Deliverable D2.2

Version: FINAL
Date: 27/04/11
Author: BOKU
Dissemination status: PU
Document reference: Deliverable_D2.2

Metadatabase of selected biodiversity databases (M1-18)

STATUS: FINAL VERSION

Project acronym:	BIOFRESH
Project name:	Biodiversity of Freshwater Ecosystems: Status, Trends, Pressures, and Conservation Priorities
Call and Contract:	FP7-ENV-2008-1
Grant agreement no.:	226874
Project Duration:	01/11/2009 – 30.04.2014 (54 months)
Co-ordinator:	Leibniz-Institute of Freshwater Ecology and Inland Fisheries at Forschungsverbund Berlin e.V., Germany
Partners:	RBINS, Royal Belgian Institute of Natural Sciences, Belgium
	BOKU, Universität für Bodenkultur Wien, Austria
	ICLARM, International Center for Living Aquatic Resources Management, Malaysia
	IRD, Institut de Recherche pour le Développement, France
	UDE, Universität Duisburg-Essen, Germany
	IUCN, International Union for Conservation of Nature, Switzerland
	UOXF.AC, Oxford University, UK
	UB, Universitat de Barcelona, Spain
	UFZ, Helmholtz Zentrum für Umweltforschung, Germany
	UCL, University College of London, UK
	EAWAG, Eidgenössische Anstalt für Wasserversorgung, Abwasserreinigung und Gewässerschutz, Switzerland
	UCBL, Université Claude Bernard - Lyon 1, France
	UPS, Université Paul Sabatier- Toulouse 3, France
	ECOLOGIC, Ecologic GmbH Institut für Internationale und Europäische Umweltpolitik, Germany
	EC-ERC, Commission of the European Communities - Directorate General Joint Research Centre, Italy
	UD, University of Debrecin, Hungary
	NRM, Naturhistoriska riksmuseet, Sweden

BIOFRESH

Biodiversity of Freshwater Ecosystems: Status, Trends, Pressures, and Conservation Priorities

Project no. 226874

Large scale collaborative project

Deliverable number	D2.2
Deliverable name	Metadatabase of selected biodiversity databases, 1st period (Months 1-18)
WP no.	WP 2
Lead Beneficiary (full name and Acronym)	University of Natural Resources and Life Sciences, BOKU
Nature	Database (DB), written report
Delivery date from Annex I (project month)	M18
Delivered	2011-04-30
Actual forecast delivery date	
Comments	

Project funded by the European Commission within the Seventh Framework Programme 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 is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union.

Name of the Authors	Name of the Partner	Logo of the Partner
Schmidt-Kloiber Astrid Vogl Robert	BOKU	

In case the report consists of the delivery of materials (guidelines, manuscripts, etc.)

Delivery name	Delivery file name	From Partner	To Partner

Introduction

This deliverable corresponds to the Task 2.1 of WP2: Review of existing biodiversity databases and generation of a metadatabase. This reports summarises the work so far, introduces the structure and the functionality of the metadatabase as well as the current content. Two further reports on the metadatabase will follow in month 36 and 51.

Aim and approach of the metadatabase

Metadata is loosely defined as “data about data”. A metadatabase therefore should gather information on datasets in order to allow data visibility and assessment. For the data producer/provider metadata are meant to document data 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’.

Regarding BioFresh, the aims of the metadatabase are

- to support the data portal (WP1) in terms of availability, format and quality of datasets
- to help WP3 with the identification of gaps
- to support the science WPs (WP4-7) in terms of availability and comparability of datasets for analysis

The metadatabase consists of two main parts:

- an online questionnaire to fill in data
- an online query page to find data

Both parts can be reached from the main BioFresh website using the menu: results/metadatabase or following this direct link: http://www.freshwaterbiodiversity.eu/index.php/metadatabase.html.

The construction of the BioFresh metadatabase was based on the experiences gained during the development of other metadatabases within the framework of other EU funded projects, e.g. WISER (www.wiser.eu). In a first step biodiversity relevant fields of the WISER metadatabase were used for the construction of the first draft online metadata questionnaire. In a second step these fields were extended with “standard” metadata fields of other relevant biodiversity databases (e.g. GBIF) and those proposed by the GCMD-DIF (Global Change Master Directory, Directory Interchange Format). In a final step fields that reflected the needs of the BioFresh science Workpackages (WP4-7) were included.

