Dataset on European diadromous species distributions from 1750 to present time in Europe, North Africa and the Middle East

Betsy Barber-O'Malley, Géraldine Lassalle, Patrick Lambert, Eric Quinton

Data in Brief 40 (2022) 107821

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset on European diadromous species distributions from 1750 to present time in Europe, North Africa and the Middle East

Betsy Barber-O'Malley[a,b,1], Géraldine Lassalle[a,*], Patrick Lambert[a], Eric Quinton[a]

[a] INRAE, UR EABX, 50 Avenue de Verdun, Gazinet-Cestas Cedex 33612, France
[b] University of Maine, 168 College Avenue, Orono, ME 04469, United States

ARTICLE INFO

Article history:
Received 5 June 2021
Revised 22 December 2021
Accepted 7 January 2022
Available online 11 January 2022

Keywords:
Diadromous species
Geographical distribution
Presence/absence
Abundance
Life-history trait
Population
European catchments

ABSTRACT

EuroDiad version 4.0 is a set of data tables that store information about the presences/absences and population functionality of diadromous species (lampreys and fish) populations in selected catchments in Europe, the Middle East, and North Africa from 1750 to present time. This database contains distribution and life-history trait information for twenty-eight European diadromous species and geomorphological data for each of the selected catchments, though not every species has data for every catchment and time period. EuroDiad was originally created in 2005–2006 (EuroDiad 1.0 and 2.0), and contained data for 196 catchments and two time periods (1851–1950 and 1951–2010). It underwent a major update in 2009–2010 (EuroDiad 3.2) through a validation process by European fisheries experts. Version 3.2 included the addition of 63 small-sized catchments (< 10,000 km^2) and an additional time period (1751–1850) for select species and catchments. This database underwent a second validation process in 2019–2020 and was updated to v 4.0, with the primary goal of providing infor-

DOI of original article: 10.1016/j.biocon.2008.02.010
* Corresponding author.
E-mail address: geraldine.lassalle@inrae.fr (G. Lassalle).
Social media: (P. Lambert)
[1] Present address: Gomez and Sullivan Engineers, 41 Liberty Hill Road, Henniker, NH 03242, USA.

https://doi.org/10.1016/j.dib.2022.107821

mation for a new generation of species distribution models, referred to as hybrid models, which incorporate both habitat suitability and population dynamics within their framework. Secondary objectives of this update were to: (a) incorporate new catchments for which information was provided by additional experts, (b) validate existing information about the presences or absences of diadromous species and categorize their population functionality within a catchment, and (c) perform data hygiene to prepare the database for broad dissemination. Information on the life history, morphology, and phenology of four emblematic species (i.e. eel, salmon, lamprey and shad) were added in this occasion. Data for this update were validated by DiadES project partners (www.diades.eu) and local experts. This update was focused on catchments located in the Atlantic Area for use in the DiadES project. Data were divided by country, and validation was performed for catchments in Ireland, the U.K., Spain, Portugal, and France under the supervision of national organisations in fisheries and environmental management. DiadES project partners were asked to validate geomorphological information for the catchment (location of the outlet, surface area of the drainage basin, length of the main watercourse, elevation at the headwaters), as well as the presences/absences information and population functionality categories for all species already present in EuroDiad for their country. If possible, verification was done for each of the three time periods. Partners were also asked to provide data for any other catchments for which they had access to information on fish population status. EuroDiad 4.0 now stores data for 350 catchments (of which 292 have population functionality records) and three time periods, though the precision of information varies and not every species has information for each time period. This validation process strengthened the usefulness of EuroDiad, which is now updated and available for use by the research community.

Specifications Table

Subject	Ecology
Specific subject area	The EuroDiad database provides information about the distribution and life history of diadromous species to use in species distribution modelling studies.
Type of data	Table
How data were acquired	The date came from a thorough review of the primary literature on diadromous species. Data were validated twice by different groups of regional diadromous species experts. The database was initially developed in Microsoft ACCESS, and then was reformatted in PostgreSQL.
Data format	Raw
Parameters for data collection	Data extracted from the primary literature had to pertain to three major categories: (1) species distribution within selected catchments, (2) catchment-specific descriptive information, and (3) species traits in the selected catchments.

(continued on next page)

Description of data collection	Catchment-specific values for the surface area of the drainage basin, length of the main watercourse, and altitude of the source were compiled from Wikipedia and online and physical Atlases. Species presences/absences, population functionality and life-history traits were recorded from the primary scientific literature. Digital libraries were a key tool to prospect for archives materials. Data on distributions and catchment physical characteristics were reviewed and validated twice by experts in the field of diadromous species biology and ecology. The database was reformatted from Access to PostgreSQL database.
Data source location	For each entry, all the primary data sources could be found in the field "bibliography" in the "basin", "abundance" and "species traits" tables. Full references were provided in the online repository under two formats RIS and BibTeX. These references were published by organisations from different countries: Albania, Algeria, Azerbaijan, Belgium, Bulgaria, Croatia, Denmark, United Kingdom, Egypt, France, Finland, Germany, Georgia, Greece, Iceland, Iran, Ireland, Italy, Kazakhstan, Latvia, Lithuania, Morocco, Netherlands, Norway, Poland, Portugal, Romania, Russia, Spain, Sweden, Tunisia, Turkey and Ukraine.
Data accessibility	Data is hosted on a public repository and accessed through a DOI. Repository name: data.inrae.fr Data identification number: doi:10.15454/IVVAIC Direct URL to data: https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/IVVAIC Instructions for accessing these data: The data is provided in three formats, two PostgreSQL database dumps and a data extract in CSV format. Information to obtain the data are described into this file: https://data.inrae.fr/api/access/datafile/106161
Related research article	Lassalle G., Béguer M., Beaulaton L. and Rochard E. (2008) Diadromous fish conservation plans need to consider global warming issues: an approach using biogeographical models. Biological Conservation 141(4): 1105–1118. doi: 10.1016/j.biocon.2008.02.010

