Deliverable 3.1

Version: Version 3 Final
Date: 2014/04/30
Author: ICLARM
Dissemination status: PU
Document reference: Deliverable D3.1._M45

Analysis reports of the consultations, including a priority list of data resources to mobilise in the different stages of the project, and data to be encoded (M45)

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 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 FIN, FishBase Information and Research Group, Inc.

BIOFRESH

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

Project no. 226874

Large scale collaborative project

Delilverable number	D3.1
Delilverable name	Analysis reports of the consultations, including a priority list of data resources to mobilise in the different stages of the project, and data to be encoded
WP no.	WP3
Lead Beneficiary (full name and Acronym)	WorldFish Center (=ICLARM) [Aquatic Biodiversity Informatics Office, Philippines]
Nature	Written report
delivery date from Annex I (proj. month)	45
Delivered	2014-04-30
Actual forecast delivery date	
Comments	Delivery delayed to include requirements by Policy makers.

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 has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 226874

Name of the Authors	Name of the Partner	Logo of the Partner
Bailly Nicolas	WorldFish (formerly ICLARM)	
De Wever Aaike	RBINS	

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

Delivery name	Delivery file name	From Partner	To Partner
Report on the data requirements and the gap analyses (this document)	BF_WP3_D3.1c_DataRequirements_ Report1_140430_v4_final.doc	WorldFish, RBINS, and All	All

General Introduction

This deliverable corresponds to the Task 3.1 of the WP3 to analyse the data and information requirements from the research workpackages WPs 4-7. We have re-analyzed the data requirements collected during the first period of the project in the light of the contingency funded projects. This is the third report under the task T3.1.

The first two internal iterations of this report were the basis of the gap analysis report (D3.3, M20). However, the reviewers of the first mid-term review held in April 2012 stated that this data gap analysis focused on scientific requirements just confirmed what was known before and that our following recommendations were not new (they were reported in the gap analysis report.

Recommendation: **recommend future funding for massive data encoding efforts.**

Recommendation: **Important need for funding the establishment/extension of authoritative freshwater species lists including an exhaustive overview of "synonyms".**

Recommendations:
- recommend European policies on IPR issues related to aggregated data.
- explore the options for a standardized embargo period after which publicly funded scientific data should be released.

One can argue that if the issues were spotted some time ago and are still flagged in more recent studies, it is because no good solutions were found yet or that previous recommendations were not followed, and still need to be addressed.

Anyway, reviewers requested that we perform the same analysis from a policy maker point of view, a demand driven data gap analysis and develop data mobilization strategies accordingly. This activity was not planned in the DoW, we could only partially fulfill this new demand.

The report is thus divided in two sections, one corresponding to the report 3.1 as it should have been according to the DoW, and a second section that corresponds to the request from the reviewers.

I Requirements from scientists

Introduction

This report attempts to give a general overview of the complex data requirements by the research work packages and proposes a methodology for further identifying data gaps. This analysis and an informal preliminary investigation during the kick-off meeting allowed us to address the first gaps (Stygofauna Mundi, European fishes, odonates, …), as well as further requests along the project.
For the second period, we analysed how much the contingency funded projects covered the data requirements.

Methodology

The procedure to get the data requirement from partners was conducted in several steps:

- Explanations given during general meetings (the kick-off meeting for this first report) of the methodology and the types of information requested.
- Sending of a questionnaire to partners of research work packages to get their data requirements.
- Synthesis in a worksheet table (eventually to be converted into a real database).
- Textual synthesis of the requirements (this report, being the deliverable)

We requested partners to prepare slides for the kick-off meeting to present their data requirements. No indication was given to standardize the requirements, but a first assessment was done under the form of a worksheet (not provided here), which allowed to immediately start the search and the computerisation of missing data.

During the kick-off meeting, we explained how to rationalize the requirements, which were initially delivered under various formats. The recommendations were based on the fact that any piece of data or information on biodiversity includes 5 components that correspond to one type of question each:

- Taxonomic component; About?
- Geographic component; Where?
- Time component: When?
- Topic component: What?
- Reference component: How and who?

The 5th one, the reference component, is different of the others since it provides the source of the data/information covered by the four other components, and in some cases, explains the methodology used to get the data, that are actually metadata, which is important in some work packages that may use data or not depending on the methodology used.

For the four first ones, it is important to know the scale (the level or the range) and the resolution (the precision), e.g. for the geography:

Scale: continent, region, country, sub-country, freshwater body / catchment, ecoregion, any other area with definitions given.

Resolution: from “by full area” to “by point data”.

As a next step, the questionnaire was set up (see Annex I) and sent to all partners. The colleagues responsible for work packages, tasks and subtasks, and deliverables were requested to centralise the requirements by task.

The questionnaire was based on the rationalisation above. For each component, the scale and the resolution is necessary, which leads in a way to atomise each piece of the requirements, and to remove certain ambiguities due to the use of long sentences in natural language. In essence, a five-line requirement in natural languages can lead to 10 different requirements when each of the components is detailed.
In addition, we asked for the priority level, and the type of electronic format for the data.

The questionnaire, in Annex I, included explanation and example to illustrate what is meant by scale and resolution for each component.

Contingency funded projects

For the second period report, a simple overview of the data targeted by the contingency funded projects was performed. These projects were presented by partners from WPs 4-7, and thus represented a part of the gaps that needed to be filled.

Other information was gathered from discussions with partners, but also from the participation to other projects like Catalogue of Life.

The list of the projects is extracted from the mid-tern review report for the first period and reported in Annex II.

Results

Data requirements

From all relevant partners and tasks, we received 10 text files corresponding to the questionnaire and 4 worksheet files with additional details on data types. It corresponds to at least 20 different requirements that were listed in a worksheet prepared in the framework of this report[1]: some of the requirements could be split more but then, only slight variations would be recorded and would make the worksheet quite confusing. The worksheet also contains verbatim all the details sent under worksheet format, and the text files are compressed together in a separate file[2].

Taxonomy
In general, the data requirements of the research workpackages related to the "usual suspects" in biodiversity and conservation: vertebrates; and in freshwater ecology: several orders of insects (e.g., odonates, mayflies, stoneflies, etc.), several orders/families of crustaceans, and a few other benthic invertebrates. A few specific tasks targeted planktonic groups and plants/algae.

Geography
There was no general trend in the required geographical scale and resolution: they ranged from global and regional scales, down to catchment for the scale, and from ecoregion down to site level for the resolution.