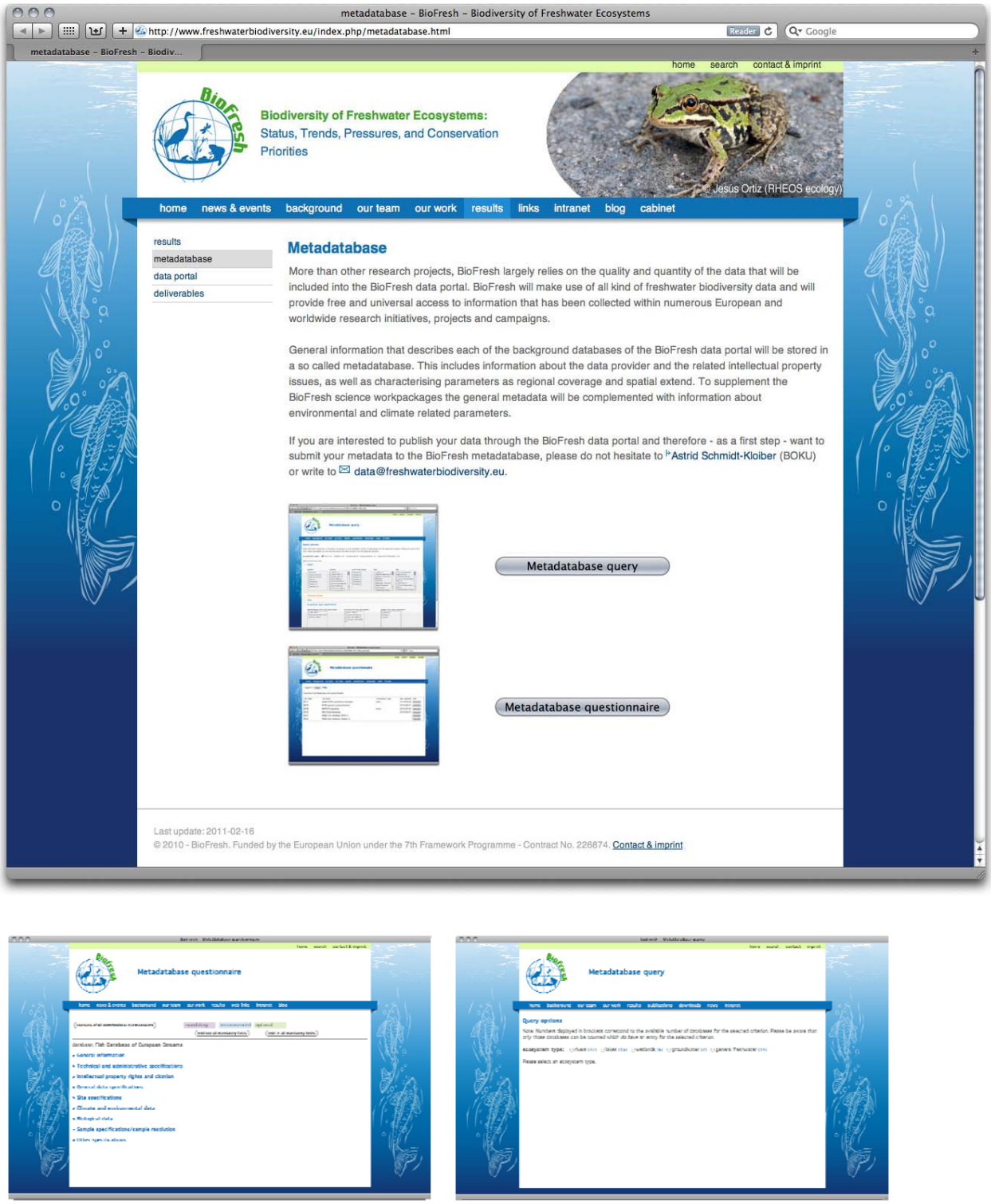

Figure 1. Screenshots of the metadatabase website; top: general overview site; bottom left: starting page of the questionnaire; bottom right: starting page of the query tool.

The metadata questionnaire

Structure of the metadatabase

The final BioFresh metadatabase and the accompanying questionnaire consist of nine main information blocks containing a variety of data characterising fields. The information blocks include:

- general information, e.g. database name, database type, keywords
- technical and administrative specifications, e.g. data format, operating system, data language, update level, contact details
- intellectual property rights and citation, e.g. database owner, citation, criteria for using the data
- general data specifications, e.g. regional coverage, countries, world climatic regions, freshwater ecoregions, ecosystem type
- site specifications, e.g. coordinate system, ecosystem type classification, number of sites, site coding
- climate and environmental data, e.g. climate related data, environmental data, physico-chemistry data, stressors
- biological data, e.g. data origin, organism groups addressed
- sample specifications/sample resolution, e.g. sample information, taxonomic resolution, taxonomic coding, sample specifications
- other specifications, e.g. GIS layers, photos, maps, quality control procedures

Different information blocks are displayed depending on the database type. The following database types can be chosen from a selection list when filling the questionnaire:

- species distribution database
- species (taxonomic group) per site database including environmental information
- biological or ecological traits/information database
- species/taxa database (e.g. red list database)
- environmental characteristics database
- climatic information database
- protected area database
- museum collection data (historical data)
- palaeoecological database
- data evaluation database ("results" database)
- other database type

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 are placed at disposal. The fields in the questionnaire are categorised into mandatory, recommended and optional fields. Further, tooltips are provided when moving the mouse over a certain field to give help while entering data.