Value of the Data

- This dataset is a unique compilation of presences/absences and functionality information (reproductive capacity) for diadromous species populations in 350 catchments in Europe, the Middle East, and North Africa for three time periods (1751–1850, 1851–1950, and 1951-present times).
- Scientists studying diadromous species and freshwater biodiversity at large scale can use this dataset.
- This dataset can be used in a range of studies focusing on diadromous species distributions, including estimating the effects of climate change and other anthropogenic pressures in relation with connectivity, changes in population dynamics, and exploring ecosystem services related to fish populations in Europe.
- This dataset also contains life-history trait information for four valuable species (Salmo salar, Alosa alosa, Petromyson marinus, Anguilla anguilla) across their distribution range. The entire life cycle was covered including the shift from continental to marine habitats.
- Data were produced after an extensive scientific literature review based on digital libraries and other electronic resources, and two expert validation processes.

1. Data Description

1.1. General dataset structure

EuroDiad v. 4.0 is a set of data tables that store information pertaining to the presences/absences, life-history traits and functionality of populations in a selection of catchments in

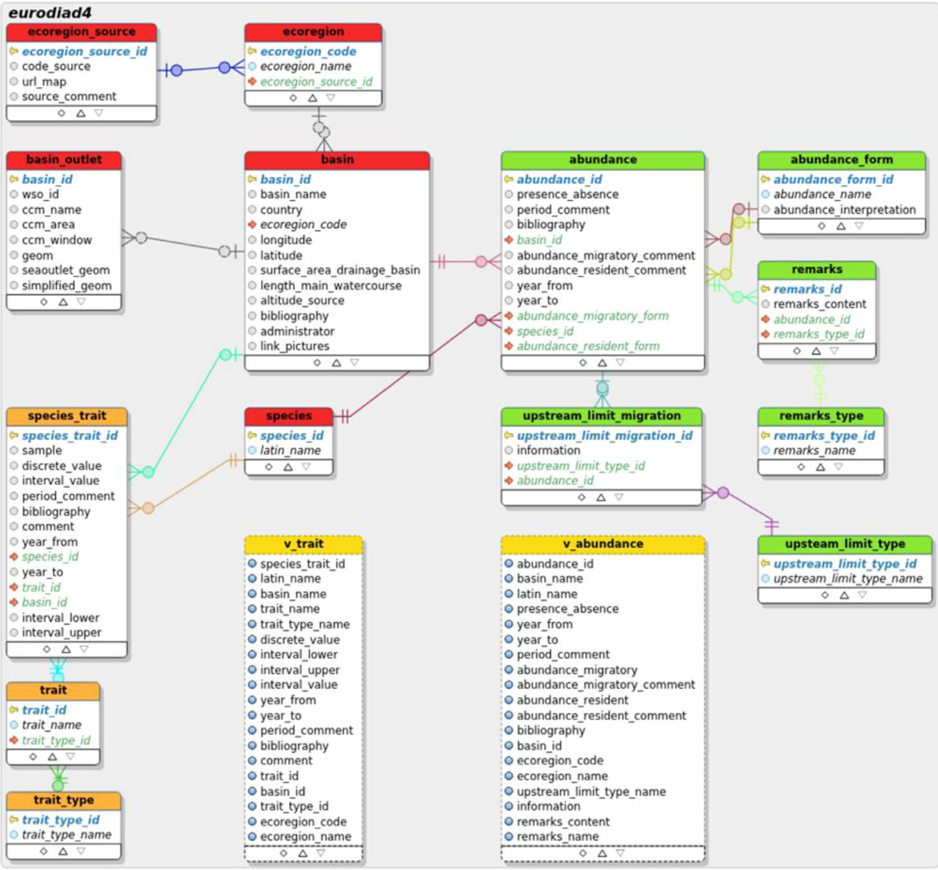

Fig. 1. EuroDiad 4.0 relational schema. Arrows indicate connections/relationships between tables. Views are represented with yellow, and are designated with a "v" before the table name.

Europe, the Middle East, and North Africa (Fig. 1). Here, functionality refers to the reproductive capacity of a population, which is discussed in more detail in the next section. Throughout the data collection and updating process, values for the surface area of the drainage basin, length of the main watercourse, and altitude of the source were compiled using primary literature, specialised websites, and expert validation. Presence/absence designation was also taken from the primary literature and represented three 100-year time periods (1751–1850, 1851–1950, 1951-present time). Life-history traits were extracted from the primary literature but also from reports and personal communications of Diadfish network experts (www.diadfish.org). The time range covered by the data is 1905–2006. All the data sources were made available in dedicated fields named "bibliography" with full references in the online repository.