[1] BF_WP3_D3.1_DataRequirements_Data_yymmdd.xls
[2] BF_WP3_3.1a_ DataRequirements_Questionnaires.zip

The use of ecoregions was requested in many cases. A few tasks requested the use of (or conversion to) UTM-coordinates, or other grid schema.
Although presence/absence by country was a requirement in one case only (because the ecology research focuses on catchments and ecoregions independently of administrative borders), we knew by experience in previous projects that it could also be important from the natural resource management point of view, so we included this level by default.

Nonetheless, there was a special emphasis on Europe in general, and on the three European catchments in particular that constituted the in-depth studied areas of the project.

Time
In general, the most recent data were requested, but some tasks require time-series and historical data.
In general, the more global was the geographic requirement, the less detailed was the time requirement.

Topic
The topics covered by the data requirements were extremely diverse, ranging from specific detailed topics such as Gross Domestic Product for lake catchments up to very broad topics such as “any data for endemic species”, and thus hard to summarize in a few lines.

We could split the tasks in two fuzzy categories with respect to how the topic data are requested:

- Tasks where the type of data was well defined, e.g. for models and activities that were already well developed, running almost as routines (e.g., WP7 on IUCN assessment);
- Tasks that were more exploratory and experimental, testing various possibilities for the models, and reacting on data availability rather than on strictly specified requirements.

For the models with precise requirements, data were often already with the partners for some taxa and regions, but additional taxa or geographical extensions were requested.
It was foreseen that to answer the needs for the experimental ones could be difficult, and we had defined with some focus and limits with the requesting partners.

References
Several data and information sources are provided; those are mainly electronic ones, but also refer to resources to be digitized (e.g., computerisation of the book Stygofauna Mundi for the stygobious fauna = cave and underground waters).

Few partners had, however, spontaneously came up with lists of references (and provided pdfs or hard-copies) that they would like to transform into databases. One notable exception is WP7, which provided about 900 references on odonates to encode data for IUCN assessments. And later on the project, other requests with references were sent (see the reports on data encoding).

Priority
Put together, the priorities requested by partners were not realistic, both for data acquisition and data digitization. Priorities were elaborated and followed for the data encoding.

Contingency funded projects

At the date of the second report when we performed this analysis, 18 contingency funded projects were accepted (3 and 3a are counted together): 12 were about acquisition of geo-referenced occurrence data as part of or the whole project, 4 compiled global or European lists of species, 2 were about geographic areas definitions. Some projects cover 2 topics (e.g., red list assessment), only the main ones were taken into account there. Threat and red list assessments required occurrence data as well.

Occurrence data

The contingency funded projects concerned occurrence data for two thirds of them. This corresponds to an important gap of data that was not depicted as such in the first analysis where requirements seemed to be more related to presence/absence per geographic area. Note that 7 over 12 projects were about freshwater fishes of Europe.

For most of the species, we were too optimistic that a high level density of data would exist in GBIF for freshwaters (as demonstrated by our attempt to get point data for *Salmo trutta* in Ebro river for WP4: about a dozen records found but not georeferenced and with no precise locality).

Taxonomy

Four contingency funded projects aimed at establishing lists of freshwater species, 3 global: caridean shrimps, freshwater aquatic and wetland plants, Macroinvertebrate taxa from med-rivers over the world; 1 European: riparian ground beetles (Carabidae).

They do not constitute Global Species Database as defined in Catalogue of Life: all species in a taxonomic group for the entire world. Rather they are only parts of GSDs, where some families or lower group would occur entirely in freshwaters.

Discussion

Data requirements

As explained above, we started to address the first known gaps mentioned during the kick-off meeting, those for which the resources to be digitized were immediately available.

In terms of availability of taxonomy, presence/absence in given geographic area, and occurrence data, we expected large discrepancies between geographic regions.

On the one hand, the distribution by country (or by state/province) in some regions (North America, Japan, Australia-New Zealand) are relatively well known like in Europe with the results of the FP5 Fauna Europaea. But still the distribution by ecoregion/basin/catchment may require some work, and could be the result of some automatic estimations by overlapping country and ecoregion/catchment.This particular exercise done by partners finally helped to quickly fill gaps in data, but required some quality control procedures. Otherwise, this type of data should be restricted to the use for partners or made available upon request with warnings related to their usage, to ensure that the users understand the limitations in the data.

Areas like China, Russia, South and Central America are in intermediary situations.

In terms of availability of environmental data, we have relied first on the knowledge of partners and their contact with local and national agencies to locate potential sources. Although we suggested it in the first internal report, we did not explore the datasets produced by the European Space Agency and the remote sensing data in general.

In terms of availability of topic data, aside the datasets with partners, the source principally lied down in the literature. And finally no request for topics came; partners had the possibility to manage them by themselves. In any case, we realized a posteriori that the demand would have been difficult from a skill point of view to satisfy..

On the other hand and at the opposite, Africa and South and South-East Asia are regions where little is known, and this will require to get in contact with colleagues there, and to create new database for the computerisation of data.

Areas like China, Russia, South and Central America are in intermediary situations.

In order to structure the way we wanted to fill the gap, and to answer in parallel to several requests (but note that we could not do everything at the same time) we have considered that a dataset could be known or unknown on the one hand, and that it was already digitized into a database or not on the other hand. Therefore, we could follow 4 working pathways in parallel to fill the gaps:

- Databases were known to the project: try to get the datasets to be incorporated in the data portal under the provider agreement. A particular effort was be made in Europe to gather various datasets to make a unique European repository (e.g., the various fish database results of the successive European FP projects on the freshwaters: FAME, EFI+, WISER to name some; see the report of a discussion between A. Melcher and N. Bailly). Unfortunately, unexpected issues on IPR were raised, and for the majority, these datasets could not be made available.
- Datasets are known but paper-published: encode the data and define with the partner the structure and the controlled vocabulary of the database. The localisation of the known datasets will come mainly from partners; in particular, the sources listed in the proposal should be exhausted first in accordance with priorities data requirement analysis.
- Datasets are unknown: This is the primary result of the gap analysis conducted with the help of the requirement database / tables as described in the previous paragraph.
    - The dataset exists (digitized or not): If we find a database, we tried to get it under the provider agreement and integrated it in the data portal.
    - The dataset did not exist: If we did not find a database or a paper-published compilation, then the dataset had to be created with the partner; we then proposed a project as long as the compilation and the encoding of data in a database was realistic in the timeframe of the project.

Contingency funds projects

Note: the recommendations listed below were then reported in the gap analysis report in D3.3.

