011).

This leads to (2): The fear that data might be used for an incorrect purpose. By providing sufficient metadata and/or an accurate documentation that describes how data were collected and where they might have their limitations, it is up to the users of the data to be sufficiently competent to use and interpret them (Costello, 2009). We emphasise that publishing metadata has no effect on the "real" data itself (i.e. the primary research data), how – if at all – they are made publically available and how they might further be used. A metadatabase only makes the existence of a dataset visible, raises the awareness of other scientists of possibly available datasets and therefore also may serve a scientist in becoming established or forming collaborations.

(3) Finally, some scientists believe that releasing data – even metadata – means abandoning intellectual property rights (Heidorn, 2008). Within the WISER metadatabase we tried to address this issue by providing a comprehensive section dealing with these rights, specifying to whom data should be available (i.e. WISER partners only, GIGs, public) and criteria for using them, for example, whether the data provider must be offered co-authorship if the data are used. In the end respecting the intellectual property rights of other scientists' data and fulfilling the responsibility to refer to these data correctly should be scientific code of honour.

The fact that only 8 out of 114 datasets in the WISER metadatabase are defined as accessible for the public (compare Moe et al., 2012), illustrates the problem of general accessibility of environmental data. This demonstrates that there is much educational work still to be done on the way to open access to primary research data as suggested by the OECD or the Aarhus convention. In the molecular sciences open access to primary data is already common practice, as sequences must be submitted to GenBank prior to publication. Also for primary biodiversity data this intention is now supported by a variety of freshwater and ecology journals that encourage the deposition of these data in public repositories when publishing in their journals (De Wever et al., 2012). Similar efforts are undertaken by Dryad[19], an international repository of data underlying peer-reviewed articles in basic and applied biosciences, and other initiatives (see Penev et al., 2011, for an overview). Costello (2009) has extensively summarised the benefits of online publication of primary research data and Chavan & Penev (2011) advertise the "data paper" as a mechanism to incentivize data publishing. Papers connected to publicly available data get significantly more cited because the data become available for inclusion in broad-scale

[19] http://datadryad.org/

analyses (Piwowar et al., 2007). We generally support that view and in that sense we also think that publishing metadata in a dedicated (online) metadatabase and/or metadata paper makes research data more visible to other scientists. This will again help in creating more recognition of the scientists' work.

The authors, although acknowledging the difficulties in sharing of metadata, have hope that this situation is changing, as evidenced by the good practices used by many scientific disciplines.

Acknowledgements

We want to thank Jan Karud and Roar Brænden for their help with the initial development of the metadatabase. Further we are grateful to all WISER partners for contributing metadata. Thanks to Daniel Hering for coordinating the project and for valuable help during the metadatabase development and manuscript writing. We further want to thank Aaike De Wever for fruitful discussions and two anonymous reviewers for their comments. The WISER project was funded by the European Union under the 7th Framework Programme, Theme 6 (Environment including Climate Change), contract No. 226273.

References

Angeler, D., R. Johnson, D. Hering & S. J. Moe, 2012. Response of aquatic assemblages to human-induced change: use of variability within and among habitats in the design cost-effective monitoring programs. Hydrobiologia: this issue.

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.

Borja, A., M. Elliott, N. Marba & P. Hernriksen, 2012. Comparison of different biological quality elements for assessment of ecological status in coastal and transitional waters. Hydrobiologia: this issue.

Caroni, R. & W. van de Bund, 2012. Combination of BQEs in water body assessment. Hydrobiologia: this issue.

Chavan, V. & L. Penev, 2011. The data paper: a mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12 (Suppl 15): 2.

Costello, M. J., 2009. Motivating Online Publication of Data. BioScience 59 (5): 418–427.

Devarakonda, R., G. Palanisamy, B. E. Wilson & J. M. Green, 2010. Mercury: reusable metadata management, data discovery and access system. Earth Science Informatics, 3 (1-2), 87–94. doi:10.1007/s12145-010-0050-7.

De Wever, A., A. Schmidt-Kloiber, M. O. Gessner & K. Tockner, 2012. Freshwater Journals Unite to Boost Primary Biodiversity Data Publication. BioScience (in print).