This dataset was compiled and entered in 2005–2006 in Microsoft Access format (EuroDiad 1.0 and 2.0), and underwent a major update in 2009–2010 (EuroDiad 3.2). The database was then reformatted to PostgreSQL (The PostgreSQL Global Development Group, 1996–2020), which is an open-source relational database management system. The most recent version (EuroDiad 4.0) consists of multiple tables, all stored within one schema in PostgreSQL (Fig. 2). These tables are connected using primary keys, which provide unique identifiers (UIDs) for each data record within a table. Queries can be written to extract specific information from one or multiple tables, and a "view" can be created that permanently combines information from multiple tables. When

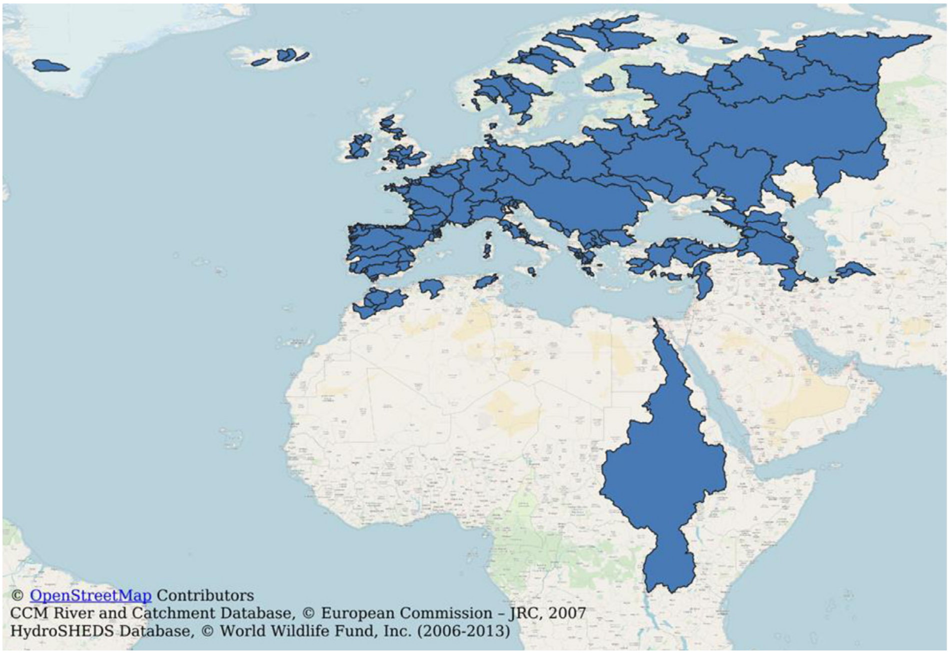

Fig. 2. Map of the 350 catchments in Europe, North Africa and the Middle East in EuroDiad 4.0.

new data is added to a table, the information in this view is updated along with any other tables whose columns are included in the updated information.

Data were divided into multiple tables (Fig. 2) which were grouped by intended use. These included a table for catchment information, distribution records, a table for species-specific life history traits, and one listing the upstream limit of captures within each catchment. In many instances, field values are coded as a numeric index rather which was associated with an explicit wording in a separate table. In the different "views", the numeric index with its explicit description is given so that the user does not have to refer back to the associated tables. Two views have been created and permanently stored in EuroDiad 4.0, one to highlight distribution information and one to highlight life-history trait information (Fig. 2).

1.2. Catchment–specific information

Literature-derived information associated with each catchment was initially stored in the tables "basin" included the geographic location (ecoregion, country, latitude and longitude of the outlet), as well as surface area of the drainage basin, length of the main watercourse, and altitude of the source (Table 1). Recently, more precise information on the catchment sea outlets and surface areas were derived in accordance with two mapping products, i.e. (1) the Catchment Characterisation Model (CCM2) database [1] (https://ccm.jrc.ec.europa.eu/), and (2) HydroSHEDS [2] (https://www.hydrosheds.org/), and stored in "basin_outlet" (Table 2).

For the majority of catchments, the point used for longitude and latitude of the outlet was defined as being aligned with the coastline, rather than further upstream at the head of tide (Table 2). However, several catchments consisted of two large rivers that merged upstream of what would be considered the outlet using this definition. In these instances, where conjoined rivers were considered as separate catchments, the point of confluence was defined as the outlet for both.

Table 1
"Basin" table with information recorded for each catchment (referred to as basin) in the Atlantic Area, as well as several catchments in Europe, the Middle East, and North Africa.

Column name	Description of data
basin_id	Unique key associated with each basin
basin_name	Name of the basin
country	Country at the outlet (mouth)
ecoregion_code	Ocean or sea at the outlet (mouth) according to MEOW ecoregions
longitude	Longitude in decimal degree at the outlet (mouth)
latitude	Latitude in decimal degree at the outlet (mouth)
surface_area_drainage_basin	Surface area of the drainage basin in km^2
length_main_watercourse	Length of the main watercourse in km
altitude_source	Altitude (elevation) of the source in m
bibliography	References used to complete the fields above
administrator	People in charge of adding (or updating) the basin information
link_pictures	Pictures taken during field trips

Table 2
"basin_outlet" table with information related to the GIS representation of catchments. Catchment polygons were extracted from: (1) the Catchment Characterisation Model (CCM2) database [1] (https://ccm.jrc.ec.europa.eu/), and (2) HydroSHEDS [2] (https://www.hydrosheds.org/). Sea outlets and surface areas should be preferentially taken from this table.