<u>Occurrence data</u>

To understand the level of lack of occurrence data, one has to take into account the importance of the specific needs with respect to spatio-temporal scales where the developed models require a high density of point data at local spatial scale (e.g., one river, one basin, one sub-ecoregion, ...), and sometimes for restricted time range

(monitoring, seasonality studies, ...). The issue of density of data was lacking in the questionnaire for the first analysis, which is usually needed as very high at the scale and resolution requested by partners.

The paucity of point data could be linked to the rarity of species in general in ecosystem. This hypothesis should be further tested in a dedicated work as long as there are large enough and available datasets with general abundance of species in ecosystems.

Another consideration is that museum data that constitute the major source for most of the (rare?) species are not enough to answer the needs for point data from all scientific domains. Survey data are important to get in as already done in marine environment for some fisheries surveys. BioFresh has an important role beyond the end of the project, to (help to) make these data available eventually in GBIF for freshwaters. The OBIS (Ocean Biogeographic Information System) initiative model is then to be considered for BioFresh, maybe as a continuation after the EU-FP7 funded project.

The gaps listed in annex through the project list will be solved, but they are the mark of a huge gap worldwide and beyond the limited number of taxonomic groups targeted here.

Recommendation: **recommend future funding for massive data encoding efforts.**

Seven of the projects on occurrence data concern fishes, and it is possible that it is regarded as an over-emphasized focus for the project. One possible explanation is that occurrence data for fishes are already largely available because the major fish collections in Europe are now computerized, and routine surveys are or will be soon made available. The projects concern mainly countries where there is no big collections, and that completed in a very remarkable way the freshwater fish occurrence data in Europe.

Another consideration is that fishes are extensively covered from a taxonomic and name point of view by a number of databases or initiatives (Catalog of Fishes, FishBase, FishNet, Fish-BoL, ...), that work altogether, where new species and new revisions are updated quite actively. They constitute a very powerful tool that is barely available for other groups of that importance (e.g. birds where 4 different global lists are available). The gap is progressively being filled up but it will take time to reach the point where fishes are already.

The massive efforts for other groups should follow the same pattern and focused first on European species to constitute “complete” datasets.

Taxonomy

The four projects correspond to some of the gaps that are known in Catalogue of Life, although they do not constitute GSDs: Crustaceans, Plants, and Beetles are known to be incompletely covered by existing databases about these groups. BioFresh/FADA has to advertise these gaps and work in collaboration with existing initiatives to ensure that freshwater species are extensively covered.

For the macroinvertebrates in Mediterranean-like rivers of the world, the gap may be also due to the fact that some global species database do not indicate if the taxa they list occur in marine, freshwater or terrestrial habitat. A solution was started in CSIRO, Australia, where Tony Rees elaborates a dataset called IRMNG to flag what is marine and extent from what is not marine and fossil. However, the freshwater flag is not yet set up. This flag will be added soon to the catalogue of Life as far as it is possible.

Beyond these 4 gaps that are punctually solved by the contingency funds, it is obvious that there is a need to complete the freshwater species world list (e.g., through the FADA initiative in collaboration with others). However, the gap could be partially addressed if existing databases would flag the environment of the species when relevant.

Recommendation: **recommend future collaboration with Catalogue of Life and IRMNG to flag freshwater species in global species databases and lists.**

FADA could play a lead role with the support of BioFresh, but also beyond as long it can be sustained and supported in a proper way. There is a need to work closely with WoRMS (World Register of Marine Species) that stores already some freshwater species either when a small freshwater group belongs to a group in majority and/or primary marine, or when species are diadromous, or when species inhabits areas called “Seas” but where they leave actually in freshwater (Baltic Sea, Caspian Sea, etc.).

Another marker in taxonomy gap is the number still elevated of new species described from Europe in general (result of Fauna Europaea for both terrestrial and freshwater species). If we take the number of fish species as an indicator, it would be interesting to analyse the recent new descriptions restricted to freshwaters. This would give an idea of the level of knowledge that we have about the taxonomy in the European freshwaters and where are the source of new species: from new exploration and/or museum specimen discovery, or the re-evaluation and splitting of new species with new methodologies, genetics/sequences in particular, like in the case of a new pike species described in Italy. What could be the impact on species richness models would be also an interesting insight.

Other data

IUCN use Hydrosheds as a basic ecoregion structure. This dataset was not complete. Other systems exist and may constitute gaps if they are not complete, e.g., the Pfaffstetter system use in AquaMaps.

Another request on KBAs is not a mark of a gap.

Two partners submitted other type of data to be computerized including biological and environmental parameters. Their scope is somewhat limited and it is not clear yet if they constitute a gap at a larger scale.

Conclusions

Datasets requested/encoded

The datasets that were requested and encoded are reported the data encoding reports as deliverables 3.2a-c.

Data used

We compared in the end which data were required and which were actually used during the project in the various WPs, including the datasets that partners had already or created by them during the project (see details in D4.1-7.1). The match was almost perfect (see D4.1-7.1) if point data were implicitly requested 9see below in Lessons learned).

However, our thoughts about the country scale were not verified (see D4.1-7.1 for further comments about that, as well as on species traits).

Lessons learned

Colleagues from WPs 4-7 had correctly expressed their needs, there was no surprise. However, the need for point georeferenced data was not clearly expressed after the first analysis. It may be explained by the lack of this precise topic in the questionnaire. Retrospectively, there might have also been a misunderstanding about the meaning of “Biodiversity data”, a locution that is abusively used as a synonym of “point data” (while biodiversity is supposed to cover gene/population/species/ecosystem domains, but biodiversity is a buzz word …).

Even with the addition of the contingency funded projects, there was not time enough to encode a large amount of data in the time allocated before the analyses performed by the other work packages. For the design of new proposals, it is important than in the preparatory phase, needs are expressed more precisely. However data encoded are not lost and can be used by other colleagues or initiatives.

The challenge was to face the high diversity of claims, for which, we had to adapt our methodology each time with new challenges. But changing processes implies each time a non-negligible learning curve in terms of human resources.

II Requirements from Policy makers

Introduction

After our first report on the gap analysis, the reviewers of the first mid-term review held in April 2012 stated that our scientific data gap analysis delivered nothing that was unknown before.

In our report, three recommendations were made as following (the first one from the data requirements report):

Recommendation: **recommend future funding for massive data encoding efforts.**

Recommendation: **Important need for funding the establishment/extension of authoritative freshwater species lists including an exhaustive overview of "synonyms".**

Recommendations:
- recommend European policies on IPR issues related to aggregated data.
- explore the options for a standardized embargo period after which publicly funded scientific data should be released.

