Knowl. Manag. Aquat. Ecosyst. 2022, 423, 25
© P. Irz *et al.*, Published by EDP Sciences 2022
https://doi.org/10.1051/kmae/2022021

www.kmae-journal.org

Knowledge &
Management of
Aquatic
Ecosystems

Journal fully supported by Office
français de la biodiversité

DATA PAPER OPEN ACCESS

A long-term monitoring database on fish and crayfish species in French rivers

Pascal Irz[1,*], Thibault Vigneron[1], Nicolas Poulet[2], Eddy Cosson[3], Thierry Point[3], Erick Baglinière[1,2,3] and Jean-Pierre Porcher[1,2,3]

[1] OFB, Direction régionale Bretagne, 84 rue de Rennes, 35510 Cesson-Sévigné, France
[2] OFB, Direction de la recherche et de l'appui scientifique, 2 allée du Pr Camille Soula, 31400 Toulouse, France
[3] OFB, Direction de la surveillance, de l'évaluation et des données, 5 square Félix Nadar, 94300 Vincennes, France

Received: 17 May 2022 / Accepted: 7 September 2022

Abstract – The data presented here cover ~4 decades of electrofishing surveys in mainland France rivers. Despite some changes in field protocols and sampling site locations over time, the ASPE database offers a unique set of 47,869 georeferenced and standardised sampling surveys. Fish individuals are identified to species, measured and weighted. Additional information encompasses habitat features as well as "river health" assessment by two fish-based indices. Though preferentially focused on fishes, the surveys also report crayfish data. Numerous studies have been already published using these data, mainly on the fields of macroecology, community and/or population ecology and bioassessment. However, we believe much more knowledge can be gained from the database in fields such as population dynamics or conservation practices in the context of global change.

Keywords: Database / freshwater habitat / rivers / fish / crayfish / mainland France / long-term monitoring / population dynamics / macroecology / time series / bioassessment / FAIR / electrofishing

1 Introduction

Freshwater systems spread over less than 1% of the Earth's terrestrial area (Allen and Pavelsky, 2018) but host c. 12% of all described species (Reid *et al.*, 2019). Fishes alone represent a quarter of this diversity (Su *et al.*, 2021). In other words, freshwaters are biodiversity hotspots. However, they undergo diverse and severe anthropogenic pressures, including climate change, pollution, flow alteration including withdrawal, habitat degradation and/or fragmentation, overexploitation and exotic species introductions (Dudgeon *et al.*, 2006; Vörösmarty *et al.*, 2010; Dudgeon, 2019; Oberdorff, 2022).

In this context, freshwater fishes are particularly at threat with nearly one third of the species at risk of extinction globally (Hughes, 2021). In mainland France, 39% of the freshwater fishes are considered near threatened or worse (IUCN *et al.*, 2019) and only 12.5% of the freshwater fish species have a good conservation status according to the EU Habitats Directive (PatriNat, 2019). The understanding of species responses to global change, and the assessment of species conservation status in order to establish efficient management plans, both require long-term monitoring. Such long-term monitoring data are globally rare (Comte *et al.*, 2021).

The ASPE ("Application de Saisie des données Piscicoles et Environnementales") database gathers together biological and environmental information on monitoring sites covering the French territory since the sixties. ASPE is unique in spatial extent (mainland France, ~550,000 km^2), number of georeferenced sampling sites (~6000) and time span (40+ years). The Office Français de la Biodiversité (OFB, https://www.ofb.gouv.fr/) currently administrates the database.

ASPE was the main dataset used for the production of species distribution maps for the French atlas of freshwater fishes (Keith *et al.*, 2020). The ASPE dataset also supported research in the fields of macroecology (*e.g.* Reyjol *et al.*, 2007; Hugueny *et al.*, 2011), species ecology (*e.g.* Cattanéo *et al.*, 2002; Poulet *et al.*, 2011; Bret *et al.*, 2016, 2017; Santos *et al.*, 2021), community ecology (*e.g.* Oberdorff *et al.*, 1998, 2001a; Tales *et al.*, 2004), bioassessment (*e.g.* Oberdorff *et al.*, 2001b, 2002; Marzin *et al.*, 2012, 2014; Clavel *et al.*, 2013), climate change (*e.g.* Buisson *et al.*, 2008; Buisson and Grenouillet, 2009; Comte and Grenouillet, 2013; Comte *et al.*, 2014, 2016), biological invasions (Manné and Poulet, 2008), etc. Nonetheless, we believe much more research can be achieved due to the recent improvements in the quality of the

---

*Corresponding author: `aspe.datapaper@ofb.gouv.fr`

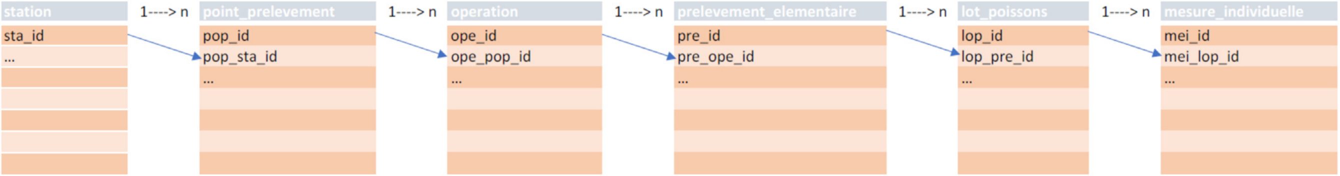

Fig. 1. Summary of the database Physical Data Model indicating how to join the six 'core' tables.

data, the stability of the sampled sites network, the standardisation of the sampling schemes, and the overall amount of surveys reported.

Here, we publish for the first time the ASPE database.

Our workflow was designed to ensure reproducibility (Peng, 2011). It includes a data processing step to produce the figures, maps and plots, followed by an authoring step to assemble these elements with the text with R Markdown (Xie *et al.*, 2018). The R (Ihaka and Gentleman, 1996) and R Markdown (Xie *et al.*, 2018) files used for data preparation and manuscript production allow any reader wishing to reproduce or modify our analyses to do so. This is intended to respect, as much as possible, the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles of scholarly publishing (Wilkinson *et al.*, 2016). All the data files and the scripts are available from the Zenodo repository https://10.5281/zenodo.6480315.

Table 1. Summary statistics of the dataset (counts).

Statistics	#
Sampling sites	6584
Sampling sites surveyed 10 years or more	1073
Sampling points	14,486
Sampling points surveyed 10 years or more	1084
Sampling surveys	47,869
Occurrences reported	358,157
Taxa reported	125
Individuals reported	20,304,413
Individual lengths measured	8,668,520
Individual weights measured	3,638,793

2 Data description

2.1 Infrastructure

The data presented here come from a database powered by PostgreSQL and hosted at the French Geological Survey (BRGM, https://www.brgm.fr/en). The web interface allows data entry, editing, basic quality control, as well as some exports for tabular data and survey reports.