Column name	Description of data
basin_id	Unique key associated with each basin
wso_id	CCM code when exists
ccm_name	Catchment name in CCM (if exists)
ccm_area	Surface area, in km^2 (calculated from CCM)
ccm_window	CCM window (used in CCM name files)
geom	Basin shape (from CCM or HydroSHEDS). Multipolygon geometric object
seaoutlet_geom	Basin outlet (from CCM or HydroSHEDS). Point object
simplified_geom	Simplified basin shape to speed up graphical display. Polygon object

Ecoregion was determined according to the Marine Ecoregions of the World [3] (Table 3), which classifies coastal and continental shelf waters. A catchment's ecoregion was defined as the marine region where its outlet was located.

References for each catchment were also recorded in the database under "bibliography", as well as the person in charge of updating the record in the event of future questions ("administrator").

1.3. Species distributions

Presence/absence and population functionality categories were stored in the table "abundance" (Table 4). For each record, the species was designated as either being present (1) or absent (0) for a particular time period (1751–1850, 1851–1950, 1951-present time) in a particular catchment. However, not every species and catchment combination had information available for all three time periods. Time periods associated with a record were designated in three columns within the abundance table. The time period range was divided into two columns that recorded the first ("year_from") and last ("year_to") year of the range. In addition, the time period between "year_from" and "year_to" was also designated in the "period_comment" column as either 1800 (for 1751–1850), 1900 (for 1851–1950), or 2000 (for 1951-present time). References associated with each record were recorded in the "bibliography" column. The columns "abundance_migratory_comment" and "abundance_resident_comment" recorded whether information for the migratory or resident population was updated as part of the 2019–2020 validation, as well as any other comments pertinent to the category designation of population functionality. Many records had additional longer comments associated with the data, and so another table

Table 3
Marine Ecoregions of the World (MEOW) included in the database.

Ecoregion code	Ecoregion name
452	Caspian Sea
20,002	North and East Iceland
20,004	West Greenland Shelf
20,018	North and East Barents Sea
20,019	White Sea
20,020	South and West Iceland
20,022	Southern Norway
20,023	Northern Norway and Finnmark
20,024	Baltic Sea
20,025	North Sea
20,026	Celtic Sea
20,027	South European Atlantic Shelf
20,028	Saharan Upwelling
20,030	Adriatic Sea
20,044	Black Sea
25,031	Aegean Sea
25,032	Levantine Sea
25,034	Ionian Sea
25,035	Western Mediterranean
25,036	Alboran Sea

Table 4
"Abundance" table with distribution information associated with each species, basin, and time period.

Column name	Description of data
abundance_id	Unique key for each record
presence_absence	presence (1)/ absence (0)
period_comment	Lists the time period included between year_from and year_to (1800, 1900, or 2000)
bibliography	References used to complete the fields above and the tables "upstream limit of migration" and "remarks"
basin_id	ID associated with each basin, which connects to information in Table 1 (Basin)
abundance_migratory_form	Population functionality category for migratory population associated with each record (1–4; see Table 6 for description of categories)
species_id	Identifies the species associated with the record (See Table 5)
abundance_resident_form	Population functionality category for resident population associated with each record (1–4; see Table 6 for description of categories)
abundance_migratory_comment	Information associated with migratory abundance and when record was updated
abundance_resident_comment	Information associated with resident abundance
year_from	First year of the time period
year_to	Last year of the time period

named "remarks" was linked to the abundance table using the unique abundance id. These tables can be combined through a query to see all of the comments, notes, and references associated with a particular distribution record.

Four of the columns storing distribution information (Table 4) contained data that were categorized using numeric rather than character designations. These included basin_id, species_id, abundance_migratory_form, and abundance_resident_form. Each of these four columns was connected to a separate table that associated the numeric category with its written description. Basin_id was linked to the "basin" table, which provided the name and other information

Table 5
"Species" table, with unique code associated with each species.

Species id	Scientific name
1	*Acipenser gueldenstaedtii*
2	*Acipenser naccarii*
3	*Acipenser stellatus*
4	*Acipenser sturio*
5	*Alosa algeriensis*
6	*Alosa alosa*
7	*Alosa caspia*
8	*Alosa fallax*
9	*Alosa immaculata*
10	*Alosa kessleri*
11	*Alosa tanaica*
12	*Alosa volgensis*
13	*Anguilla anguilla*
14	*Caspiomyzon wagneri*
15	*Coregonus oxyrinchus*
16	*Coregonus* sp.
17	*Huso huso*
18	*Lampetra fluviatilis*
19	*Chelon ramada*
20	*Osmerus eperlanus*
21	*Petromyzon marinus*
22	*Platichthys flesus*
23	*Salmo caspius*
24	*Salmo labrax*
25	*Salmo salar*
26	*Salmo trutta*
27	*Salvelinus alpinus*
28	*Vimba vimba*

Table 6
"Abundance_form" table, with population functionality categories used to describe both the migratory and resident populations.