Reviewers mentioned these recommendations were not new. One can argue that if the issues were spotted some time ago and are still flagged in more recent studies, it is because no good solutions were found yet or that previous recommendations were not followed, and still need to be addressed.

Reviewers wanted us to produce a new deliverable oriented towards the policy makers, a demand driven data gap analysis and develop data mobilization strategies. This activity was not planned in the DoW, and we detail hereafter how we have answered partially to this new demand.

DIK(U)W Hierarchy, and Method

DIK(U)W stands for Data, Information, Knowledge, Wisdom, and Understanding, which designates any process set up to transform data into information, information into knowledge, etc. (see figures 1,2). In our domain, wisdom should be understood as management (of natural resources). This theoretical framework about information management will help to clarify the role of researchers, and why our answer to reviewers is necessarily partial and to be given at several levels and in different perspectives.

The first perspective, a practical one, is that the Global Freshwater Biodiversity Atlas, being constructed as a part of the BioFresh web presence, gives answers to many questions by synthesizing graphically data into knowledge as maps. It answers the basic questions at global level about the distribution of freshwater species richness and how it is structured through a number of parameters, the threats to the freshwater biodiversity, possible impacts of climate change, etc.

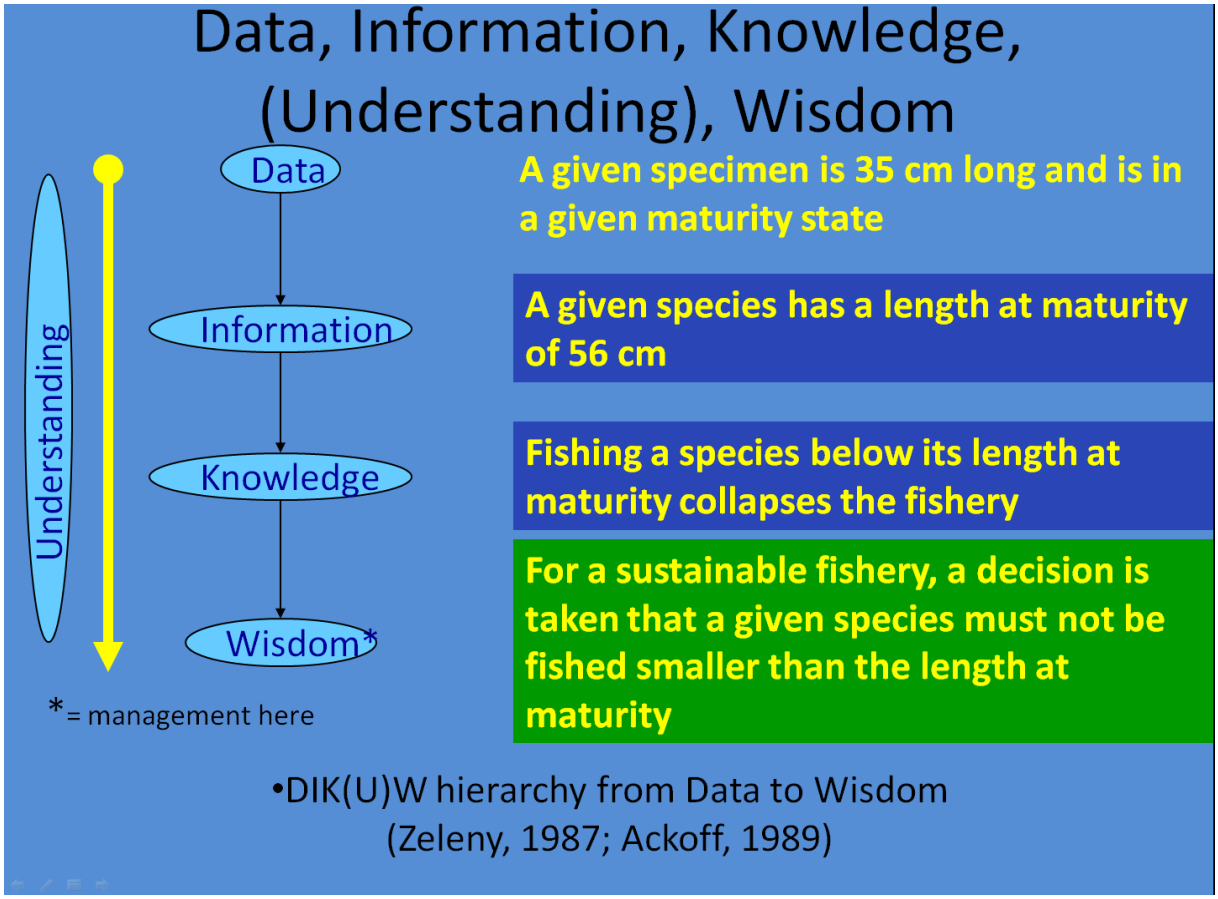

Fig. 1. The data to management hierarchy for fishery management, e.g., in continental waters.