Dudley, B., S. J. Moe, A. Schmidt-Kloiber & L. Carvalho, 2012. Extraction of data from WISER databases. Poster Presentation at the WISER Final Conference, Tallinn, Estonia, 25-26 January 2012. Book of Abstracts. ISBN 978-9949-484-19-5.

Edwards, P. N., M. S. Mayernik, A. L. Batcheller, G. C. Bowker & C. L. Borgman, 2011. Science friction: Data, metadata, and collaboration. Social Studies of Science 41 (5): 667–690.

Heidorn, P. B., 2008. Shedding Light on the Dark Data in the Long Tail of Science. Library Trends, 57 (2): 280–299. doi:10.1353/lib.0.0036

Hering, D., A. Borja, J. Carstensen, L. Carvalho, M. Elliott, A. Heiskanen, R. K. Johnson, S. J. Moe, D. Pont, A. L. Solheim, W. van de Bund & C. K. Feld, 2012. Assessment and restoration of European water bodies: A synopsis of the European WISER project. Hydrobiologia: this issue.

Hey, T. & A. Trefethen, 2003a. The data deluge: An e-science perspective. In: Berman F., G. Fox & A. J. G. Hey (eds), Grid Computing: Making the Global Infrastructure a Reality. Chichester. Wiley: 809–824.

Hey, T. & A. Trefethen, 2003b. e-Science and its implications. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 361 (1809): 1809–1825.

Kirchner, T., H. Chinn, D. Henshaw & J. Porter, 1995. Documentation standards for data exchange. In Ingersoll, R. & J. Brunt (eds), Proceedings of the 1994 LTER Data Management Workshop. Long-Term Ecological Research Network Office, University of Washington, Seattle, Washington, USA: 5–8.

Michener, W. K., R. J. Feller & D. G. Edwards, 1987. Development, management, and analysis of a long-term ecological research information base: example for marine macrobenthos. In Boyle, T.P. (ed), New Approaches to Monitoring Aquatic Ecosystems. ASTM STP, Vol. 940. American Society for Testing and Materials, Philadelphia: 173–188.

Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner & S. G. Stafford, 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7 (1): 330–342.

Michener, W. K., 2006. Meta-information concepts for ecological data management. Ecological Informatics 1 (1): 3–7.

Michener, W. K. & M. B. Jones, 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology & Evolution 27 (2): 85–93.

Moe, S. J., B. J. Dudley, A. Schmidt-Kloiber & D. Hering, 2012. The WISER way of organising ecological data from European rivers, lakes and transitional/coastal waters. Hydrobiologia: this issue.

Penev, L., D. Mietchen, V. Chavan, G. Hagedorn, D. Remsen, V. Smith & D. Shotton, 2011. Pensoft Data Publishing Policies and Guidelines for Biodiversity Data. Pensoft Publishers, http://www.pensoft.net/J_FILES/Pensoft_Data_ Publishing_Policies_and_Guidelines.pdf

Piwowar, H. A., R. S. Day & D. B. Fridsma, 2007. Sharing detailed research data is associated with increased citation rate. PLoS ONE 2: e308.

Shorish, Y., 2010. The Challenges to Data Sharing in the Sciences: Implications for Data Curation. Term paper, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Available at http://people.lis.illinois.edu/~shorish1/courses.html (accessed 05 December 2011).

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

Vandepitte, L., F. Hernandez, S. Claus, B. Vanhoorne, N. De Hauwere, K. Deneudt, W. Appeltans & J. Mees, 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.