Code	Description of data
1	The species was never recorded in the catchment
2	Occasional vagrants were recorded in the catchment (but there was no functional population)
3	Functional populations were present in the catchment
4	Functional populations were present in and numerically dominant in the freshwater community

for each catchment. Species_id was linked to the "species" table, where the scientific name for 28 diadromous species was listed (Table 5). The abundance_migratory_form, and abundance_resident_form columns were linked to the table "abundance_form". There were four population functionality categories used to describe both the migratory and resident populations of a species, where applicable (Table 6). In this context, a "functional" population refers to one that has a high enough abundance to reproduce. Not every record in the distribution table was associated with a population functionality category, as information was not available for certain species within particular catchments.

1.4. Species traits

The table "species_trait" lists information, where available from the primary literature and reports and personal communications of the Diadfish network experts, about the life history, morphology, and phenology of four species (Alosa alosa, Petromyzon marinus, Salmo salar, and

Anguilla). When multiple samples were available for a particular trait, each value was listed using a unique species trait id. Traits were designated using a numeric code (Table 7) that was linked to each species id and basin id (Table 8). The general type of trait (life history, morphological, or phenological) was also listed using a numeric code (Table 9). The table "upstream_limit_migration" lists, for each catchment with available information as defined by basin id, the furthest upstream point that a diadromous population successfully accessed.

2. Experimental Design, Materials and Methods

2.1. Digital libraries and electronic resources

Data collection and updates were produced following a careful literature review. The main digital libraries were searched for references on diadromous species across time, e.g. Gallica (https://gallica.bnf.fr), The digital library of the Real Jardin Botanico of Madrid (https://bibdigital.rjb.csic.es), World Digital Library (https://www.wdl.org), Biodiversity Heritage Library (https://www.biodiversitylibrary.org/), JSTOR (https://www.jstor.org/), Google Book (https://books.google.fr/). Searches were made by crossing a species Latin or vernacular name with a catchment name. The request was made in various languages as species and basin names might change across countries. The same request was made on various platforms until all the time periods for a given catchment-species couple were filled with information on presence/absence to a minimum. Trait data were needed for the most recent period only. As such, Google Scholar was used as the main platform for queries. Main grey references for each of the four species were also inspected for life-history trait data.

2.2. Updates to the dataset spatial and temporal extent

This dataset was enriched through a series of data updates and external validations. The successive versions were used in several published manuscripts and grey literature (Table 10). Three main updates were performed on the dataset between 2005 and 2020 to increase the spatial and temporal extent and improve the accuracy of the data. The first update involved a validation of the original dataset in 2005–2006 (EuroDiad 2.0) and solicited the help of six diadromous species experts across six different countries (Table 11). Data was divided by country, or broader geographical entity (e.g. all the catchments flowing into the Azov sea), and sent to 20 diadromous species experts across Europe; six answered. Participants were asked to validate the presence/absence information and population functionality categories for all species present in EuroDiad for the catchments included in their list and for the two time periods. The second update to EuroDiad 3.2, which added data but did not include an external validation process, occurred in 2009–2010. It was performed with the help of a dedicated staff on this task over four months. The third update to EuroDiad, performed in 2019–2020, added additional catchments and involved a data validation process with 49 participants (Table 12) leading to the last version.

EuroDiad 2.0 included 196 catchments: 35 large catchments (>50 000 km^2), 47 medium-sized catchments (>10 000 km^2) as identified in the same list, and 114 small-sized catchments (<10 000 km^2) that were chosen to avoid any obvious geographical bias. The large catchments together drain two thirds of the continent according to the list of major European river catchments provided by the European Environment Agency (www.eea.europa.eu). In this version of the dataset, species distributions were only recorded for two time periods: 1851–1950 and 1951–2010. The update to EuroDiad 3.2 addressed gaps in spatial coverage, adding diadromous species distributions for 63 small-sized catchments located in Russia, Norway, France, Greece, Spain, Italy, Sicily, Sardinia, and Morocco. Temporal coverage of the dataset was also improved by updating the historic distribution of each species within the 63 newly entered catchments to also

Table 7
Numeric codes to designate different species traits. Trait type is a numeric code provided in Table 9. 1SW stands for 1 sea-winter individuals, 2SW for 2 sea-winter individuals and 3SW for 3 sea-winter individuals.

Code	Description of species trait	Trait type
1	Sex ratio at reproduction 3SW	1
2	Male length at reproduction	1
3	Peak period of upstream migration	3
4	Weight of adult migrant	1
5	Ammocoete length age 4	1
6	Weight of female migrants	1
7	Length at smolting	1
8	Male length at yellow stage	1
9	Male age at silver stage	1
10	Length of 2SW	1
11	Female repeat spawners	1
12	Gonadosomatic index	1
13	Relative residual fecundity	1
14	Eye index for yellow stage	1
15	Length at glass eel stage	1
16	Female age at yellow stage	1
17	Female length at yellow stage	1
18	Female eye index	1
19	Maximum branchiospines	2
20	Male length at silver stage	1
21	Adult migrant length	1
22	Dorsal fin length	2
23	Male repeat spawners	1
24	Ammocoete length age 3	1
25	Length at silver stage	1
26	Weight at silver stage	1
27	Age range for reproductive males	1
28	Weight at metamorphosis	1
29	Age at yellow stage	1
30	Sex ratio at reproduction 2SW	1
31	Female migrant length	1
32	Weight at yellow stage	1
33	Weight of male migrants	1
34	Length of 2+SW	1
35	Adult length at reproduction	1
36	Length of 3SW	1
37	Period of metamorphosis	3
38	Male age at reproduction	1
39	Male weight at silver stage	1
40	Sex ratio at silver stage	1
41	Repeat spawners	1
42	Length at metamorphosis	1
43	Condition factor for adult migrants	1
44	Peak period of upstream migration of glass eels	3
45	Weight at glass eel stage	1
46	Age at silver stage	1
47	Male eye index	1
48	Spawning period	3
49	Period of downstream migration	3
50	Ammocoete length age 5	1
51	Age range for reproductive females	1
52	Female weight at yellow stage	1
53	Female age at reproduction	1
54	Ammocoete length	1
55	Ammocoete length age 2	1
56	Anal fin length	2
57	Ammocoete length age 6	1