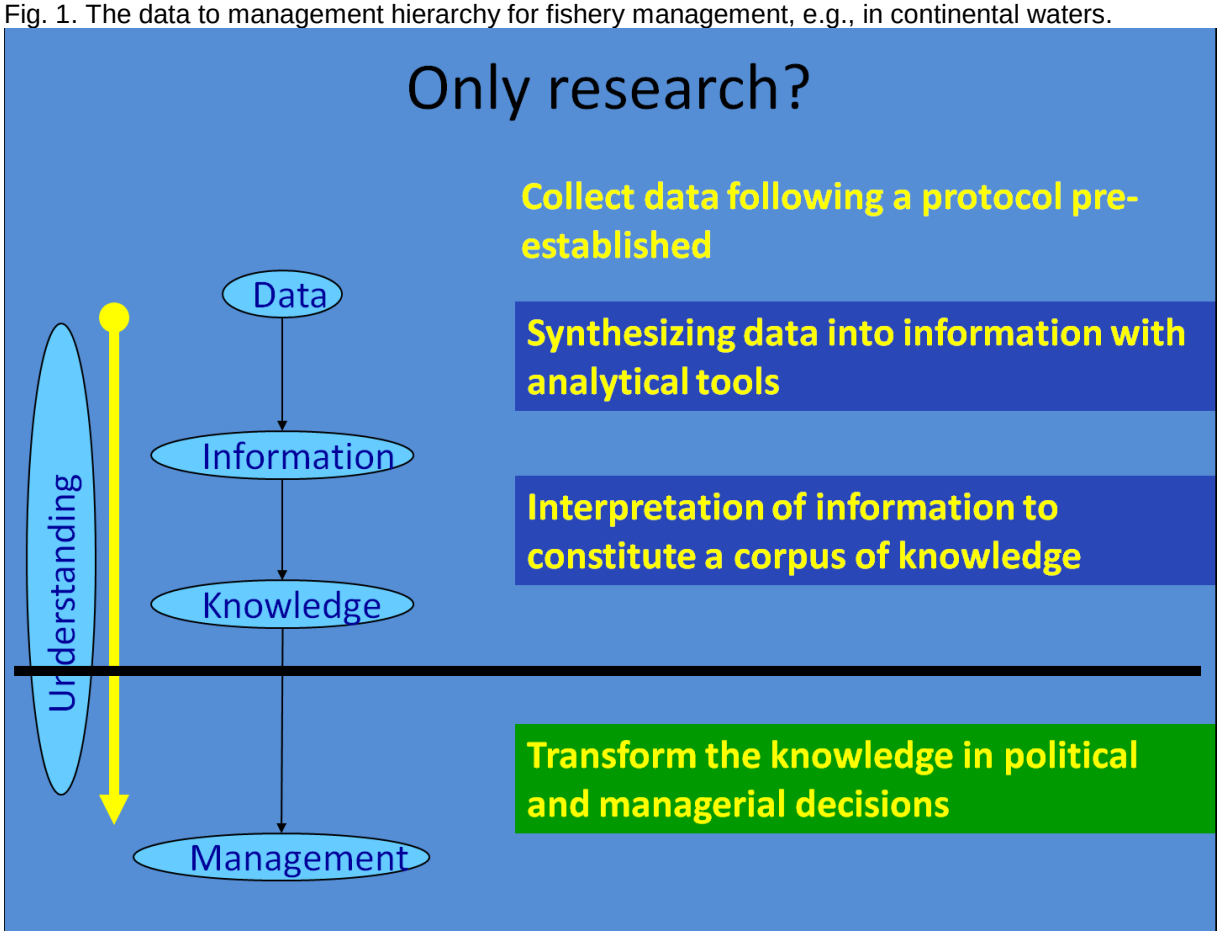

Fig. 2. Generalization of the data to management hierarchy for natural resources management.

We believe that the Global Freshwater Biodiversity Atlas already provides part of the solution, and is part of the strategy for mobilizing data demonstrating how shared, citable and traceable back to publication data can be presented for the benefice of the society. Implicitly the message is: the more good data we mobilize, the better will be the maps and their interpretation towards a better sustainable management of natural resources.

The second perspective is more theoretical but explains why the Atlas is part of the answer for the data mobilization needed for information. There are steps between raw data and the decisions to be taken for management – Information and Knowledge (Figs. 1, 2). The BioFresh project and its portal are primarily focused on the data and information steps (Fig. 2) according to the DoW, less about the knowledge step. It also addresses the understanding processes for primarily synthesizing data into information, and, secondarily, interpreting information into knowledge. In our view, the Atlas constitutes the start of the establishment of the knowledge step. The interpretation (as in Fig. 2) in the Atlas is materialized by the selection of a number of thresholds for plotting maps for instance. If the quality of data can be assessed and controlled up to a certain point, the interpretations are obviously more arguable.

What is expected from researchers and engineers are 1) for researchers: to identify important factors in general, and to establish theories (e.g., the IUCN RedList assessment), and 2) for engineers: to provide local data to which these theories will be applied at local level (with possible adaptations due to local peculiarities) (e.g., the 3 European basins targeted). Depending on the level of management, stakeholders will require both, but separately (see the section below Think globally, act locally).

The datasets that are or will be available in the portal fit completely the first point above and are used in the second and third understanding phases of the Figure 2 (synthesizing, and interpretation) for establishing theories, and accumulate knowledge. They fit less the local level (with exceptions of the local areas targeted by the project).

In other words, we may receive requirements from a number of stakeholders at too low level, i.e., for local solutions, rather than general considerations for establishing theories. The BioFresh consortium is not organized to answer the former, except the ones coming from the partners in the project.

Finally, it is beyond this project to address directly the fourth understanding phase (from knowledge to management).

The process that starts by asking the requirements from the management back to the needed original raw data is quite long, and should have started with the project. Moreover, it cannot address directly the data or even the information step (that targets still technico-scientific people to transform this information into knowledge for the managerial and decision level).

For example, in fisheries, one question is: what regulations should be put in place to develop a sustainable fishery on a given species in a given area? One regulation is to set up a minimum size of capture, which requires to get the information of the size at maturity for that species (which potentially varies with the latitude where the stock lives), which requires to have many raw data as couplets [length, maturity stage]. Incidentally and historically, it is not this top down approach that was followed, but rather the bottom up one.

As it was not planned in the DoW since the beginning, we considered after the first mid-term review that the effort to organize consultations with many stakeholders at various levels of the DIK(U)W hierarchy would lead to too

many changes in the DoW, in the budget, and in the partnership. This decision was also driven by bad experiences in previous European projects where, although the consultation was planned, the methodology was not elaborated at the beginning of the project, and the meeting was not well prepared due to lack of time. Hence, the meeting delivered very few results and a lot of frustration on both sides. But as we were looking for proxies, the symposium Water Lives held in January 2014 (as decided early 2013) offered an interesting opportunity. As it was shaped we realized that it constituted the meeting of stakeholders that one workpackage could have not organized by gathering many policy makers in Europe.

The presentations and discussions were extremely frank, both on the policy makers and the scientific side, revealing honestly the huge gap between the two worlds, which was not the rule in our previous experiences of such meetings.

We have reorganized and summarized the discussions of that meeting as the gap analysis for policy makers, in addition to the Atlas as part of the solution. For each point we have made a conclusion in the form of a recommendation for the next calls in Horizon 2020. Some points were added from a consultation held with S. Condé from the EEA/Topic Center for Biodiversity at MNHN, Paris in September 2013 (a summary of the discussion is put in Annex III).

Water Lives discussion summary

Outline

Gaps for policy makers

1. **Policy makers expect socio-economic justifications for biodiversity conservation**
2. **Policy makers expect solutions (possibly under several scenarios) in addition to problem exposure**
3. **Policy makers read and listen to short, concise and precise statements**
4. **Policy makers do not like uncertainty**
5. **The main geographic unit for management seems to be the catchment (like in France and Italy)**
6. **Although decisions are mainly driven by socio-economics, interest and power, there is still space for ethical issues**
7. **Policy makers need to be educated about freshwater biodiversity**

Gaps for scientists

8. **Scientists do not know how to push their results in the policy making pathways**

Gap Details

Gaps for policy makers

1. **Policy makers expect socio-economic justifications for biodiversity conservation**

Results of projects should argue for possible impact of biodiversity conservation over currently established economy.