2.2 Physical data model (PDM)

The comprehensive PDM and the data dictionary are respectively provided as Appendices A and B. Though the overall number of tables exceeds one hundred, the use of the database is rather simple (see Fig. 1) with six "core tables" hierarchically joined. Most other tables are reference tables connecting primary keys to labels and associated features. The complexity of the PDM is a direct consequence of the richness of the data stored with diverse sampling protocols.

In summary, these core tables contain the following information:

- "station": sampling site (name, location). Typically, a site is a stretch spreading several kilometres. It has an Id in the Service d'Administration Nationale des Données et Référentiels sur l'Eau (SANDRE, the water data and frames of reference national administration service; https://www.sandre.eaufrance.fr/). This table has more entries than reported in Table 1 because directly inherited from the SANDRE, and hence encompasses 27,000+ entries without fish data.
- "point_prelevement": sampling point (name, location). Each fish data is related to a georeferenced sampling point. Each "point_prelevement" is not necessarily related to a "station" (sampling site), so, the sampling point is the scale to consider for most of the spatially non-aggregated analyses.
- "operation": sampling survey (date, protocol, objectives, funder, field coordinator). Each survey is related to a point and has a date.
- "prelevement_elementaire": elementary sample. Depending on the protocol, the elementary sample pools together the catches of a pass, of a set of spots, or of a stratified sampling (see details in Sect. 3.1).
- "lot_poissons": fish batch (species, length extrema, batch weight, type of measured length, *e.g.* fork length).
- "mesure_individuelle": individual measurements (length, weight). This table alone represents over 50% of the whole database volume.

2.3 Content

Summary statistics of the database are provided in Table 1. A total of 47,869 surveys were carried out between 1966 and 2021. They report 358,157 taxa occurrences on 14,486 sampling points (1084 of which surveyed at least 10 years) located in 6584 sampling sites (1073 of which surveyed at least 10 years). Hence, there is considerable potential for time-series analyses. Total or fork length was measured in the field from 8,668,520 individuals out of the 20,304,413 ones reported.

The surveys are not spread evenly throughout France (Fig. 2). The variability in the number of surveys between hydrological units is related to unit size and river network density. For example, in southwestern France, the high-permeability sandy soils do not

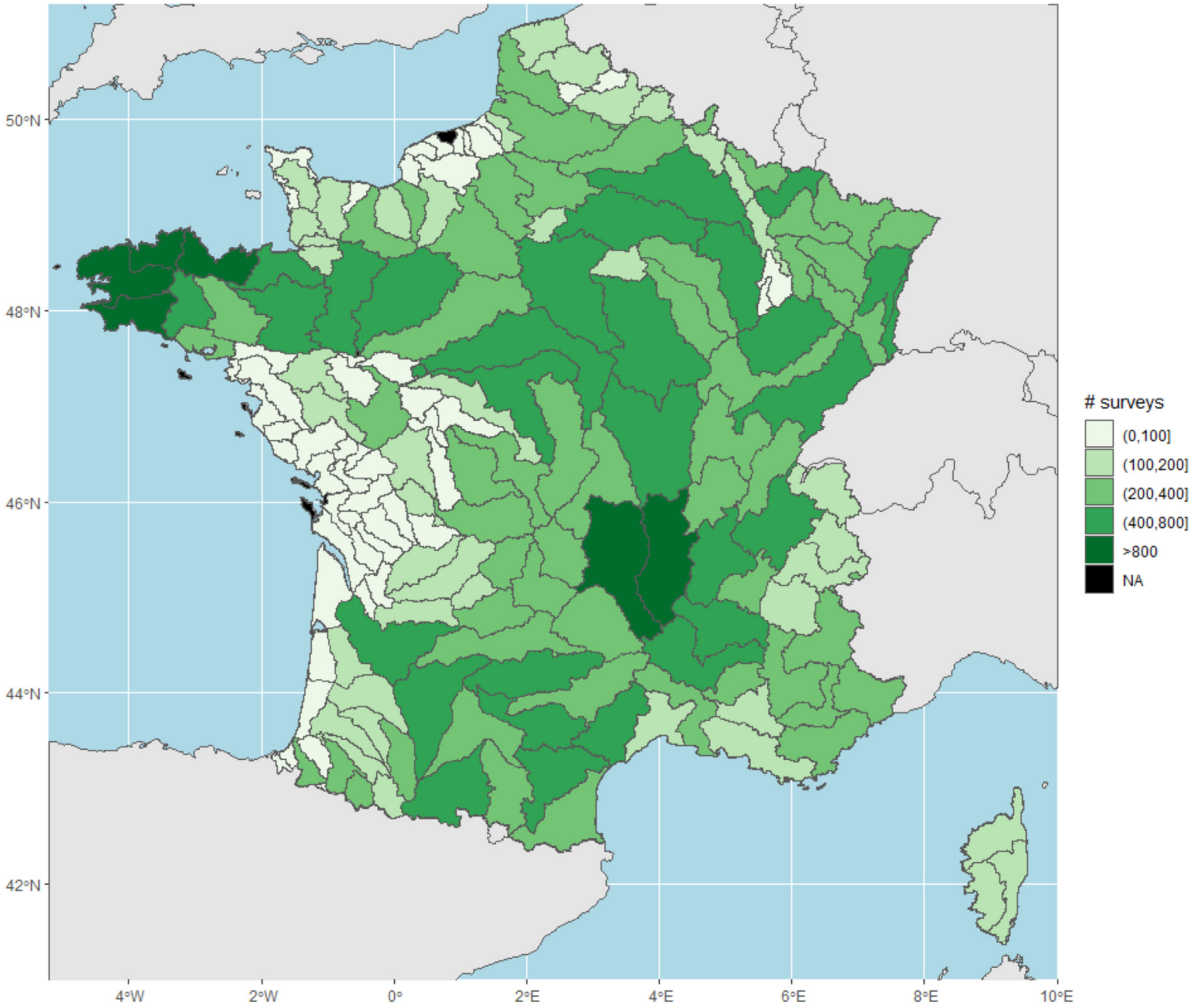

Fig. 2. Map of the overall number of electrofishing surveys carried out by hydrographic unit. The delimitation of hydrographic units was downloaded from the SANDRE WFS service https://services.sandre.eaufrance.fr/geo/sandre, layer 'SecteurHydro_FXX'.

allow sufficient runoff to feed a permanent hydrographic network. These areas, displayed in light green on the map, are less densely sampled only because having basically fewer rivers.

The vast majority of the sampling surveys occurs from May to October with a peak in September, during the low flow season.