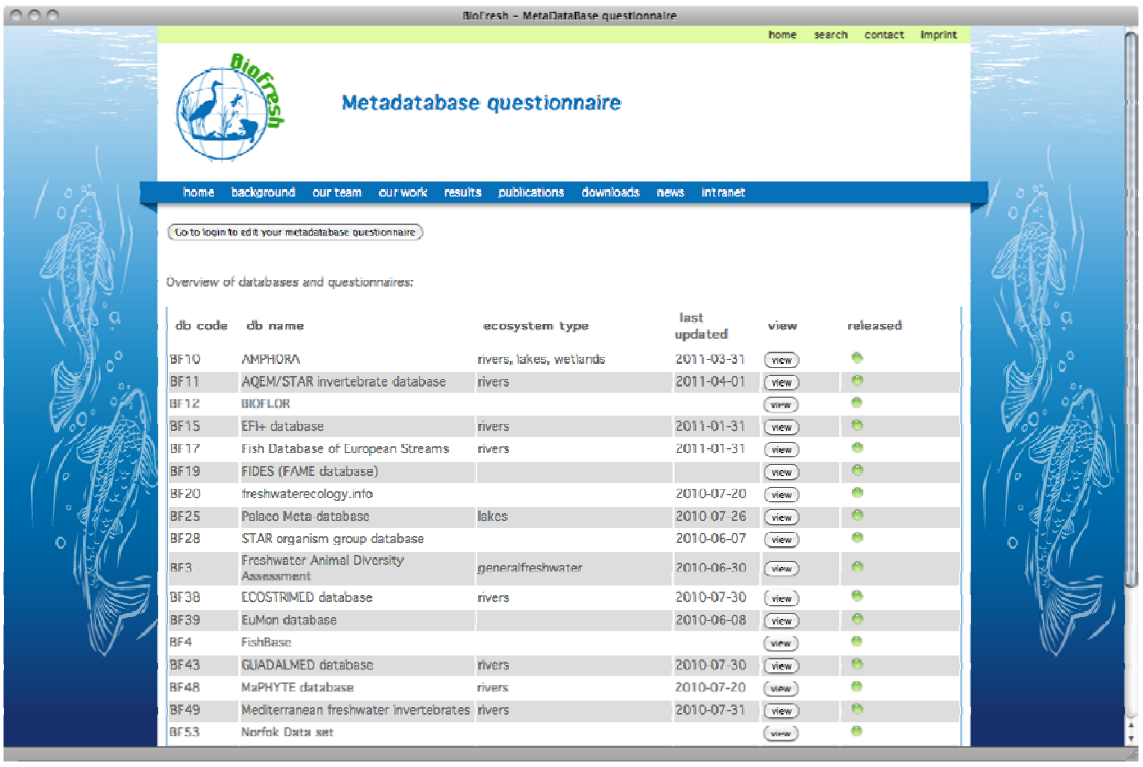

Figure 2. Overview page of the available metadata questionnaires.

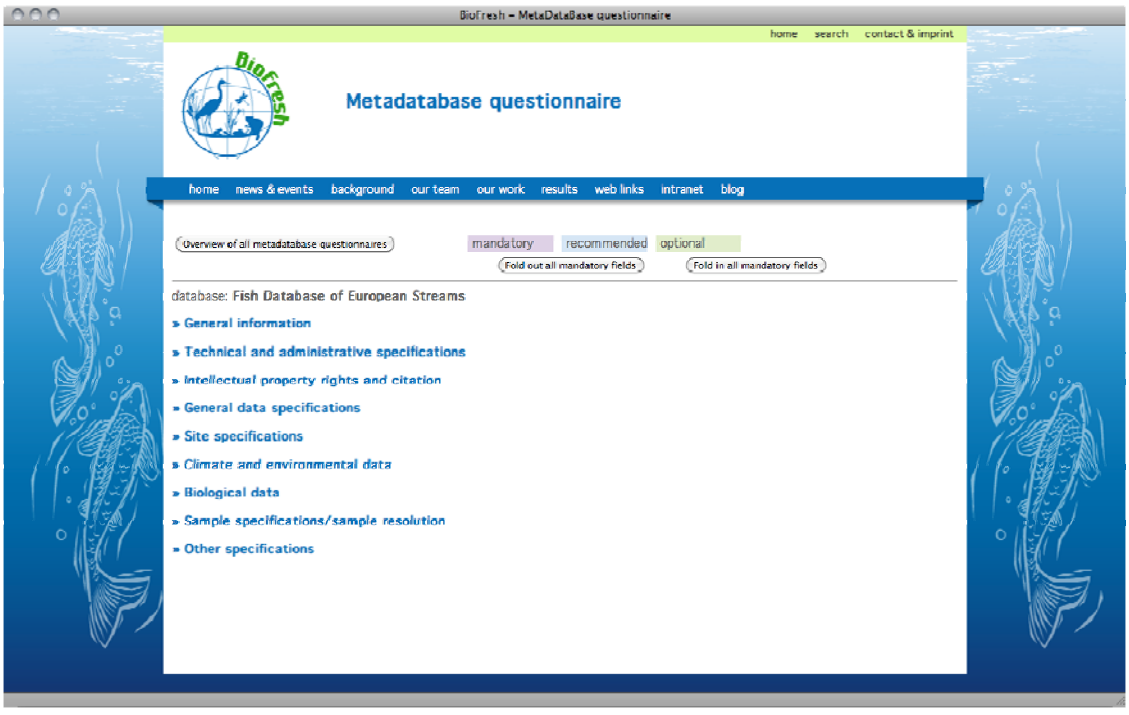

Figure 3. Overview of the main information blocks within the metadata questionnaire.