In particular:

- What are the benefits from ecosystem services?
- What would be the cost of lost ecosystem services (what would be the cost of artificial mechanisms set up)?
- What are the thresholds for irreversibility?
- What are the costs associated with irreversibility?

Conclusion 1: For every proposal about biodiversity conservation, there should be (at least) one partner and one workpackage on socio-economics.

Remarks: It was clear during the symposium that natural science, cultural/patrimonial and ethic justifications have few chances to trigger any action from the policy making point of view. Successes happened though, but these cases should be understood as exceptions and not as the basic strategy that should be re-oriented to socio-economic justification (see point 6 however).

2. Policy makers expect solutions (possibly under several scenarios) in addition to problem exposure

A good example is the proposal to better manage the riparian biodiversity, in particular the plant/tree cover where economic justification are relatively easy to draw by comparison of well managed banks against deteriorated biodiversity in these areas.

Conclusion 2: In relevant work packages in proposals, there should be clear description of tasks about the elaboration of scenarios under various politico-socio-economic situations resulting with different political decisions.

Remarks: In the Fisheries Centre of the University of British Columbia, a model has been used to elaborate a game that shows the impact of political and regulatory decisions over fish stocks and fishing grounds. Policy makers were invited in a room with 8 computers connected in a network and a big screen displaying the evolution of various indicators after decisions and actions were taken [and stocks were always depleted after a series of rounds ...].
With the knowledge that ecology has from the ecosystems, it might be possible to repeat these experiences on freshwater ecosystems having a visualization of the resulting landscape in 3D. For example, the drought and the refilling of Aral Sea should not be too difficult to model now, integrating at least the principal factors, with an impacting visual effect: fishing boats in the middle of a dried sea. Results about the riparian biodiversity should have a strong visual impact too.

3. Policy makers read and listen to short, concise and precise statements

Policy briefs seem to be a main vehicle from science to policy makers as long they can be read quickly and easily by them, meaning that any technical jargon must be translated in an understandable speech.

Conclusion 3: For every proposal about biodiversity conservation, the workpackage on communication should explicitly list policy briefs deliverables. Involve more the community that is already at the interface science-policy to write down these short papers.

Remarks: In proposals, communication services should be involved when they are already used to produce briefs and other vulgarization work.

4. Policy makers do not like uncertainty

The solution is to revise the scientific language used to export results. This is also linked to the points 2 and 3.

Conclusion 4: Express uncertainty through the proposal of several solutions under several scenarios, which will mask positively a part of the uncertainty of scientific answers.

5. The main geographic unit for management seems to be the catchment (like in France and Italy)

The management at catchment level seems to be emphasized by countries. BioFresh paved the way with the 3 catchments workpackage.

Conclusion 5: Most of the proposals should include a workpackage on integration of data, information, knowledge, management, and models at catchment level.

Remarks: For European projects, inter-country catchment basins should be prioritized.

6. Although decisions are mainly driven by socio-economics, interest and power, there is still space for ethical issues

As it happened in the past (e.g. for stopping slavery, child work, etc.), it can be shown by scientists to policy makers that there are other philosophical (ethical, cultural, aesthetic, spiritual, moral, …) drivers.

This is not a request from policy makers, but a gap in the way they think and that was felt and expressed by scientists during the discussions.

Conclusion 6: As most as possible, there should be a partner in philosophy and a workpackage on other ways to value biodiversity than monetary.

Remarks: It is not necessarily a large workpackage, but some person-months could be used to produce a document summarizing the most recent philosophical thoughts about the topic(s) of the project. Discussions could be planned during the general meetings. Epistemology is not the primary target here, but is not necessarily excluded.

7. Policy makers need education about freshwater biodiversity

This is not a request from policy makers, but a gap in the way policy makers integrate issues on biodiversity that was felt and expressed by scientists during the discussions.

Conclusion 7: As most as possible, there should be a partner and a workpackage specialized on production of education material, in particular for high engineer, administration and managerial schools.

Gaps for scientists

8. Scientists do not know how to push their results in the policy making pathways

After a policy brief is written, scientists must be in contact with people who will put them forward.

Conclusion 8: Proposal should associate people who know pathways into policy making, if possible at the beginning of the proposal making. There is also a need to build a strong scientific advocacy to mitigate the impact of economy lobbying.

Remarks: Following the policy maker colleagues who were present during the symposium, the difficulty is not to obtain a first appointment, but to obtain the second one! The first contact must be like a policy brief, short, concise and precise, and emphasizing solutions rather than problems. It is not sure if all scientists are well trained for this exercise. Partnerships with NGOs advocating for the defense of environment could propose a solution by inserting a third party between the science and the politics domains. The staff of these NGOs should know both domains, and be able to translate the knowledge produced by scientist in statements that could trigger decisions and actions by the policy making domain.
One sociological research topic could be to understand what are the speeches, conditions, proposals, or demonstrations, etc. that would have the highest chances to trigger a decision by a policy maker.

Think globally act locally

As a third perspective, the well-known statement “Think globally act locally” could be applied for data needs for policy making if we think about two different scales.

Data are needed for two different exercises:

- At large (global) scale: to establish generalizations if not theories about the possible threats to freshwaters (which requires data to understand the functioning of ecosystems). Data are needed for research, which will establish the framework at large for policy making. Already we know a lot about

what constitute the potential threats to any freshwater body picked at random (Fig. 1). The amount of data collected behind this sketch is certainly huge (but mostly data are not available).

- At low scale: to apply theories to local conditions within the framework at large, which is the (important) role of engineering, not research anymore. Local data are needed usually with a higher density and frequency. Small and Medium Enterprises (SMEs) should be involved at that scale and bring to the local policy makers the knowledge about the freshwaters under their close jurisdiction inferred from the analyses of locally collected data. Such SME survival should be supported by policy making at large scale (e.g., enforce freshwater environment impact assessment any time that a possible threat is put in motion by any project), and by local policy making and enforcement where SMEs would be hired to deliver a given impact assessment. And make available data to test and fine-tune theories.

As we realize, the data needs for policy making are quite different depending on the scale that is addressed, global theories (think globally) and particular applications (act locally). All the points from the symposium above are true at both scales, but the precise nature and the amount of data to be collected may vary significantly.