3 Data acquisition methods

3.1 Electrofishing protocols

The field sampling protocols have evolved throughout years due to (1) shifts in the objectives pursued, (2) scientific improvement of sampling techniques, (3) attempts to standardise sampling at national and European scales to gain comparability and (4) sampling increase of large river systems (Fig. 3).

Long-term network monitoring of fish assemblages requires standardised methods that provide comparable quality data. All the surveys reported here were carried out by electrofishing following standardised procedures. Two kinds of survey are reported in the database: (1) species-specific surveys (the "abundance indexes", see Fig. 3) that should not be used for the interpretation of other taxa (a taxon not reported does not indicate its absence); (2) multi-species inventories. Among multi-species inventories, complete sampling was performed for small wadable streams with one (single-pass), two, or occasionally three or more passes (multi-pass, see Belliard *et al.*, 2012). For deeper and/or wider streams, several methods have been used, using boats when required; *i.e.* banks continuous sampling (Belliard *et al.*, 2012), banks stratified habitat sampling (Pouilly, 1994) and systematic point sampling (75 or 100 spots sampled, respectively, for medium and large rivers, see Tomanova *et al.*, 2013).

The length of the sampling point is set to include at least one riffle/pool sequence or two meanders, *i.e.* length exceeding 12 to 15 times the mean wetted width

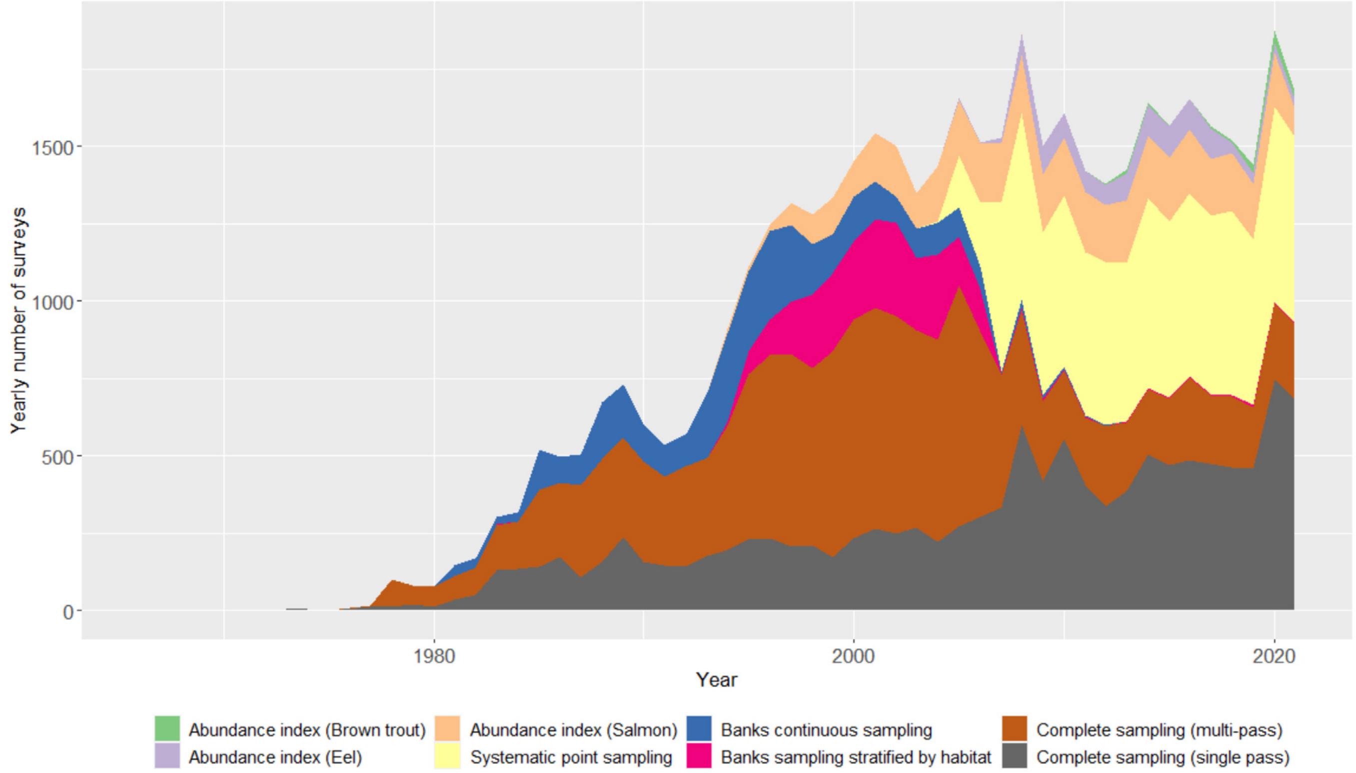

Fig. 3. Temporal changes in the electrofishing sampling protocols.

Table 2. Minimal length (curvilinear longitudinal extent measured along the river central axis) of sampling points, after Belliard *et al.* (2012). Reading: for rivers between 3 and 30 m wide, the length of the sampling point has to exceed twenty times its mean wetted width.

Mean wetted width (MWW)	Minimal length of the sampling point
<3 m	60 m
From 3 m to 30 m	20 × MWW
From 30 m to 60 m	600 m
>60 m	10 × MWW

(variables “odp_longueur” and “odp_largeur_lame_eau”, respectively, in the Table “operation_description_peche”). As required by the European Norm CEN-EN-14011 on sampling of fish with electricity, and to minimise the gaps between the threshold values proposed, the minimal length of sampling point was set after Belliard *et al.* (2012), (Table 2).

The Table “operation_description_peche” documents the conditions of the surveys including the gear type and tuning, along with measures of water conductivity and temperature. Over 90% of the surveys were carried out with direct current.

The “elementary sample” (table “prelevement_elementaire”) is the finest grain of sound faunistic list entered in the database. Depending on the protocol implemented in the field, it pools together the catches from a set of fish batches and individuals according to different rules: (1) In complete single-pass sampling or in banks continuous sampling, the elementary sample is the whole catch for the survey. (2) In complete multi-pass sampling, the elementary sample gathers the catch for a pass. (3) In banks sampling stratified by habitat, it gathers all the catches in a given habitat strata (*e.g.* the riffle habitats). (4) In systematic point sampling, a compulsory elementary sample gathers the catches on the 75 (or 100) spots of the design, and an optional elementary sample gathers the catches on the so-called “supplementary points” that are not included in the base protocol, but that field operators may find useful to sample because of their fine-scale originality (*e.g.* tributary outlet, fallen tree).

3.2 Individuals’ description

Out of the 20,304,413 individuals reported, 20,159,871 (99.29%) were identified to the species or sometimes below (*e.g.* the *Cyprinus carpio* variants). The others were either identified at the genus (or family) level or considered hybrids.

