Assessing and enhancing ecosystem services provided by diadromous fish in a climate change context

Deliverable title: EuroDiad 4.0 User's Guide
Deliverable reference: WP 5 – Action 1

Contributors and affiliations:

Betsy BARBER – INRAE
Géraldine LASSALLE – INRAE
Patrick LAMBERT – INRAE
Eric QUINTON -INRAE
and all PARTNERS listed in Annex 1

Subsidy contract n°	EAPA_18/2018
Programme priority	4. Enhancing biodiversity and the natural and cultural assets
Start date of project	1st February 2019
End date of the project	31st January 2022
Work package n°	5. Development of new species distribution models
Indicator	Number of technical and scientific publications produced: 1
WP Leader	
Partners involved	INRAE – partner n°1 AZTI – partner n°2 EHEC-USC – partner n°3 MARE-UÉ – partner n°5 CMVNC – partner N°6 IFI – partner n°7 Cefas – partner n°9 MNHN – partner n°10
Submission date (month)	April 2020

Dissemination level		
PU	Public	X
PP	Restricted to other programme participants (including the Programme Authorities)	
RE	Restricted to a group specified by the consortium (including the Programme Authorities)	
CO	Confidential, only for members of the consortium (including the Programme Authorities)	

Table of contents

1. Executive summary in the four languages of the programme

EuroDiad is a set of data tables that store information about the presence/absence and functionality (or reproductive capacity) 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 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. The most recent version, EuroDiad 4.0, is available in PostgreSQL and underwent a second validation process in 2019-2020. Data for version 4.0 were validated by DiadES project partners and local experts. EuroDiad 4.0 now stores data for 350 catchments (of which 292 have abundance records) and three time periods, though the precision of information varies and not every species has information for each time period.

The primary goal of the 2019-2020 update was to provide information 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 project partners, b) validate existing information about the presence or absence of diadromous species and categorize their population functionality within a catchment, and c) perform data hygiene to prepare the database for broad dissemination. This update was focused on catchments located in the Atlantic Area for use in DiadES projects. Data were divided by country, and validation was performed for catchments in Ireland (IFI), the UK (Cefas), Spain (AZTI and USC-EHEC), Portugal (CMVNC and MARE-UÉ), and France (INRAE and MNHN). 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 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. Partners were also asked to provide data for any other catchments for which they had access to information on fish population status. This validation process strengthened the usefulness of EuroDiad, which is now updated and available for use by project partners. The database is available on request to betsy.barber@inrae.fr, patrick.mh.lambert@inrae.fr and geraldine.lassalle@inrae.fr.

Résumé exécutif dans les quatre langues du Programme

EuroDiad est constituée d'un ensemble de tables qui stockent des informations sur la présence/absence et la fonctionnalité (ou capacité reproductrice) de populations d'espèces migratrices amphihalines (lamproies et poissons) dans des bassins versants sélectionnés en Europe, au Moyen-Orient et en Afrique du Nord de 1750 à nos jours. Cette base de données contient des informations sur la distribution et les traits d'histoire de vie de 28 espèces migratrices amphihalines et des données géomorphologiques pour les bassins versants sélectionnés, néanmoins toutes les espèces n'ont pas des données saisies pour tous les bassins versants et toutes les périodes. EuroDiad a été créée en 2005-2006 (EuroDiad 1.0 et 2.0), et contenait des données pour 196 bassins versants et deux périodes temporelles (1851-1950 et 1951-2010). Elle a subi une mise à jour majeure en 2009-2010 (EuroDiad 3.2) suite à une procédure de validation par des experts européens en halieutique. La version 3.2 a été enrichie de 63 « petits » bassins versants (<10,000 km²) et d'une période temporelle supplémentaire (1750-1850) pour les espèces et bassins versants sélectionnés. La version la plus récente, EuroDiad 4.0, est disponible sous PostgreSQL et a subi une seconde procédure de validation en 2019-2020. Les données de la version 4.0 ont été validées par les partenaires du projet DiadES et des experts locaux. EuroDiad 4.0 contient désormais 350 bassins versants (dont 292 ont des enregistrements relatifs à la distribution des espèces) et trois

périodes temporelles, néanmoins la précision des informations varie et toutes les espèces n'ont pas des informations pour chaque période temporelle.