(*continued on next page*)

Table 7 (*continued*)

Code	Description of species trait	Trait type
58	Fecundity	1
59	Ammocoete length age 1	1
60	Sex ratio at reproduction	1
61	Sex ratio at reproduction grilse	1
62	Age of smolt at capture	1
63	Average number of branchiospines on the first gill arch	2
64	Male migrant length	1
65	Female length at silver stage	1
66	Reproductive age range	1
67	Female length at reproduction	1
68	Female weight at silver stage	1
69	Male age at yellow stage	1
70	Fecundity of grilse	1
71	Coefficient of condition for glass eels	1
72	Period of upstream migration	3
73	Sea age	1
74	Grilse length	1
75	Length at yellow stage	1
76	Relative fecundity	1
77	Peak period of downstream migration	3
78	Minimum branchiospines	2
79	Adult age at reproduction	1
80	Female age at silver stage	1
81	Male weight at yellow stage	1

Table 8
"Species_trait" table.

Column Name	Description of data
species_trait_id	Unique id associated with each record that lists a numeric value for a species trait
sample	Indication of group membership within a basin; includes group designation and basin name, with the possibility of multiple groups within a basin
discrete_value	Numeric value of the trait when it is an individual discrete number (example: size, weight, age)
interval_value	Description of an interval or set of discrete values (example: time interval, season interval)
period_comment	Description of years associated with the observation period
bibliography	Precise bibliographic reference, with page number
comment	Unit of trait, as well as comments or notes
year_from	First year associated with the study where the trait was recorded
species_id	Code to designate which species is associated with the trait
basin_id	Code to designate which basin is associated with the study for the trait
trait_id	Code to designated the type of trait (linked to Table 9)
year_to	Last year associated with the study where the trait was recorded
interval_lower	Lower limit of the interval; for time periods, designated as a number for the month (ex: mid-December = 12.5)
interval_upper	Upper limit of the interval; for time periods, designated as a number for the month (ex: mid-December = 12.5)

Table 9
General category of species trait type associated with description of traits.

Code	Trait type
1	Life history
2	Morphological
3	Phenological

Table 10
List of key papers and grey literature based on the different EuroDiad versions.

Document type	EuroDiad version	Refs.
Master thesis	EuroDiad 1.0 (original version)	Béguer [4]
Academic paper	EuroDiad 1.0 (original version)	Béguer et al. [5]
PhD thesis	EuroDiad 2.0 and 3.2	Lassalle [6]
Academic paper	EuroDiad 2.0	Lassalle et al. [7]
Academic paper	EuroDiad 2.0	Lassalle and Rochard [8]
Academic paper	EuroDiad 3.2	Lassalle et al. [9]

Table 11
List and affiliations of diadromous species experts who participated in the validation of EuroDiad 2.0 in 2006.

Expert name	Country	Institute
Miran Aprahamian	UK	Environment Agency
Giorgio Bianco	Italy	University of Napoli, Biology Department
Panos Stavros Economidis	Greece	Aristotle University, Zoology Department
Mejdeddine Kraïem	Tunisia	National Institute of Marine Sciences and Technologies (INSTM)
Ion Navodaru	Romania	Danube Delta National Institute for Research and Development
Panu Orell	Finland	Natural Resources Institute Finland (LUKE)

include 1751–1850. This was not done for all catchments within the database, however. For the 196 catchments originally entered into the database, the time period of 1751–1850 was only added in version 3.2 for four emblematic species: *Alosa alosa, Acipenser sturio, Anguilla anguilla and Salmo salar*.

For the 2019–2020 update resulting in the most recent version of this dataset, EuroDiad 4.0, effort was focused on catchments located in the Atlantic Area of Europe. This was primarily for use in several ongoing projects focusing on species distributions in this particular geographic location. The EuroDiad 4.0 update validated the existing data for the Atlantic Area and improved the spatial extent of the dataset by adding several more small-sized catchments, increasing the total number of recorded catchments to 350. While this is a large improvement over EuroDiad 1.0, the precision of available information for each country still varies (Table 13) and the proportion of European catchments included in the dataset could still be improved (Table 14). The 2019–2020 update also improved the temporal extent of the dataset to include all three time periods for all catchments where this information was available.

2.3. *Last external validation process*

Morphological and distribution data associated with catchments in the Atlantic Area were validated in 2019–2020 by DiadES project partners. Data was divided by country, and validation was performed for catchments in Ireland, the UK, Spain, Portugal, and France, by national organisations involved in fisheries and environmental management (Table 12). In the U.K., local experts were available to validate data in all regions except Scotland.