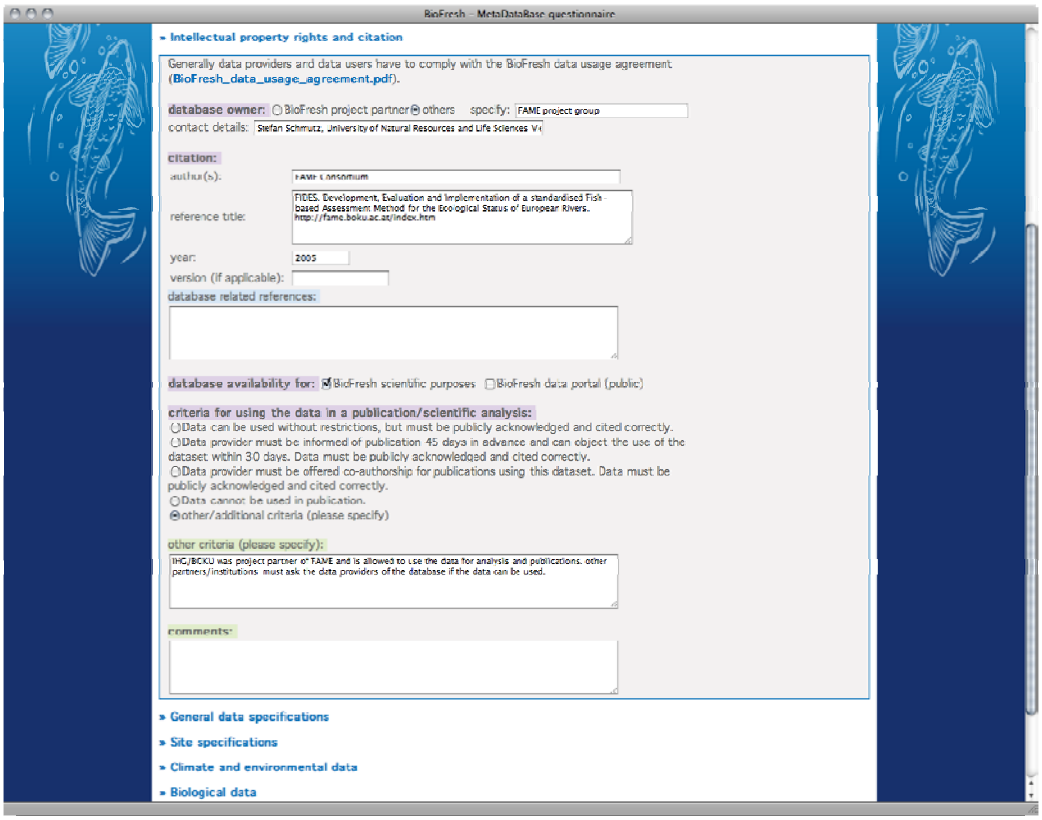

Figure 4. Detailed view of the information block "Intellectual property rights and citation".

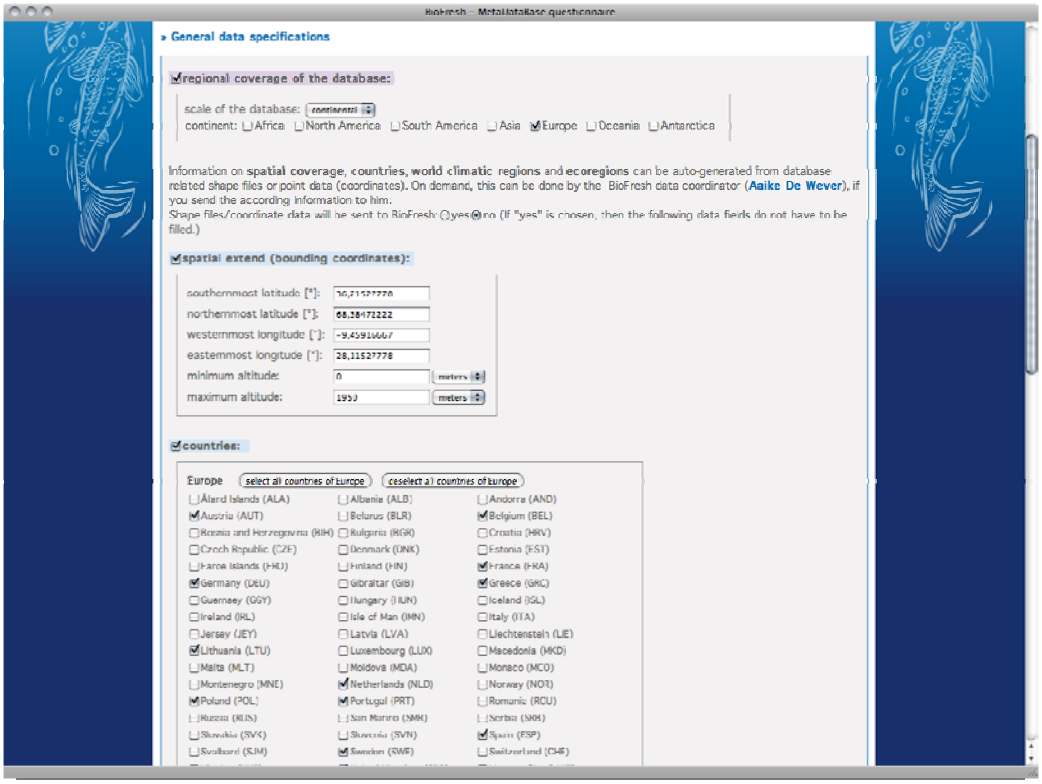

Figure 5. Detailed view of the information block "General data specifications".