Le but premier de la mise à jour de 2019-2020 était de fournir des informations pour une nouvelle génération de modèles de distribution d'espèces, appelés modèles hybrides, qui prennent en compte la favorabilité d'habitats et la dynamique de population dans un même cadre. Les objectifs secondaires de cette mise à jour étaient : a) d'inclure de nouveaux bassins versants pour lesquels les informations ont été fournies par les partenaires, b) de valider les informations existantes sur la présence ou absence des espèces migratrices amphihalines et catégoriser la fonctionnalité de leur population dans un bassin versant, et c) de réaliser un nettoyage approfondi de la base de données avant sa diffusion plus large. Cette mise à jour a été centrée sur les bassins versants localisés dans l'Espace Atlantique pour utilisation dans le projet DiadES. Les données ont été validées par pays, et la validation a été faite à l'échelle du bassin versant en Irlande (IFI), au Royaume-Uni (Cefas), en Espagne (AZTI et USC-EHEC), au Portugal (CMVNC et MARE-UÉ), et en France (INRAE et MNHN). Les partenaires du projet DiadES ont dû valider les données géomorphologiques pour les bassins versants (localisation de l'embouchure, surface du bassin versant, longueur du cours d'eau principal et altitude de la source), de même que les données de présence/absence et les catégories de fonctionnalité de population pour toutes les espèces présentes dans EuroDiad pour le pays en question. Si possible, la vérification a été faite pour chacune des trois périodes temporelles. Les partenaires ont dû aussi fournir des données pour d'autres bassins versants pour lesquels ils avaient des informations sur le statut des populations de poissons. La procédure de validation a renforcé l'utilité d'EuroDiad qui est maintenant mise à jour et disponible pour utilisation par les partenaires du projet. La base de données est disponible sur demande auprès de betsy.barber@inrae.fr, patrick.mh.lambert@inrae.fr et geraldine.lassalle@inrae.fr.

Sumário Executivo nas Quatro Línguas do Programa

O EuroDiad é uma base de dados que armazena informação sobre a presença/ausência e funcionalidade (ou capacidade reprodutiva) de populações de espécies diádromas (lampreias e peixes), em bacias hidrográficas selecionadas na Europa, Médio Oriente e Norte de África, desde 1750 até ao presente. Esta base de dados contém informação sobre a distribuição e ciclos de vida de vinte e oito espécies diádromas e dados geomorfológicos para cada uma das bacias hidrográficas selecionadas, apesar de nem todas as espécies terem dados para todas as bacias hidrográficas e intervalo temporal. O EuroDiad foi originalmente criado em 2005-2006 (EuroDiad 1.0 e 2.0) e continha dados de 196 bacias hidrográficas e dois períodos temporais (1851-1950 e 1951-2010). Foi submetido a uma grande atualização em 2009-2010 (EuroDiad 3.2) através de um processo de validação por especialistas europeus em recursos pesqueiros. A versão 3.2 incluiu a adição de 63 bacias hidrográficas de pequena dimensão (<10.000 km^2) e um período adicional (1751-1850) para determinadas espécies e bacias hidrográficas. A versão mais recente, EuroDiad 4.0, está disponível no PostgreSQL e passou por um segundo processo de validação em 2019-2020. Os dados que integram a versão 4.0 foram validados pelos parceiros do projeto DiadES e especialistas locais. O EuroDiad 4.0 inclui agora dados de 350 bacias hidrográficas (das quais 292 possuem registos de abundância) e três intervalos temporais, embora a precisão da informação varie e nem todas as espécies possuam informação para cada período.

O objetivo principal da atualização de 2019-2020 consistiu em fornecer informação para uma nova geração de modelos de distribuição de espécies, denominados modelos híbridos, e que incorporam na sua estrutura a adequabilidade do habitat e a dinâmica populacional. Os objetivos secundários desta atualização foram: a) incorporar novas bacias hidrográficas para as quais foi fornecida informação pelos parceiros do projeto, b) validar a informação existente sobre presença ou ausência de espécies diádromas e categorizar a sua funcionalidade populacional na bacia hidrográfica; e c) executar a limpeza e preparação da base de dados para disseminação alargada. Esta atualização foi dirigida às bacias hidrográficas localizadas na Área Atlântica para uso em projetos DiadES. Os dados foram divididos por país e a validação foi realizada para bacias hidrográficas

da Irlanda (IFI), Reino Unido (CEFAS), Espanha (AZTI e USC-EHEC), Portugal (CMVNC e MARE-UÉ) e França (INRAE e MNHN). Solicitou-se aos parceiros do DiadES que validassem as informações geomorfológicas da bacia hidrográfica (localização da Foz, área da bacia de drenagem, comprimento do curso de água principal, elevação nas cabeceiras), bem como as categorias de informação de presença / ausência e funcionalidade populacional para todas as espécies já presentes no EuroDiad para o seu país. Sempre que possível, a verificação foi feita para cada um dos três períodos . Foi também solicitado que os parceiros fornecessem dados de outras bacias para as quais tivessem acesso a informação sobre o estado populacional de espécies piscícolas. Esse processo de validação reforçou a utilidade do EuroDiad, que agora está atualizado e disponível para utilização pelos parceiros do projeto. A base de dados está disponível mediante solicitação a betsy.barber@inrae.fr, patrick.mh.lambert@inrae.fr e geraldine.lassalle@inrae.fr.