For the validation process, species experts were identified regionally using partners for the DiadES project and their local affiliations. Participants were asked to validate morphological information for the catchment (location of the outlet, surface area of the drainage basin, length of the main watercourse, elevation at the headwaters), as well as the presence/absence information and population functionality categories for all species already present in EuroDiad for their country. If possible, verification was done for each of the three time periods. Participants were also asked to provide data for any other catchments for which they had access to information on diadromous species population status.

Several questions were raised and addressed over the course of the validation process to ensure that each participant had the same interpretation of data categories. This was especially

Table 12
List and affiliations of diadromous species experts who participated in the validation of EuroDiad 4.0 in 2019–2020.

Validator	Affiliation
DiadES Project Partners	
Jimmy King (retired)	IFI
William Roche	IFI
Tea Bašić	Cefas
Gordon Copp	Cefas
Estibaliz Diaz	AZTI
Iker Azpiroz	EKOLUR
Aitor Lecuona	Diputación Gipuzkoa
Guillem Chust	AZTI
David José Nachón García	USC/EHEC
Fernando Cobo Gradin	USC/EHEC
Rufino Vieira Lanero	USC/EHEC
Carlos Antunes	CMVNC
Catarina Mateus	UE/MARE
Ana Filipa Belo	UE/MARE
Pedro R. Almeida	UE/MARE
Carlos Alexandre	UE/MARE
Esmeralda Pereira	UE/MARE
Local Experts	
Robert Rosell	AFBI
Trevor Harrison	DAERA
Michael McNeill	DAERA
Richard Kennedy	AFBI
Andy Gowans	EA
Charles Crundwell	EA
Darryl Clifton-Dey	EA
John Foster	EA
Rob Hillman	EA
Nicole Bryson	EA
Jody Armitage	EA
Philip Rudd	EA
Justin Mould	EA
Emma Woods	EA
Paul Hyatt	NRW
Matt Buck	EA
Paul Greest	NRW
Richard Cove	NRW
Alberto Aguirre	ANBIOTEK (Basque Country)
Ainhize Uriarte	AZTI (Basque Country)
Ignacio Ferrando	Vaersa (Valencia)
Francesc Jesús Gómez	Government of Catalonia
Jerónimo de la Hoz	Government of Asturias
Marc Ordeix	CERM-UVic-UCC (Catalonia)
José Peñalver;	Government of Murcia
Josu Elso	Government of Navarra
Carlos Fernández Delgado	University of Córdoba
Francisco J. Oliva Paterna	University of Murcia
Maria Mar Torralva	University of Murcia
Belén Muñoz	Ministerio para la Transición Ecológica y el Reto Demográfico
Gema Campillo	Government of Cantabria
Pablo Caballero	Xunta de Galicia

true for the 2019–2020 update with a large number of participants. The first question related to handling uncertainty in the accuracy of presence or absence data, with a focus on the consequences of false presences or false absences in the output of habitat suitability modelling. The decision was made to leave the presence/absence records in the abundance table blank (NULL) when there was reasonable doubt about the reliability of presence/absence data. This was done

Table 13
"Countries" with information in EuroDiad 4.0, with the total number of catchments included in the "basin" table and the total number of records for which presence or absence (P/A) information for a species in a catchment for a time period is available in the "abundance" table, divided by country. The British Isles were considered as Ireland, England, Scotland and Wales.

Country	Number of catchments	Total number of P/A records
Albania	2	120
Algeria	1	60
Azerbaijan	2	120
Azerbaijan/Georgia	1	60
Bulgaria	1	58
Croatia	1	60
Denmark	4	116
Egypt	1	60
England	18	778
Finland	2	120
France	60	3390
Georgia	4	231
Germany	6	302
Greece	22	1644
Greenland	1	60
Iceland	13	240
Iran	6	359
Ireland	11	420
Italy	23	1594
Kazakhstan	1	60
Latvia	1	60
Lithuania	1	60
Morocco	5	347
Netherlands	2	0
Norway	33	677
Poland	1	60
Portugal	12	386
Romania	1	60
Russia	21	1358
Scotland	10	240
Spain	48	2016
Sweden	7	416
Tunisia	2	120
Turkey	18	1062
Ukraine	4	238
Wales	4	11

Table 14
Comparison of EuroDiad 4.0 catchments with distribution information to the Catchment Characterisation Model (CCM2) database [1] (https://ccm.jrc.ec.europa.eu/). EuroDiad catchments outside of the CCM coverage were excluded for the present analysis.

Drainage area (km^2)	Number of catchments in the CCM	Number of catchments in EuroDiad 4.0	Cumulative drainage area in the CCM (km^2)	Cumulative drainage area in EuroDiad (km^2)
< 10	8703		36 654.60	
< 100	8534	6	259 617.1	417.8
< 1000	2061	62	638 347.2	33 013.6
< 5000	423	107	881 577.9	246 588.9
< 10,000	56	29	373 723.6	194 537.3
< 50,000	74	38	1 508 510.4	811 291.8
>= 50,000	45	37	8 205 620.2	6 635 297.5

Table 15
Change in categorical designation of population functionality between EuroDiad 2.0 and 4.0. Note that the definition is the same regardless of character or numeric categories. These definitions are stored in the "abundance_form" table under "abundance_interpretation" (Fig. 1, Table 6).