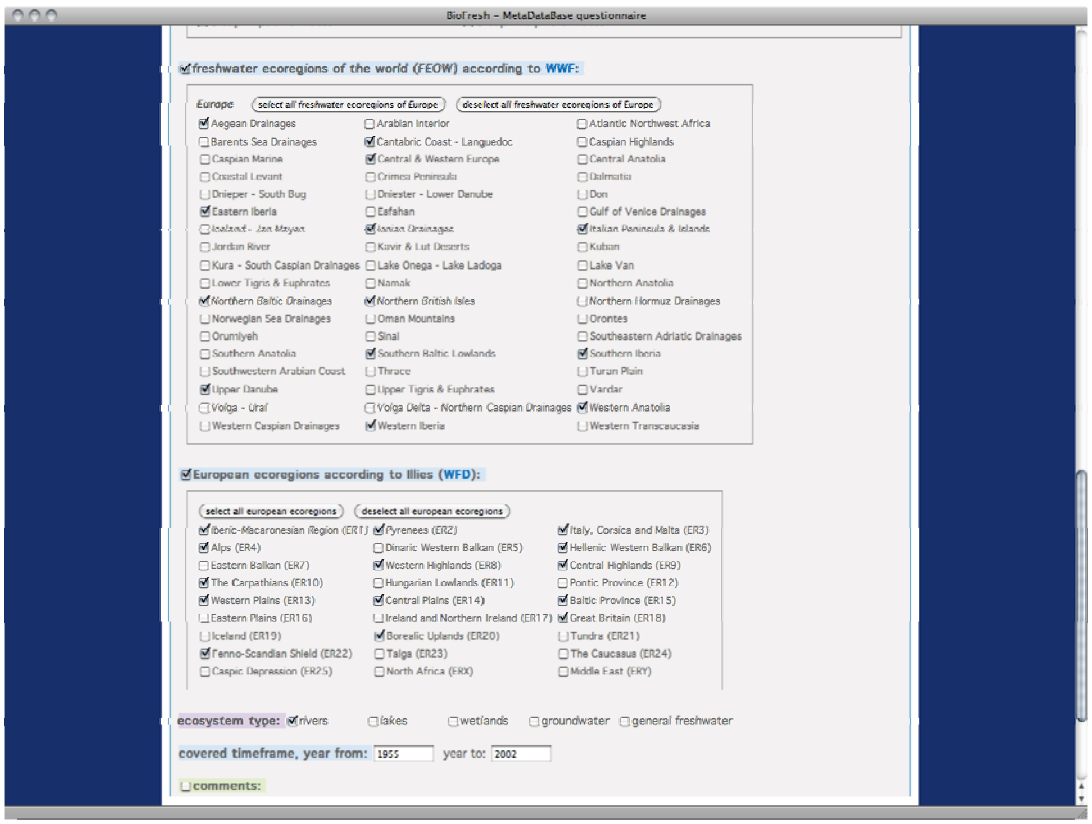

Figure 6. Detailed view of the information block "General data specifications".

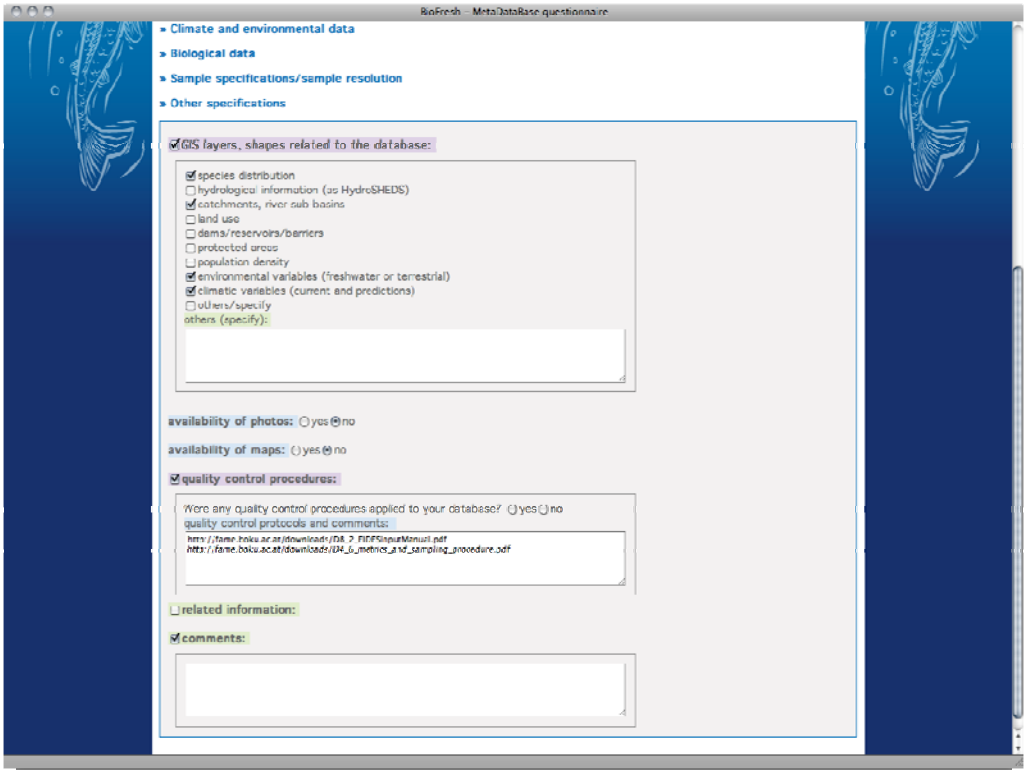

Figure 7. Detailed view of the information block "Other specifications".

The metadatabase query tool

Usability of the metadatabase

The metadata query tool is available via a web interface and should help BioFresh scientists as well as the interested public to find appropriate datasets and to gain information about these datasets.

The tool consists of six main query blocks for specifying the search:

- ecosystem type
- region
- organism group
- time
- ecosystem type classification
- climate and environmental data

The tool is designed as a semi-dynamic query immediately displaying a result table of appropriate datasets after applying the query options. During the query process only available datasets per criterion are displayed. In future the availability of the respective dataset through the BioFresh portal will be indicated in the result table using a traffic light system.