Resumen ejecutivo en los cuatro idiomas del Programa

EuroDiad es un conjunto de tablas de datos que almacenan información sobre la presencia/ausencia y la funcionalidad (o capacidad reproductiva) de las poblaciones de las especies diádromas (lampreas y peces) en determinadas cuencas de Europa, de Oriente Medio y del norte de África desde 1750 hasta la actualidad. Esta base de datos contiene información sobre la distribución y los rasgos del ciclo vital de 28 especies diádromas y datos geomorfológicos de cada una de las cuencas fluviales seleccionadas, si bien no siempre existe información de cada cuenca y período de tiempo para cada especie. EuroDiad fue creada originalmente en 2005-2006 (EuroDiad 1.0 y 2.0), y contenía datos para 196 cuencas y dos períodos de tiempo (1851-1950 y 1951-2010). Posteriormente, fue objeto de una importante actualización en 2009-2010 (EuroDiad 3.2) mediante un proceso de validación por parte de expertos europeas en materia de pesquerías. La versión 3.2 incluía la adición de 63 cuencas de pequeño tamaño (< 10.000 km2) y un período de tiempo adicional (1751-1850) para determinadas especies y cuencas. La versión más reciente, EuroDiad 4.0, está disponible en PostgreSQL y fue sometida a un segundo proceso de validación en 2019-2020. Los datos de la versión 4.0 fueron validados tanto por los socios del proyecto DiadES cómo por expertos locales. EuroDiad 4.0 almacena actualmente datos para 350 cuencas (de las cuales 292 tienen registros de abundancia para las poblaciones de peces) y tres períodos de tiempo, aunque la precisión de la información varía y no todas las especies cuentan con información para cada período de tiempo.

El objetivo principal de la actualización llevada a cabo en 2019-2020 era proporcionar información para una nueva generación de modelos de distribución de especies, denominados modelos híbridos, que incorporan tanto la idoneidad del hábitat como la dinámica de la población en su marco de trabajo. Los objetivos secundarios de esta actualización eran: a) incorporar nuevas cuencas fluviales para las que los socios del proyecto proporcionaron información, b) validar la información existente sobre la presencia o ausencia de especies diádromas y categorizar su funcionalidad poblacional dentro de una cuenca fluvial, y c) realizar una depuración o limpieza de los datos a fin de preparar la base de datos para una amplia difusión. Esta actualización se centró en las cuencas fluviales situadas en el Espacio Atlántico para su uso en el proyecto DiadES. Los datos se dividieron por país, y se realizó la validación para las cuencas fluviales de Irlanda (IFI), el Reino Unido (Cefas), España (AZTI y USC-EHEC), Portugal (CMVNC y MARE-UÉ) y Francia (INRAE y MNHN). Se pidió a los socios del proyecto DiadES que validaran la información geomorfológica de la cuenca (ubicación de la desembocadura, superficie de la cuenca de drenaje, longitud del curso de agua principal, elevación en la cabecera), así como la información sobre presencia/ausencia y las categorías de funcionalidad de la población para todas las especies ya presentes en EuroDiad para su país. De ser posible, se hizo una verificación para cada uno de los tres períodos de tiempo. También se pidió a los socios que proporcionaran información o datos sobre cualquier otra cuenca fluvial para la que tuvieran acceso a información sobre la situación de la población de peces. Este proceso de validación reforzó la utilidad de EuroDiad, que ahora está actualizada y disponible para su uso por parte de los socios del proyecto. La base de datos puede ser solicitada a betsy.barber@inrae.fr, patrick.mh.lambert@inrae.fr y geraldine.lassalle@inrae.fr.