Each entry of the Table “mesure_individuelle” (dedicated to individuals’ measurements) refers to a fish batch in the Table “lot_poissons”. Individual length is reported to the nearest millimeter (total or fork length), and weight is reported to the gram. Out of the 47,869 surveys, 95% report at least one individual fish length and 63% at least one individual fish weight measured in the field.

3.3 Fish-based bioassessment

The dataset encompasses two tables containing bioassessment information, “operation_ipr” and “operation_ipr_plus”, respectively associated to fish-based indices of “river health” initially described in Oberdorff *et al.* (2002) and Marzin *et al.* (2014). These indicators aggregate sets of complementary metrics that measure distances between observed (*i.e.* calculated from the catches) and theoretical (*i.e.* expected in reference conditions) assemblage features (richness or abundance by functional groups). These two tables contain groups of variables describing, for each survey, habitat condition, metrics observed values, metrics theoretical values, as well as the value of the final index. In the ASPE database, the calculation of the two indices is triggered by a first-level validation of the survey and implemented *via* a web service.

3.4 Habitat description

Table “operation_donnees_environnementales” contains a description of the sampling point at the time of a survey, *e.g.* mean depth (“ode_profondeur_moyenne_station”), or dominant substrate (“ode_gra_id_granulometrie_dominante”). The field “ode_hab_id” can be joined with “hab_id” from the Table “habitat” to retrieve an accurate description of the riparian vegetation and fish shelter availability. The additional environmental features required to calculate the fish-based indices (*e.g.* wetted width, slope) are provided in the dedicated Tables “operation_ipr” and “operation_ipr_plus”. The regional variables (*e.g.* surface area of the drainage upstream, distance from sources) are obtained using GIS tools, whereas the local variables (*e.g.* bed dimensions, waterline slope, substrate of the river bed, riparian vegetation) are collected directly in the field.

3.5 Quality process

Through the years, the increased standardisation of field protocols improved the quality of the data by reducing the operator bias. The data entry is followed by a two-step validation procedure. The first validation occurs when all the data for a given survey are entered and checked by the authority in charge of the field sampling. The second one is done by a network of experts who screen the data and ensure their likelihood. The Table “operation_suivi” (that can be joined to the Table “operation”) tracks the validation procedure (field “ops_eta_libelle”) and the overall reliability of the survey is tagged (field “ops_niq_libelle”).

4 Using the dataset

4.1 Licence and ownership

The dataset can be freely downloaded from a Zenodo repository https://10.5281/zenodo.6480315 and used, given that the present article is properly credited. The dataset is published under the Open Licence 2.0 (https://www.data.gouv.fr/fr/pages/legal/licences, SPDX identifier etalab-2.0) and thereby compatible with any open license requiring at least crediting, including Creative Commons Attribution (CC-BY) and Open Data Commons Attribution (ODC-BY) by the Open Knowledge Foundation.

4.2 Formats

The two open formats proposed for the dataset were chosen to keep the volume reasonable (around 200 Mo) and to address a wide array of users. The files’ names include suffixes indicating the date and time when they were created. Office suites users can open the .csv files in a compressed archive named “aspe_csv_yyyy_mm_dd_hh_mm_ss.zip”. However, it should be acknowledged that disentangling the PDM (Appendix A) can be challenging even with the data dictionary (Appendix B). The Table “mesure_individuelle” of individual measurements, with around 20M rows, can not be opened in a spreadsheet. R users will go for the two RData files provided. The first file, named “mei_yyyy_mm_dd_hh_mm_ss.RData”, contains the Table “mesure_individuelle”. The second file, named “tables_sauf_-mei_yyyy_mm_dd_hh_mm_ss.RData”, contains all the other Tables. R users will be greatly assisted in their analyses by two R packages with tutorials (see Sect. 4.5).

4.3 Pitfalls

The complexity of the PDM gives the chance to a wide array of errors when joining the Tables to assemble datasets. Hence it is recommended to take great care in checking the output of the queries. This complexity can, to some extent, be overcome by using the {aspe} R package (see Sect. 4.5). Among the potential pitfalls when using the database, we draw attention on the following peculiarities:

- One survey can contribute to several networks. Hence, adding the “objective” to a set of “operation” leads to duplicate them.
- When a data table includes individual measurements as well as the number of fishes in the batch (field “lop_effectif” of table “lot_poissons”), the latter should not be summed up because being only a replicate.
- Depending on the networks, some sites can be sampled at different frequencies.
- The distinction between a sampling site (Table “station”) and a sampling point (Table “point_prelevement”) can be tricky (but see Sect. 2.2). A sampling site can encompass several sampling points. For most uses, it is recommended to carry out analyses at the point scale. However, it can be useful to aggregate the points within sites, to maximise, for example, the length of time series when analysing temporal trends.
- The taxonomic classification used here is inherited from the 80 s. Recent studies revealed that what was earlier thought of as a species could have to be split into several ones, *e.g.* in the genus *Esox* (Denys *et al.*, 2014), *Phoxinus* (Denys *et al.*, 2020) or *Barbatula* (Gauliard *et al.*, 2019). Unfortunately, there is no correspondence included in the database with TAXREF, the French national taxonomic register (Gargominy *et al.*, 2021). However, the gateway Table “passerelle_taxo” of the {aspe} package allows joining the two nomenclatures. The species scientific names included are not all up-to-date nor standardised, so great care is advised for joining external data, such as those imported from Fishbase (www.fishbase.us).
- The naming rules applied to the internal elements of the database (tables and fields) refer to French language, which may cause understanding problems to not French speaking users.

4.4 Coordinates reference systems (CRS)

Several CRS appear in the database. All the sampling points, as well as most of the sampling sites, are georeferenced. Their longitude and latitude coordinates can be found in the

Table 3. Number of individual fishes reported by species, along with their Red List status (IUCN, 2019). CR: critically endangered, EN: endangered, VU: vulnerable, NT: near threatened, LC: least concern, LR/lc: lower risk / least concern.

Species name	Nb indiv.	Red List FR	Red List World
Alosa alosa	65	CR	LC
Anguilla anguilla	482,229	CR	CR
Cottus petiti	596	CR	VU
Zingel asper	271	EN	CR
Squalius laietanus	262	EN	LC
Misgurnus fossilis	2143	EN	LC
Petromyzon marinus	3134	EN	LC
Esox lucius	38,060	VU	LC
Lota lota	7989	VU	LC
Lampetra fluviatilis	937	VU	LC
Thymallus thymallus	16,190	VU	LC
Alosa fallax	34	NT	LC
Barbus meridionalis	57,953	NT	NT
Gobio lozanoi	4115	NT	LC
Cobitis taenia	16,368	NT	LC
Salmo salar	298,818	NT	LR/lc
Parachondrostoma toxostoma	53,067	NT	VU
Phoxinus bigerri	348	NT	LC
Leuciscus burdigalensis	9315	NT	LC

Tables “station” (“sta_coordonnees_x” and “sta_coordonnees_y”) and “point_prelevement” (“pop_coordonnees_x” and “pop_coordonnees_y”), respectively. Each pair of coordinates is associated to a CRS in fields “sta_typ_id” and “pop_typ_id”. The correspondence between the CRS Id, its name (*e.g.* WGS 84) and its EPSG code (4326 for WGS84) is provided in the Table “ref_type_projection”.