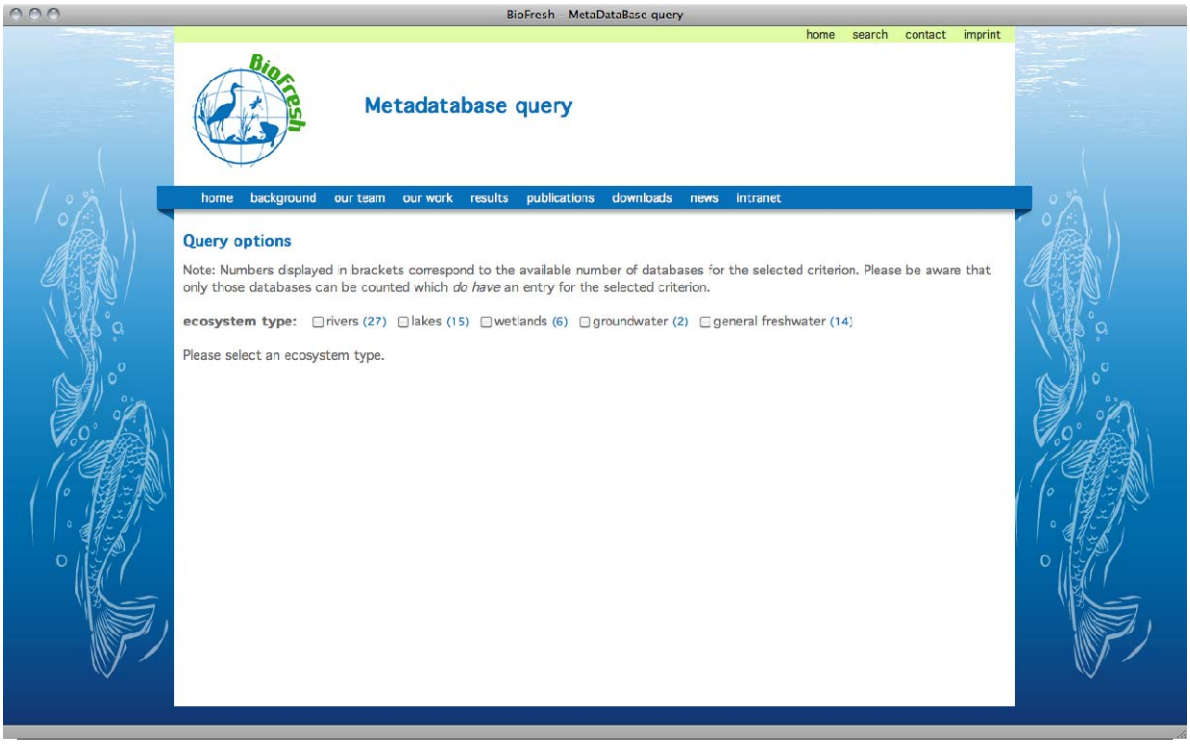

Figure 8. Overview of the BioFresh metadatabase query tool, showing the first query option on the ecosystem type.

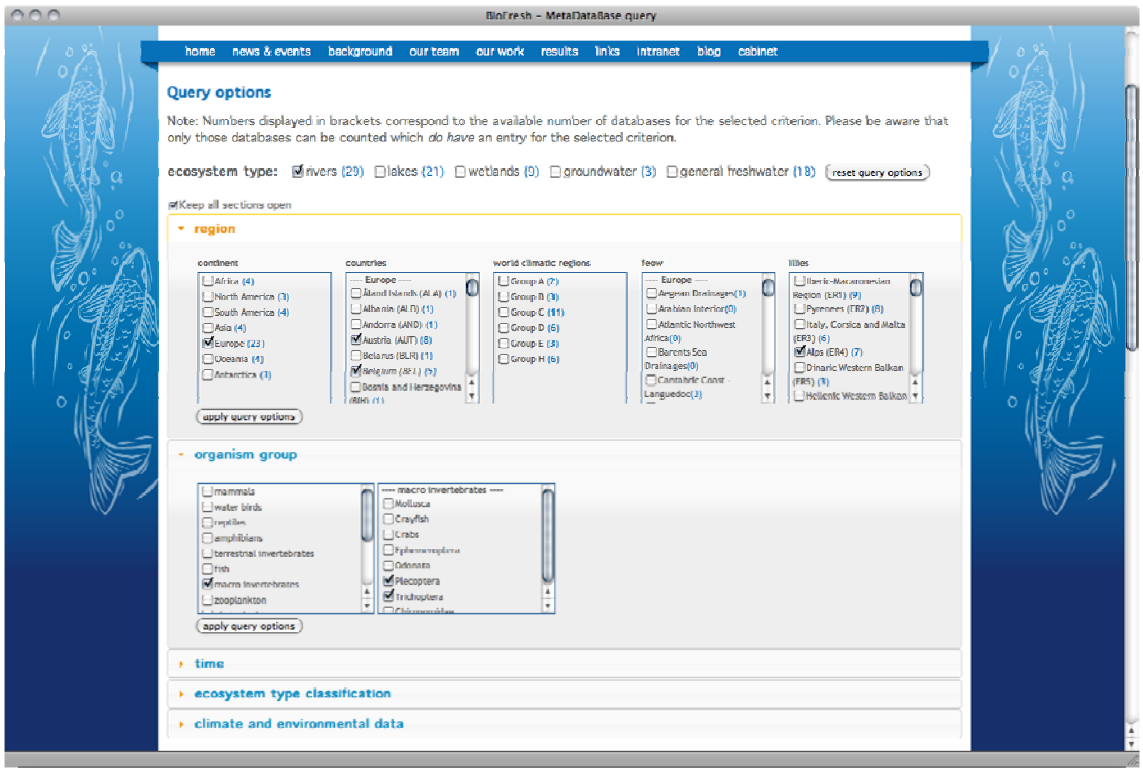

Figure 9. Additional query options for the region and the organism group.