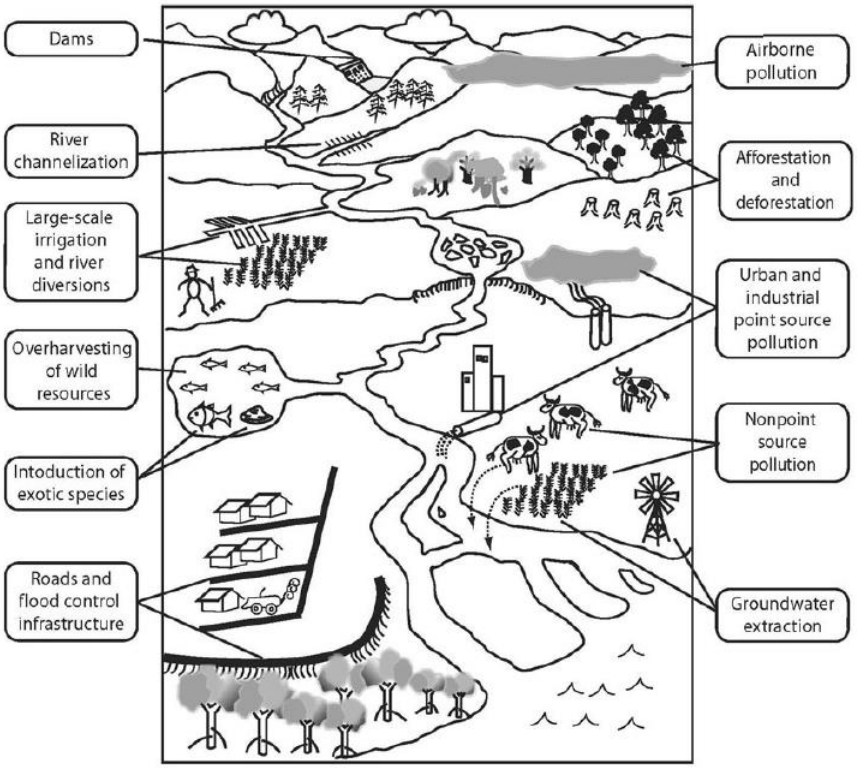

Figure 1. - Many interacting drivers, and interacting responses, affect the condition of freshwaters on any particular landscape. Mining activities should be added (gravel extraction). After Carpenter & Biggs (2011).

Further work

A questionnaire (Annex IV) was distributed during the symposium but too few were returned by scientists, and even less by policy makers. It did not add any other considerations but supported many of the points detailed above, in particular the difficulty to communicate between scientists and policy makers.

An opinion paper should be produced before the end of 2014.

Annex I

I Worksheet file gathering all data requirements
As a separate.

II Questionnaire for the requirements

Template

–About? Taxa (scale/resolution)
–Where? Biogeopolitical areas (scale/resolution)
–When? Time (scale/resolution)
–What? Topic incl. Environmental data (scale/resolution)
–Who / How? References / Data sources / Data contacts.

–What priority level?
–How do you want the data structured OR we provide simple structure (XLS)?

Please refer to BioFresh metadata fields for data definition as far as possible (see separate email message).

Template explanations

We suggest first you to describe data requested as complete sentences in natural language:
Examples:

- period of spawning of crocodile species by ecoregion;
- maximal total length of fishes for Ebro, Elbe and Danube;
- temperature of water per quarter by catchment in Spain;
- etc.

Then you create templates for quintuplet (about, where, when, what, who/how) in such a way that no ambiguity remains.

Example:

- Maximum length, trophic level, and type of reproduction for insects, amphibians, odonates for Ebro, and for amphibian, fishes for Elbe requires at least 2 templates to avoid we search data for fishes in Ebro, and for insects and odonates for Elbe. Again this is a trivial example but some cases may be tricky as we already depicted in the slides from the kick-off meeting.

–About? Taxa (scale/resolution)
Scale: taxa targeted, if possible with scientific name, and the taxon levels when precision is needed or ambiguities are possible, e.g.:

- Trichoptera alone is ok;
- crustaceans may need more details, otherwise it encompasses ALL Crustacea, including macro- and micro-fauna;
- crocodiles: is it strictly the genus *Crocodylus* or the entire family Crocodylidae, including alligators, caimans, and gavials? (this example is trivial but less obvious cases may happen).

If the taxa are restricted by a biological or ecological trait, a definition may be needed, e.g., for macrofauna, precise a minimum size.

Resolution: from “by phylum” to “by species”. If not indicated, we will understand species by default.

Example: species of crocodiles (family).

–Where? Biogeopolitical areas (scale/resolution)
Scale: continent, region, country, subcountry, freshwater body / catchment, ecoregion, any other area with definitions given.

Resolution: from "by full area" to "by point data".

Examples:
- Point data in the world;
- By country (e.g. presence/absence).

–When? Time (scale/resolution)
Scale: from all data anytime to a restricted period.

Resolution: from "by the entire period" to "by year / season / month" (we don't think we will need data below that resolution …).

–What? Topic incl. Environmental data (scale/resolution)
Scale and resolution: as data requested may be extremely various, it is difficult to give a generic definition.

Resolution: when needed, the precise unit should be indicated.

Example: for instance, requesting data on reproduction is not precise enough, and will need to de detailed: fecundity in ovules/kg, period of spawning, total length at first maturity in cm, etc.

–Who / How? References / Data sources / Data contacts.
When known, references of data sources including non computerized documents can be sent.
In the maximum of cases, we will link piece of data to the original reference in order to give full credit given, which allows users to check the methodology when needed. If the methodology needs to be stored as a metadata, this must be explicitly requested.

Example: Habitats from the following reference:
Botosaneanu, L. (ed.), 1987. Stygofauna Mundi. A Faunistic, Distributional, and Ecological Synthesis of the World Fauna Inhabiting Subterranean Waters (Including Marine Intersti), Brill Academic Publ. 740 p.

–What priority level?
Indicate deadlines when necessary.

–How do you want the data structured OR we provide simple structure (XLS)?
Indicate if a specific format is necessary and if a database exists already from which we could copy the structure.

–Refer to BioFresh metadata fields for data definition (see separate email message)?

The type of data for which projects were funded with the contingency funded were analyzed to complete from an independent point of view the results obtained from the first report. They were put in the perspective of the requirements analysis to check if they fill these needs.

ANNEX II

List of contingency funded projects.