4.5 Supporting analytical tools

Two R packages (Wickham, 2015) are developed to foster the analysis of the data. Though still not definitive, they allow easy exploitation of the database without having to memorise the PDM. These packages have been used extensively to produce the present manuscript.

The first package ({aspe} package) provides a suite of tools for most of the common processing of the ASPE database including parsing a dump of the ASPE PostgreSQL database, calculation of abundances, densities, size distributions, along with graphical output and spatial processing. It is available at: https://github.com/PascalIrz/aspe. This package includes three useful dataframes named “liste_rouge”, “passerelle_taxo” and “traits_bio”, respectively giving, for each taxa, (1) its global, European and national IUCN Red List status (IUCN *et al.*, 2019; see Table 3), (2) a taxonomic gateway joining scientific and common names with the ASPE, TAXREF (Gargominy *et al.*, 2021) and SANDRE Ids and (3) a few species life-history traits. The Github page of the package indicates several links to tutorials as well as explanations on the functions’ naming rules. Assembling a custom ‘dataframe’ is assisted by a series of functions allowing joining the ASPE tables. The function “mef_creer_passerelle()” starts by joining the six “core” tables Ids, then other functions with the “mef_” prefix allow joining other Tables to add information.

The second package ({aspeQual} package) provides quality control tools for the database. So far it is limited to checking the likelihood of individual measurements and is available at: https://github.com/PascalIrz/aspeQual.

5 Usage examples

5.1 Species length-weight relationships

The aim here is to process the data in order to fit a model in the classical form for a given species:

$$Weight = a{\cdot}\mathrm{Length}^b.$$

A link to a comprehensive step-by-step tutorial is available from the {aspeQual} package Github repository (see Sect. 4.5). This tutorial also shows how to process all the taxa at once and how to confront the results with those already published in Fishbase (https://www.fishbase.se/search.php). As shown in Table 4, the number of individuals measured for some common species is quite remarkable.

5.2 Community dynamics

The length of the time-series allows insights into the fish assemblage dynamic. On Figure 4, only 6 out of the 15 species reported (Eurasian minnow *Phoxinus phoxinus*, stone loach *Barbatula barbatula*, European brook lamprey *Lampetra planeri*, bullhead *Cottus gobio*, brown trout *Salmo trutta* and gudgeon *Gobio gobio*) constantly occur on this site of the Tardes River. The roach *Rutilus rutilus* was reported all years but in 1999 and 2008. The other species are reported only occasionally. This type of plot, showing the temporal turnover of local fish assemblages, is easy to produce with the {aspe} R package.

Table 4. Example of length-weight relationships obtained from the ASPE data (total length in cm, weight in g) under a power equation form. Equation parameters (a, b), adjusted r-squared (Adj. r^2) and number of individuals measured and weighted (Nb indiv.) are given for each species.

Species	*a*	*b*	Adj. r^2	Nb indiv.
Salmo trutta	0.011	2.985	0.984	407,381
Leuciscus cephalus	0.009	3.069	0.987	160,288
Rutilus rutilus	0.007	3.161	0.972	82,330
Barbus barbus	0.012	2.908	0.989	48,240
Gobio gobio	0.014	2.890	0.886	42,167
Perca fluviatilis	0.010	3.067	0.973	27,599
Leuciscus leuciscus	0.009	3.048	0.974	20,920
Telestes souffia	0.014	2.869	0.921	20,614
Cottus gobio	0.046	2.476	0.824	19,956
Chondrostoma nasus	0.007	3.074	0.987	15,615
Barbus meridionalis	0.014	2.943	0.968	12,840
Lepomis gibbosus	0.016	3.113	0.924	11,820
Alburnus alburnus	0.029	2.462	0.800	9,706
Esox lucius	0.007	2.987	0.987	9,535
Parachondrostoma toxostoma	0.010	2.971	0.953	7,144

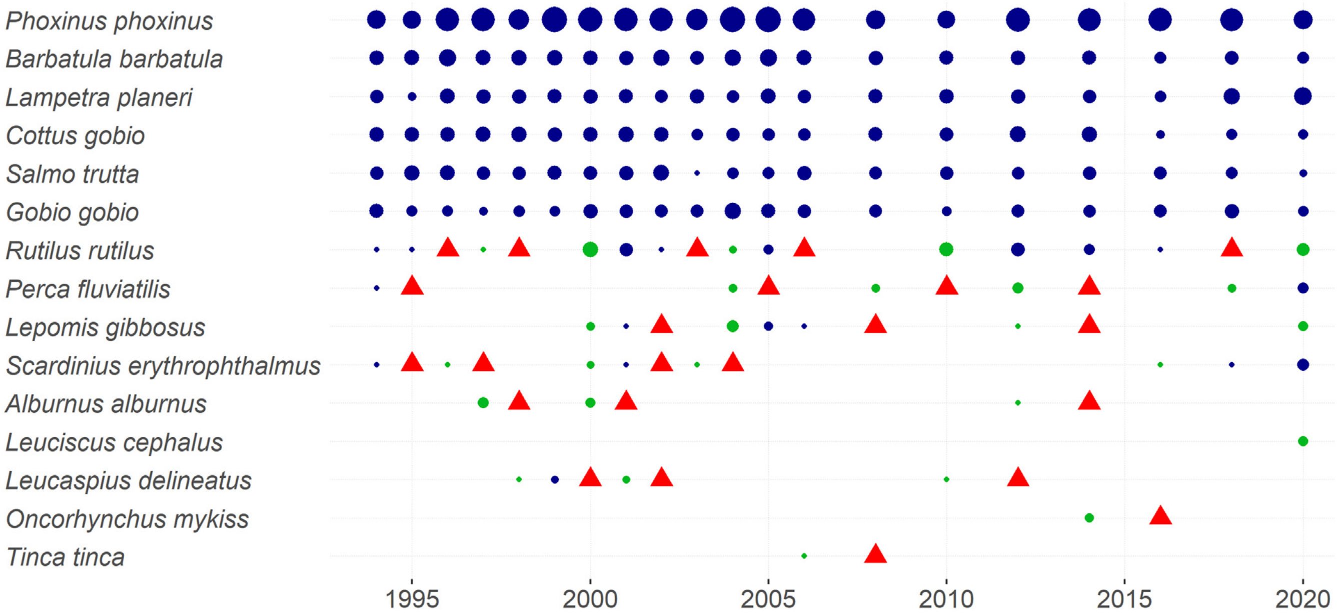

Fig. 4. Example of temporal community dynamic (1994-2020) at the 'Tardes' sampling site (center France). The red triangles indicate local 'extinctions', *i.e.* the species is absent despite its presence at the previous survey. The green bubbles are local 'colonisations'. The area of the green and blue bubbles is proportional to the number of individuals caught during each survey for each species. The species are ordered by their cumulated catch numbers over the period. Note that over 2006, the site was sampled only once every two years.