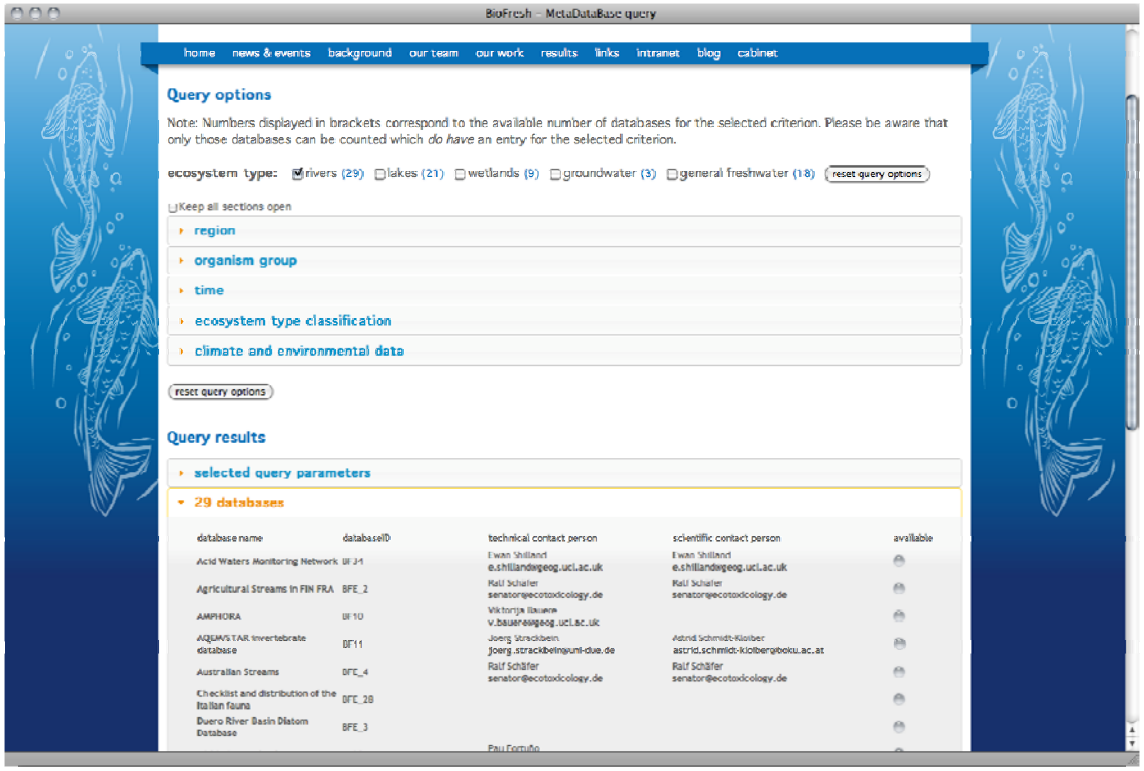

Figure 10. Result table of available datasets corresponding to the query options. The traffic light system indicates the availability through the BioFresh portal.

Quality control of the metadatabase

Data providers who want to add metadata to the BioFresh metadatabase receive a specific login and password to prevent misuse and accidental data manipulation.

Quality control of entered metadata is done in two different ways:

- The WP1-3 quality control team checks the entered data for completeness and reliability of mandatory fields and addresses the data provider if problems are detected. The quality control team directly corrects spelling and other minor errors. All quality control procedures are documented in a specially designed quality control protocol.
- Data provider themselves have the opportunity to define if and when their metadata are ready for release. This can be done with a ready-for-release-button in the internal section of the data provider.

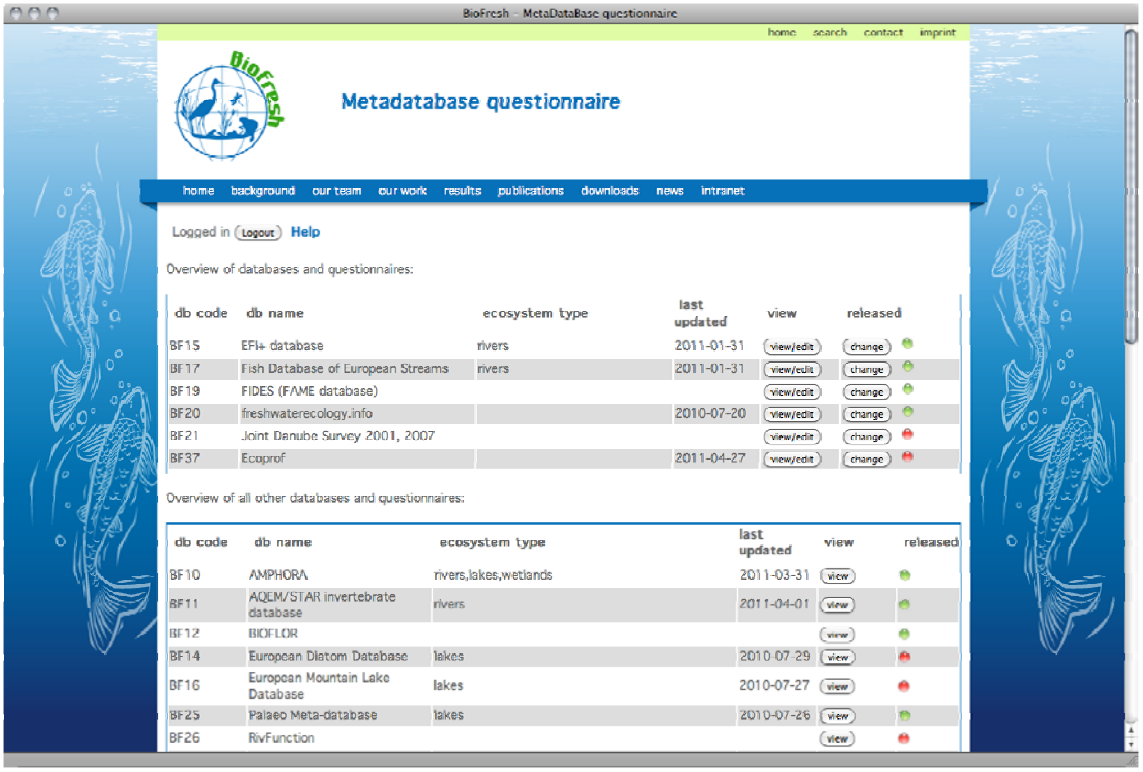

Figure 11. Internal section of a data provider. The upper part shows the datasets of this data provider and offers the possibility to change the release status. The lower part shows all other BioFresh metadata.