Character designation (EuroDiad 2.0)	Numeric designation (EuroDiad 4.0)	Definition
Missing	1	The species was never recorded in the catchment
Rare	2	Occasional vagrants were recorded in the catchment (but there was no functional population)
Common	3	Functional populations were present in the catchment
Abundant	4	Functional populations were present in and numerically dominant in the freshwater community

instead of deleting the record so that uncertainty about a species within a particular catchment could be recorded with comments and references in the event of future discussions or updates. This approach also ensured that these records with no reliable information could be excluded from queries.

The second question involved the temporal extend of the dataset. For some species and catchments, there was a mismatch between the rate of population change and this 100-year time periods, meaning that there was a change in population functionality category within a time period instead of between time periods. In these instances, the condition at the end of the time period was used to define the population status. For example, if a species was present in 1970 but absent by 2018, that species was recorded as being absent for the whole time period 1951-present time. However, in these instances, a comment was included in the abundance_migratory_comment column of the "abundance" table stating when the change in abundance occurred (if known). In addition, several questions related to particular catchments were addressed on a case-by-case basis. This included merging or separating catchments, or in rare cases changing the name of the catchment in the "basin" table. When these types of changes were made, it was recorded in the administrator column of the "basin" table, along with the name of the person updating the records.

The last question involved the four population functionality categories. These categories were designated during the development of the original versions of EuroDiad as "Missing", "Rare", "Common", and "Abundant". However, as these descriptive words can have different definitions depending on the context and a researcher's background, it was decided for the 2019–2020 dataset update to use a numeric code instead of a character code (Table 15) for these four categories in order to alleviate any confusion. The definitions of these categories was not changed between EuroDiad 2.0 and EuroDiad 4.0, only the way they were presented in the data tables within the dataset. The focus of these categories was on functional versus non-functional populations, as this is an important designation when using habitat suitability to predict future fish distributions. Less focus was put on the actual size of the populations, and numeric abundances were not recorded for any species or catchments, regardless of the availability of that data.

Ethics Statement

The manuscript adheres to Ethics in publishing standards. Experts provided their written informed consents to have their names publicly cited in the paper in Tables 11 and 12.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Betsy Barber-O'Malley: Data curation, Validation, Methodology, Writing – original draft; **Géraldine Lassalle:** Conceptualization, Methodology, Writing – original draft; **Patrick Lambert:** Methodology, Software, Supervision, Writing – review & editing; **Eric Quinton:** Methodology, Software, Writing – review & editing.

Acknowledgments

The second database validation was funded by the Atlantic Area Interreg Project DiadES (www.diades.eu). Many thanks go to all the experts that help with the validation of the EuroDiad 4.0 database (see https://diades.eu/work-packages/ and Table 12).

References

[1] J. Vogt, P. Soille, A. De Jager, E. Rimaviciute, W. Mehl, S. Foisneau, K. Bodis, J. Dusart, M. Paracchini, P. Haastrup, C. Bamps, A Pan-European River and Catchment Database, EUR 22920 EN, Luxembourg, Luxembourg, 2007, doi:10.2788/35907.

[2] B. Lehner, G. Grill, Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrol. Process. 27 (2013) 2171–2186, doi:10.1002/hyp.9740.

[3] M.D. Spalding, H.E. Fox, G.R. Allen, N. Davidson, Z.A. Ferdaña, M. Finlayson, B.S. Halpern, M.A. Jorge, A. Lombana, S.A. Lourie, K.D. Martin, E. McManus, J. Molnar, C.A. Recchia, J. Robertson, Marine ecoregions of the world: a bioregionalization of coastal and shelf areas, Bioscience 57 (2007) 573–583, doi:10.1641/B570707.

[4] M. Béguer, Analyse de l'importance relative du régime thermique et des caractéristiques des bassins versants dans la distribution des assemblages de poissons migrateurs amphihalins de l'Europe de l'Ouest, Mémoire de Master 2, Université de Bordeaux, Bordeaux, France, 2005.

[5] M. Béguer, L. Beaulaton, E. Rochard, Distribution and richness of diadromous fish assemblages in Western Europe: large scale explanatory factors, Ecol. Freshw. Fish. 16 (2007) 221–237, doi:10.1111/j.1600-0633.2006.00214.x.

[6] G. Lassalle, Impacts des changements globaux sur la distribution des poissons migrateurs amphihalins - Une approche par modélisation à l'échelle continentale, Université Bordeaux I, Bordeaux, France, 2008 Thèse de doctorat.

[7] G. Lassalle, M. Béguer, L. Beaulaton, E. Rochard, Diadromous fish conservation plans need to consider global warming issues: an approach using biogeographical models, Biol. Conserv. 141 (2008) 1105–1118, doi:10.1016/j.biocon.2008.02.010.

[8] G. Lassalle, E. Rochard, Impact of twenty-first century climate change on diadromous fish spread over Europe, North Africa and the Middle East, Glob. Change Biol. 15 (2009) 1072–1089, doi:10.1111/j.1365-2486.2008.01794.x.

[9] G. Lassalle, P. Crouzet, J. Gessner, E. Rochard, Global warming impacts and conservation responses for the critically endangered European sturgeon, Biol. Conserv. 143 (2010) 2441–2452, doi:10.1016/j.biocon.2010.06.008.