5.3 Bioassessment

Figure 5 displays the temporal trend in the Fish-based Index scores (FBI, see Oberdorff *et al.*, 2002) for a set of 326 sites regularly sampled throughout mainland France between 2001 and 2021. The FBI median scores and their 75 percentiles (below the median on the plot) improved between 2001 and 2010–2011 and then stabilized. Note that no clear signal appears for the 25 percentiles (above the median on the plot). This suggests that the overall improvement is driven by sites originally suffering the worst conditions. However, this plot is displayed only to exemplify the use of the dataset. A rigorous evaluation of the trend slope and significance would require a more refined statistical analysis.

6 Discussion

6.1 ASPE for research

The dataset is still scientifically under-exploited. All the records in ASPE being georeferenced, the spatial joining with complementary datasets offers promising multidisciplinary perspectives.

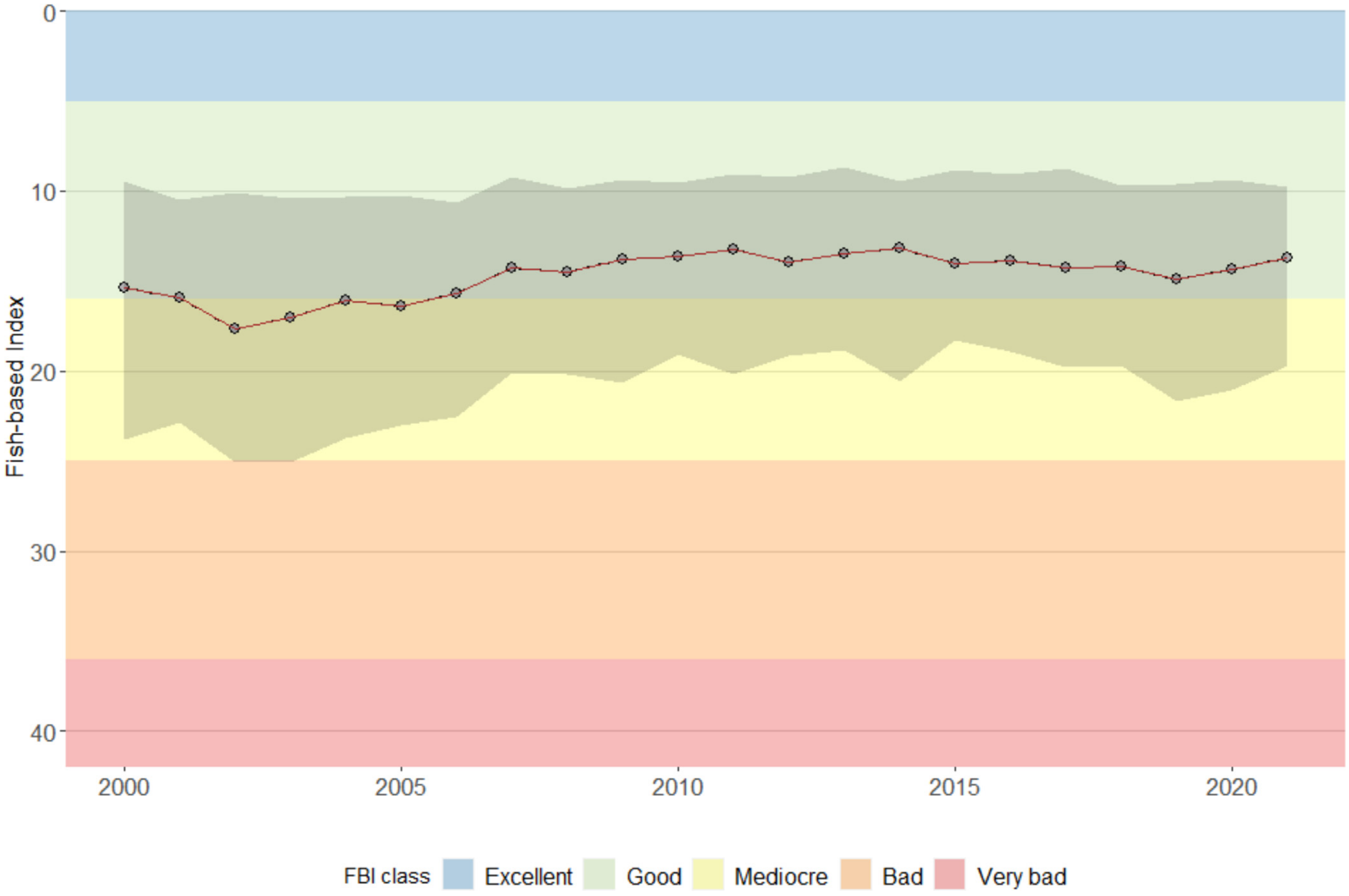

Fig. 5. Temporal bioassessment trends. The background colours indicate the local 'ecosystem health' according to Oberdorff *et al.* (2002). The line and the dots indicate the annual median fish-based index (FBI) across lowland sites in mainland France (n=326 sites under 500m in altitude, sampled at least 10 years with a first survey before 2004 and a last survey after 2017). The shaded area indicates the interquartile range.

For example, the crayfish data have hardly been explored despite a total 158, 381 individuals reported. Only 19% of them belong to native species. The most frequent exotic ones (*Pacifastacus leniusculus* and *Faxonius limosus*) have been introduced from North America. Considering autochtonous species, the stone crayfish *Austropotamobius torrentium*, assessed "Critically Endangered" by the French Red List (IUCN and MNHN, 2012) was never observed, which confirms this status. The "Endangered" noble crayfish *Astacus astacus* was very rarely observed but the white-clawed crayfish *Austropotamobius pallipes* (classified "Vulnerable") was comparatively more common.

Another hardly-exploited data (to our knowledge) is the sanitary records on individuals. Externally visible pathologies are reported for 57,387 individuals, *e.g.* ectoparasites ($n = 6481$ individuals) or fins or tegument alterations ($n = 7701$ individuals). Though this level of accuracy is far from being systematic, the data may be worth an epidemiological study. Some brutal collapses in Cyprinids populations in the Seine River have, for example, been attributed to outbreaks of bucephalosis (Lambert, 1997), a parasitic disease due to *Bucephalus polymorphus* infestation that can be identified by external inspection.