Content of the metadatabase

Currently 93 datasets are registered in the BioFresh metadatabase (Table 1). They are classified into three categories:

- internal datasets: owned, managed or maintained by BioFresh partner
- external datasets: metadatabase information acquired from external data holders
- digitised datasets: gained through digitisation of literature performed by WP3

External data providers either contact the BioFresh WP1-3 team themselves or – in case that a BioFresh partner detects an interesting dataset via web search, scientific paper, etc. – are contacted by the WP1-3 team. In the latter case, the WP1-3 team enters basic metadata information into the metadatabase. Subsequently an adjusted standard invitation letter is sent to the data holders asking to document and promote their database. The data holders then only need to check and amend the metadata.

Table 1. List of the 93 datasets registered in the metadatabase so far.

Dataset Code	Dataset Name
BFE_2	Agricultural Streams in DEU FIN FRA
BFE_39	Alpine GIG Lakes Macoinvertebrates data
BFE_34	Alpine GIG Lakes Macrophytes data
BFE_30	Alpine GIG Lakes Phytoplankton data
BF10	AMPHORA
BF11	AQEM/STAR invertebrate database
BFE_23	Australian Faunal Directory
BFE_4	Australian Streams
BF34	AWMN
BF12	BIOFLOR
BFE_17	Catalogue of Diatom Names
BFE_41	Central Baltic GIG Lakes Macoinvertebrates Chironomid Pupae data
BFE_40	Central Baltic GIG Lakes Macoinvertebrates littoral data
BFE_35	Central Baltic GIG Lakes Macrophytes data
BFE_31	Central Baltic GIG Lakes Phytoplankton data
BFE_49	Central Baltic GIG Rivers Macrophytes data
BFE_28	Ckmap - Distribution of the Italian Fauna
BFE_46	Cross GIG GIG Lakes Fish data
BFE_47	Cross GIG GIG Lakes Phytobenthos data
BFE_48	Cross GIG GIG Rivers Fish data
BFE_51	Cross GIG GIG Rivers Large Rivers data
BFE_3	Duero River Basin
BFE_45	Eastern Continental GIG Lakes Macoinvertebrates data
BFE_38	Eastern Continental GIG Lakes Macrophytes data
BFE_29	Eastern Continental GIG Phytoplankton data
BF37	Ecoprof
BF14	EDDI
BF15	EFI+ database
BFE_6	Elbe Monitoring 2006
BF16	EMERGE database
BFE_27	Ephemeroptera of the World
BF39	EuMon database
BF70	European river catchment database
BF17	FAME database
BFE_21	Fauna Iberica
BFE_25	FaunaItalia
BF19	FIDES (FAME database)
BFE_11	Fish Atlas for Germany and Austria
BF4	FishBase
BF67	FishBase occurrence data
BFE_14	Flemish sport fisheries data
BF69	Fonix
BF3	Freshwater Animal Diversity Assessment
BFE_12	Freshwater Fishes of Colombia
BFE_5	Freshwater fishes of India
BF20	freshwaterecology.info
BFE_16	Friedrich Hustedt diatom collection
BFE_10	German Dragonfly Atlas
BF5	GISD
BFE_26	Global Runoff Data
BF43	GUADALMED database
BF_dig1	Illies Fauna Europaea

Dataset Code	Dataset Name
BFE_15	Intercalibration dataset France
BF72	IUCN environmental data
BF6	IUCN Species Information Service
BF21	Joint Danube Survey 2001, 2007
BFE_9	Lake Balaton
BFE_8	Lake Vortsjarv
BF22	LEDA Traitbase
BF47	Long-term time-series metadatabase
BF48	MaPHYTE database
BF49	Mediterranean freshwater invertebrates
BFE_44	Mediterranean GIG Lakes Macoinvertebrates data
BFE_37	Mediterranean GIG Lakes Macrophytes data
BFE_33	Mediterranean GIG Lakes Phytoplankton data
BF71	Mediterranean Iberian Trichoptera database
BF53	Norfok Data set
BFE_42	Northern GIG Lakes Macoinvertebrates acidification data
BFE_43	Northern GIG Lakes Macoinvertebrates eutrophication data
BFE_36	Northern GIG Lakes Macrophytes data
BFE_32	Northern GIG Lakes Phytoplankton data
BFE_50	Northern GIG Rivers Macoinvertebrates acidification data
BFE_22	Opiebenthos
BF25	Palaeo Meta-database
BFE_13	PanTHERIA
BF65	Proasellus distributional data base
BF57	RIVA-HABEX database
BF26	RivFunction
BFE_1	SANTANDER
BF27	Shelled Gastropoda of Western Europe
BF59	SIX database
BF8	SPRICH
BF28	STAR organism group database
BF29	Statzner et al. 2007
BFE_7	Swedish lakes and Streams
BFE_24	Swiss Cartographic data on Fauna
BFE_20	Taxa Watermanagement the Netherlands
BFE_19	Terrestrial Ecosystem Monitoring Sites
BF60	UBA Projectdatabase
BF31	WFD Intercalibration database
BF62	WISER lake database (Module 3)
BF61	WISER river database (WP5.1)
BFE_18	World Lake Database

Outlook

Future activities will include the extension of the metadatabase query tool with a full text search and optimisation of existing features. A continuous enlargement of the number of datasets within the metadatabase is envisaged by further approaching external data holder as well as by linking with other relevant metadata initiatives.