Project Nr.	Decisions for data acquisition	Data/tool
1	**Dragonflies of tropical America, Australia and Asia**	4000 references of dragonflies of tropical America, Asia and Australia
2	**Distribution data of Freshwater fishes of Russia and adjacent countries**	20,000 geo-referenced distribution data of freshwater fishes of Russia and adjacent countries
3	**Distribution data of Greek freshwater fishes**	2400 geo-referenced distribution data of Greek freshwater fish species
3 a)	**Distribution data of Greek freshwater fishes**	Additional distribution of freshwater fishes from Greece: 150 geo-referenced sites
4	**North America freshwater fishes**	Completion of data sets for North American freshwater fish species. Maps and threat assessments for fishes (800 spp.).
5	**HydroSHEDS**	Development of new tools to supplement the HydroSHEDS global hydrographic database.
6	**WFD Intercalibration datasets**	Water Framework Directive Intercalibration datasets
7	**Freshwater species of New Zealand**	Red List assessments for selected freshwater macrofauna, approximately 500 species
8	**Global caridean shrimp fauna**	Distributional database and Red List data of freshwater caridean shrimps (789+ taxa)
9	**Distribution data of Freshwater fishes from Poland**	Geo-referenced distribution data of freshwater fishes of Poland from about 450 papers
10	**European caddisfly fauna**	Compilation of about 500,000 geo-referenced distribution data
11	**Central European riparian ground beetles (Carabidae)**	Taxalists of riparian ground beetles (Carabidae) of riparian areas along Central European rivers
12	**Global checklist "freshwater aquatic and wetland plants"**	Compilation of a global checklist "freshwater aquatic and wetland plants".
13	**Freshwater fishes of Serbia**	Geo-referenced distribution data of freshwater fishes of Serbia from 3-4 thousands sampling sites
14	**Freshwater fishes of the Western Balkans**	Geo-referenced database of freshwater fishes of the Western Balkan; 10 000 records
15	**Freshwater fishes of Italy**	Georeferenced database of freshwater fishes of Italy; 16 000 existing records + 25 000 new ones
16	**Mediterranean Climate Watercourses**	Macroinvertebrate taxa from med-rivers over the world (Mediterranean Basin, South Africa, Chile, California, and Australia), setting-up MacroMed database to be intergrated into BioFresh
17	**Non-marine Ostracod from Europe and Eurasia**	Improving accuracy of geographic referencing and nomenclatorial and taxonomic congruence of *c.* 45,000 records of non-marine Ostracoda from the OMEGA database,
18	**Birdlife Internation – Freshwater Key Biodiversity Areas**	Customise the existing database (World Biodiversity Database) to hold all the information on FW KBAs.

ANNEX III

Summary of a consultation held with S. Condé
from the EEA/Topic Center for Biodiversity at MNHN, Paris in September 2013

In the context of ecosystems assessment including their services, information gaps are:

- Species traits which can help for selecting indicator species in relation with state of ecosystems;
Issues:
+ Which traits?
+ Low level or high level: low level, e.g., fecundity, high level, e.g., resilience/vulnerability to climate change, indicators that would integrate many traits.

- Tipping points for freshwater ecosystems;
Issues:
+ Which characteristics of these ecosystems can support the definition of a level of irreversibility?
+ Does any scientific work progress on this question?

- Compiling data for ecosystem services mapping
+ Problem on definition of services and based on which data: issues on scale and distant services (footprint?)
+ Check IPBES and GEOBON and EBONE http://www.wageningenur.nl/en/Expertise-Services/Research-Institutes/alterra/Projects/EBONE-2.htm

- Presentation of knowledge at the Science-Policy interface
+ Help stakeholders to formulate their questions and to transform them in data requirements:
Check the EU project KNEU which is defining a Network of Knowledge called BiodiversityKnowledge
(The current gap analysis could be just to formulate the questions).
+ What are the past and existing projects?

- Information needed to solve the possible incompatibilities between various directives:
Within the WFD, problem of standardization mainly on the water bodies classifications (nearly each MS uses its own classification)
Between WFD and Nature directives, any study on possible inconsistency between a water body in a Good Ecological Status with the Favorable Status of a species or habitat listed in the Nature directive?

ANNEX II

Questionnaire: BioFresh – Water Lives – 29-30 January 2014
For a gap analysis in freshwater biodiversity knowledge/information for policy making

This questionnaire aims at exploring where are the main gaps from policy-making point of view for (primarily freshwater) biodiversity sustainable management balancing exploitation and conservation. Important: It concerns primarily your personal experience. Ultimately this analysis will provide to the scientific community some indications where efforts are to be made to collect data necessary to elaborate requested information and knowledge.

[The information below will not be reported further, just used for personal contacts if precisions are needed; however, you may skip those, just indicating which type is your institution: governmental, non-governmental, research, private sector, …]
Last and first names: Email:

Institution:

Position and role:

Do you think of yourself as a policy-maker (or participating directly to policy-making?)
Same question for your institution?
Do you work in the domain of freshwaters (100%, partly, No)? Your institution?

Did you encounter cases where missing knowledge / information prevented or delayed any further policy- or decision-making? Please tick the boxes, at least one per line, several per line allowed.

Exploitation: did you miss knowledge on impact of exploitation over biodiversity sustainability at:
Biodiversity level: Ecosystem □; Species □; Population □; Genetics □;
Geographic level: Europe □; Country □; Subcountry □; Municipality □; Hydro (basin, river, lake) □;

Conservation: did you miss knowledge for biodiversity conservation and impacts on socio-economics at:
Biodiversity level: Ecosystem □; Species □; Population □; Genetics □;
Geographic level: Europe □; Country □; Subcountry □; Municipality □; Hydro (basin, river, lake) □;
Socio-economic level: Economic □; Social □;

Ecosystem services: did you miss knowledge on their benefits, or the costs for society if they are lost at:
Biodiversity level: Ecosystem □; Species □; Population □; Genetics □;
Geographic level: Europe □; Country □; Subcountry □; Municipality □; Hydro (basin, river, lake) □;
Socio-economic level: Economic □; Social □;

Climate change: did you miss knowledge on impact of climate change over biodiversity management at:
Biodiversity level: Ecosystem □; Species □; Population □; Genetics □;
Geographic level: Europe □; Country □; Subcountry □; Municipality □; Hydro (basin, river, lake) □;
Socio-economic level: Economic □; Social □;

Detail an example where you missed knowledge / information for policy- or decision-making about (freshwater) biodiversity (please use the verso of this page if needed):

Did you encounter problems to explain your needs in terms of knowledge / information to the scientific community? Please use the verso of this page to detail an example if any.