The length of the time series, along with the availability of individuals size spectra, offer extensive perspectives in population dynamics (*e.g.* synchronicity, response to temperature or hydrologic regime changes, density-dependence).

The database also offers perspectives in the field of bioassessment. The fish-based indices that have been developed on the ancestor of ASPE (Oberdorff *et al.*, 2002; Marzin *et al.*, 2014) did the best with the data available at that time. However, in the recent years, the dataset improved in quality, with more protocol standardisation, as well as in quantity. We also, nowadays, have access to new data on habitat, *e.g.* temperature extrema (Beaufort *et al.*, 2021) and anthropogenic stressors (*e.g.* pollution monitoring). Therefore, renewing the development procedure with up-to-date data and methods could, most likely, allow substantial improvement of the existing bioassessment tools.

6.2 Riverine fish data in France tomorrow

Data collection goes on. The aims of the surveys, the sampling protocols and the networks of sites are now settled. Hence, the perspective is that of longer time series of standardised surveys of sites representative of the diversity of rivers in France.

Open access to the up-to-date data should be available through an API by the end of 2022. The ASPE database will also become more interoperable with the attribution of a Universally Unique Identifier (UUID) to each observation and to each survey from the very data entry.

The data dictionary (Appendix B), along with the PDM, can help the users find their ways among the tables. Translating the 720 entries of the data dictionary from French to English should now be a matter of high priority to simplify access for potential foreign scientific users.

6.3 Taxonomic identification: the Achilles' heel?

Due to animal welfare considerations, taxonomic identification is carried out quickly in the field, so that individuals can be released safe and sound. So, misidentification may occur, especially for small individuals or cryptic species. The electrofishing protocols implemented do not target larvae, however, juveniles are captured and can be tricky to identify. Furthermore, many fish species can hybridise. Hybrid individuals can be difficult to spot without molecular tools, *e.g. Parachondrostoma toxostoma* x *Chondrostoma nasus* (Costedoat *et al.*, 2004) or *Esox lucius* x *E. aquitanicus* (Jeanroy and Denys, 2019). Eventually, the last decade saw an important taxonomic review with the description of cryptic species that have only partially been implemented in the ASPE database. Even species known for a while, and included in the taxonomic reference of the database (table ref_espece), are not systematically identified by field staff; *e.g.* the rostrum dace *Leuciscus burdigalensis* is still often reported as the common dace *L. leuciscus*. A control with up-to-date species distribution maps such as those of Keith *et al.* (2020) should help correct most of these misidentifications.

Fundings

The database engineering and administration are funded by the Office Français de la Biodiversité (OFB, https://www.ofb.gouv.fr/). The non-OFB data providers freely feed the database.

Conflicts of interest

There is no potential conflict of interest relevant to this article.

Supplementary Material

Appendix 1: Database Physical Data Model.
Appendix 2: Data dictionary.

The Supplementary Material is available at https://www.kmae-journal.org/10.1051/kmae/2022021/olm.

Acknowledgements. Though the database does not keep track of that, more than a thousand of persons have contributed to the surveys, either directly in the field or by providing organisational and logistic supports. They are all greatly acknowledged with special thanks to the OFB field staff, survey directors, engineers in charge of the monitoring programs and data managers.

Decades ago, a few visionaries understood, before the mainstream, the virtues of establishing a perennial monitoring network to track biological stream conditions across time. Today, we owe them the legacy of a major dataset. We hope budget cuts and other short-term considerations will not hamper future data collection and subsequent fundamental, as well as applied, knowledge production.

The comments provided by an anonymous referee and by the Editor-in-Chief allowed substantial improvements of the manuscript.

References

Allen GH, Pavelsky TM. 2018. Global extent of rivers and streams. *Science* 361: 585–588.

Beaufort A, Diamond JS, Sauquet E, Moatar F. 2022. Spatial extrapolation of stream thermal peaks using heterogeneous time series at a national scale. *Hydrol Earth Syst Sci* 26: 3477–3495.

Belliard J, Ditche JM, Roset N, Dembski S. 2012. Guide pratique de mise en oeuvre des opérations de pêche à l'électricité. Onema, Vincennes, 31 p.

Bret V, Bergerot B, Capra H, Gouraud V, Lamouroux N. 2016. Influence of discharge, hydraulics, water temperature, and dispersal on density synchrony in brown trout populations (*Salmo trutta*). *Can J Fish Aquat Sci* 73: 319–329.

Bret V, Capra H, Gouraud V, *et al.* 2017. Understanding inter-reach variation in brown trout (*Salmo trutta*) mortality rates using a hierarchical Bayesian state-space model. *Can J Fish Aquat Sci* 74: 1612–1627.

Buisson L, Grenouillet G. 2009. Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. *Divers Distrib* 15: 613–626.

Buisson L, Thuiller W, Lek S, Lim P, Grenouillet G. 2008. Climate change hastens the turnover of stream fish assemblages. *Glob Chang Biol* 14: 2232–2248.

Cattanéo F, Lamouroux N, Breil P, Capra H. 2002. The influence of hydrological and biotic processes on brown trout (*Salmo trutta*) population dynamics. *Can J Fish Aquat Sci* 59: 12–22.

Clavel J, Poulet N, Porcher E, *et al.* 2013. A new freshwater biodiversity indicator based on fish community assemblages. *PLOS ONE* 8: e 80968.

Comte L, Grenouillet G. 2013. Do stream fish track climate change? Assessing distribution shifts in recent decades. *Ecography* 36: 1236–1246.

Comte L, Murienne J, Grenouillet G. 2014. Species traits and phylogenetic conservatism of climate-induced range shifts in stream fishes. *Nat Commun* 5: 5053.

Comte L, Hugueny B, Grenouillet G. 2016. Climate interacts with anthropogenic drivers to determine extirpation dynamics. *Ecography* 39: 1008–1016.

Comte L, Carvajal-Quintero J, Tedesco PA, *et al.* 2021. RivFishTIME: a global database of fish time-series to study global change ecology in riverine systems. *Glob Ecol Biogeogr* 30: 38–50.

Costedoat C, Pech N, Chappaz R, Salducci MD, Lim P, Gilles A. 2004. Étude de l'hybridation introgressive entre *Chondrostoma t. toxostoma* et *chondrostoma n. nasus* (Téléostéen, Cyprinidae) en utilisant une approche multiple. *Cybium* 28: 51–61.

Denys GPJ, Dettai A, Persat H, Hautecoeur M, Keith P. 2014. Morphological and molecular evidence of three species of pikes *Esox* spp. (Actinopterygii, Esocidae) in France, including the description of a new species (vol 337, pg 521, 2014). *C R Biol* 337: 657–659.