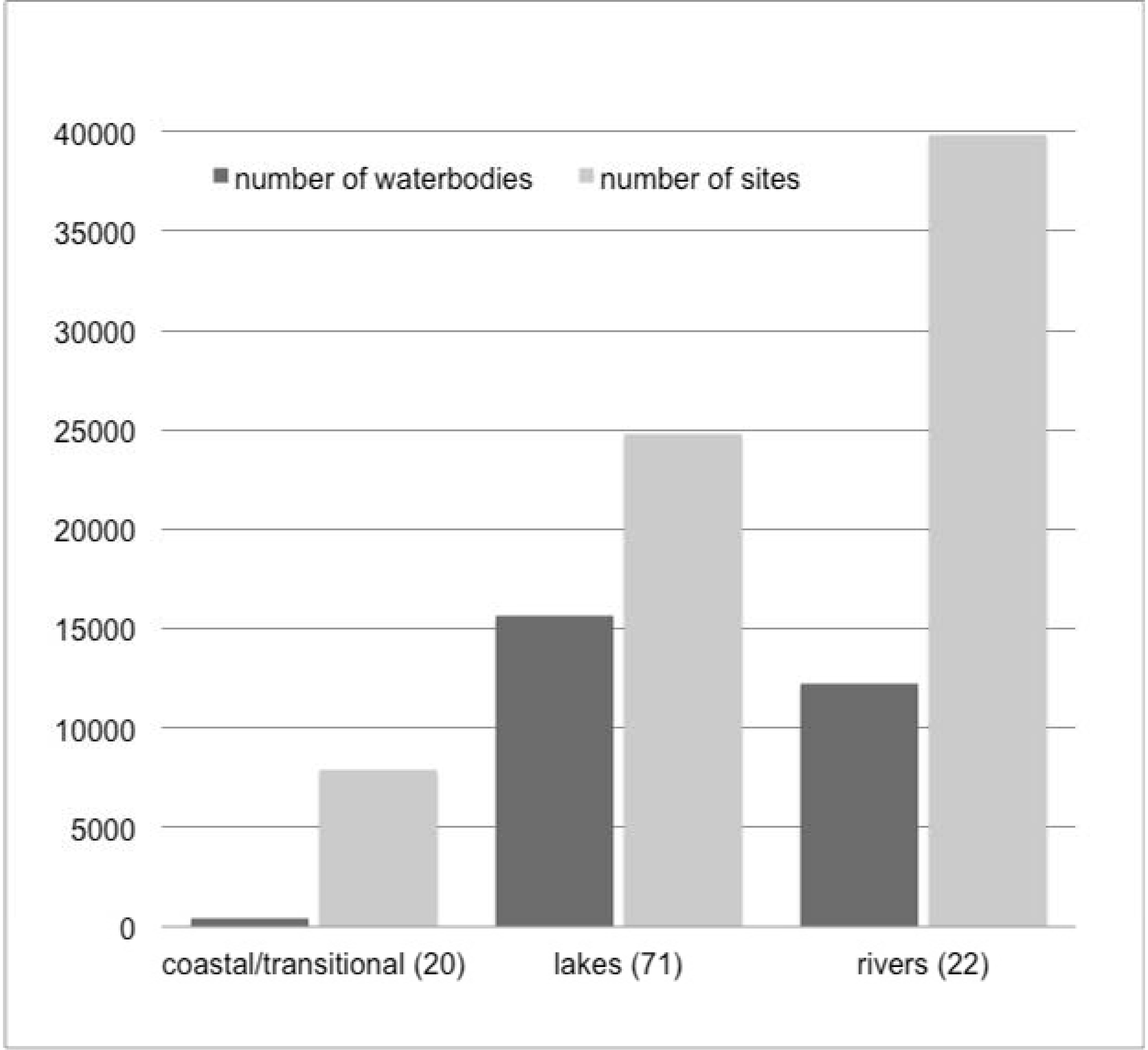
Verdonschot, P. F. M. & H. Keizer-Vlek, 2012. Response of different organism groups to pressure reduction in lakes, rivers and coastal waters. Hydrobiologia: this issue.

Wayne, L., 2005. Institutionalize metadata before it institutionalizes you. Federal Geographic Data Committee. Available at www.fgdc.gov/metadata/metadata-publications-list (accessed 06 February 2012).

Whitlock, M. C., 2011. Data archiving in ecology and evolution: best practices. Trends in Ecology & Evolution 26 (2): 61–65.

Figure captions

Fig. 1 Number of waterbodies and sites per water type (number of datasets in parenthesis) as registered by the data providers in the WISER metadatabase. Note that for 3 coastal/transitional, 1 lake and 8 river datasets no numbers of waterbodies were specified



40000
35000
30000
25000
20000
15000
10000
5000
0
number of waterbodies
number of sites
coastal/transitional (20)
lakes (71)
rivers (22)