Denys GPJ, Dettai A, Persat H, Daszkiewicz P, Hautecœur M, Keith P. 2020. Revision of *Phoxinus* in France with the description of two new species (Teleostei, Leuciscidae). *Cybium* 44: 205–237.

Dudgeon D. 2019. Multiple threats imperil freshwater biodiversity in the Anthropocene. *Curr Biol* 29: R960–R967.

Dudgeon D, Arthington AH, Gessner MO, *et al*. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev* 81: 163–182.

Gargominy O, Tercerie S, Régnier C, *et al*. 2021. TAXREF v13, référentiel taxonomique pour la France: Méthodologie, mise en œuvre et diffusion. Rapport UMS PatriNat (OFB-CNRS-MNHN), Paris, 64 p.

Gauliard C, Dettai A, Persat H, Keith P, Denys GPJ. 2019. *Barbatula leoparda* (Actinopterygii, Nemacheilidae), a new endemic species of stone loach of French Catalonia. *Cybium* 43: 169–177.

Hughes K. 2021. The World's forgotten fishes. World Wide Fund for Nature (WWF), 47 p.

Hugueny B, Movellan A, Belliard J. 2011. Habitat fragmentation and extinction rates within freshwater fish communities: a faunal relaxation approach. *Glob Ecol Biogeogr* 20: 449–463.

Ihaka R, Gentleman R. 1996. R: a language for data analysis and graphics. *J Comput Graph Stat* 5: 299–314.

IUCN, MNHN. 2012. Crustacés d'eau douce de France métropolitaine. Paris, France. La Liste rouge des espèces menacées en France.

IUCN, MNHN, SFI, AFB. 2019. Poissons d'eau douce de France métropolitaine. La Liste rouge des espèces menacées en France, Paris.

Jeanroy C, Denys GPJ. 2019. Morphological traits allow distinguishing their hybrids from the Northern pike, *Esox lucius*, and the Aquitanian pike, *Esox aquitanicus* (Actinopterygii, Esociformes). *Cybium* 43: 227–232.

Keith P, Poulet N, Denys G, Changeux T, Feunteun É, Persat H. 2020. Les Poissons d'eau douce de France. Deuxième édition. Muséum national d'Histoire naturelle, Paris & Biotope, Mèze, 704 p.

Lambert A. 1997. Introduction de poissons dans les milieux aquatiques continentaux: «Quid de leurs parasites?». *Bull Fr Pêche Piscic* 344-345: 323–333.

Manné S, Poulet N. 2008. First record of the western tubenose goby *Proterorhinus semilunaris* (Heckel, 1837) in France. *Knowl Manag Aquat Ecosyst* 1–5.

Marzin A, Archaimbault V, Belliard J, Chauvin C, Delmas F, Pont D. 2012. Ecological assessment of running waters: do macrophytes, macroinvertebrates, diatoms and fish show similar responses to human pressures? *Ecol Indic* 23: 56–65.

Marzin A, Delaigue O, Logez M, Belliard J, Pont D. 2014. Uncertainty associated with river health assessment in a varying environment: the case of a predictive fish-based index in France. *Ecol Indic* 43: 195–204.

Oberdorff T. 2022. Time for decisive actions to protect freshwater ecosystems from global changes. *Knowl Manag Aquat Ecosyst* 19.

Oberdorff T, Hugueny B, Compin A, Belkessam D. 1998. Non-interactive fish communities in the coastal streams of North-western France. *J Anim Ecol* 37: 472–484.

Oberdorff T, Hugueny B, Vigneron T. 2001a. Is assemblage variability related to environmental variability? An answer for riverine fish. *Oikos* 93: 419–428.

Oberdorff T, Pont D, Hugueny B, Chessel D. 2001b. A probabilistic model characterizing fish assemblages of French rivers: A framework for environmental assessment. *Freshw Biol* 46: 399–415.

Oberdorff T, Pont D, Hugueny B, Porcher J-P. 2002. Development and validation of a fish-based index for the assessment of 'river health' in France. *Freshw Biol* 47: 1720–1734.

PatriNat. 2019. Biodiversité d'intérêt communautaire en France : Un bilan qui reste préoccupant – Résultats de la troisième évaluation des habitats et espèces de la DHFF (2013-2018) – Note de synthèse, 4 p.

Peng RD. 2011. Reproducible research in computational science. *Science* 334: 1226–1227.

Pouilly M. 1994. Relations entre l'habitat physique et les poissons des zones à Cyprinidés rhéophiles dans trois cours d'eau du bassin rhodanien: Vers une simulation de la capacité d'accueil pour les peuplements. [Thèse de doctorat], Université Claude Bernard – Lyon I, 256 p.

Poulet N, Beaulaton L, Dembski S. 2011. Time trends in fish populations in metropolitan France: insights from national monitoring data. *J Fish Biol* 79: 1436–1452.

Reid AJ, Carlson AK, Creed IF, *et al*. 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol Rev* 94: 849–873.

Reyjol Y, Hugueny B, Pont D, *et al*. 2007. Patterns in species richness and endemism of European freshwater fish. *Glob Ecol Biogeogr* 16: 65–75.

Santos R, Poulet N, Besnard A. 2021. Life-history traits correlate with temporal trends in freshwater fish populations for common European species. *Freshw Biol* 66: 317–331.

Su G, Logez M, Xu J, Tao S, Villéger S, Brosse S. 2021. Human impacts on global freshwater fish biodiversity. *Science* 371: 835–838.

Tales E, Keith P, Oberdorff T. 2004. Density-range size relationships in French riverine fishes. *Oecologia* 138: 360–370.

Tomanova S, Tedesco PA, Roset N, Berrebi dit Thomas R, Belliard J. 2013. Systematic point sampling of fish communities in medium- and large-sized rivers: Sampling procedure and effort. *Fish Manag Ecol* 20: 533–543.

Vörösmarty CJ, McIntyre PB, Gessner MO, *et al*. 2010. Global threats to human water security and river biodiversity. *Nature* 467: 555–561.

Wickham H. 2015. R Packages: Organize, test, document, and share your code, O'Reilly Media, 201 p.

Wilkinson MD, Dumontier M, Aalbersberg IjJ, *et al*. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3: 160018.

Xie Y, Allaire JJ, Grolemund G. 2018. R Markdown: the definitive guide, Chapman; Hall/CRC, London, 338 p.

Cite this article as: Irz P, Vigneron T, Poulet N, Cosson E, Point T, Baglinière E, Porcher J-P. 2022. A long-term monitoring database on fish and crayfish species in French rivers. *Knowl. Manag. Aquat. Ecosyst.*, 423, 25.