2. Introduction

This document is a user's guide to the EuroDiad 4.0 database, which was developed to provide information about the distribution of diadromous species (i.e. jawless fish and fish) found in continental waters in Europe, the Middle East, and North Africa from 1750 to present time. Information was gathered for 28 diadromous species (i.e. anadromous and catadromous) in this geographical area. Data for version 4.0 were validated by DiadES project partners. The primary goal of this update was to provide information for a new generation of species distribution models, referred to as hybrid models, which incorporate both habitat suitability and population dynamics within their framework. The objectives of this update were to: a) incorporate new catchments for which information was provided by project partners, b) validate existing information about the presence or absence of diadromous species and categorize their abundance or population functionality within a catchment, and c) perform data hygiene to prepare the database for broad dissemination.

3. Brief description of EuroDiad 4.0

EuroDiad is a set of data tables that store information about the presence/absence and functionality of populations in selected catchments in Europe, the Middle East, and North Africa. Here, functionality refers to the reproductive capacity of a population, which is discussed in more detail below. This data was originally 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) (Table 1). 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 (Figure 1 and 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 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. The data used in EuroDiad has been through multiple validation processes, once in 2006 by experts in the field of fish biology and ecology (Table 2), and most recently in 2019 by project partners and experts as part of WP 5 in the project DiadES (Annex 1).

Table 1. List of key papers and grey literature based on the different EuroDiad versions.

Document type	EuroDiad version	Reference
Master thesis	EuroDiad 1.0	Béguer, 2005
Academic paper	EuroDiad 1.0	Béguer et al., 2007
PhD thesis	EuroDiad 2.0 and EuroDiad 3.2	Lassalle, 2008
Academic paper	EuroDiad 2.0	Lassalle et al., 2008
Academic paper	EuroDiad 2.0	Lassalle and Rochard, 2009
Academic paper	EuroDiad 3.2	Lassalle et al., 2010

7

Table 2. List of experts asked to validate 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
Ion Navodaru	Romania	Danube Delta National Institute for Research and Development
Panu Orell	Finland	Finnish Game and Fisheries Research Institute
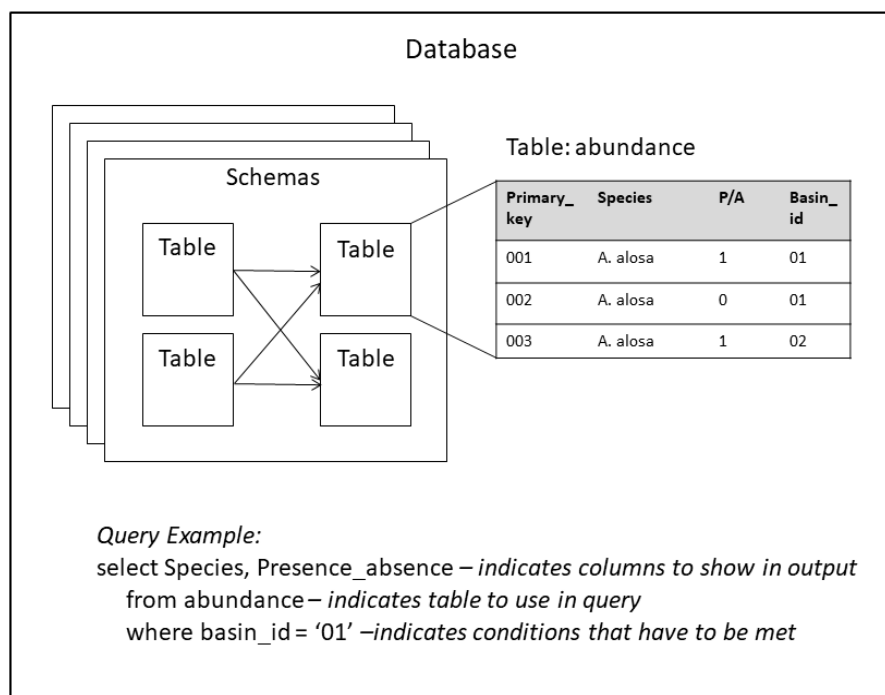
Figure 1. General setup of database in PostgreSQL. A database can contain multiple schemas, each with multiple tables and connections. Data in tables are organized into columns of variables and rows of unique data records. Queries can access specific information in tables within a schema, or combine information from multiple tables.

3.1. EuroDiad 4.0 coverage

3.1.1. Spatial and temporal extent

After the first validation process in 2006, EuroDiad 2.0 included 196 catchments: 35 large catchments (> 50 000 km²) that 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), 47 medium-sized catchments (> 10 000 km²) as identified in the same list, and 114 small-sized catchments (< 10 000 km²) that were chosen to avoid any obvious geographical bias. EuroDiad 3.2 was developed in 2009-2010 to incorporate two major improvements in the spatial and temporal coverage of the database. To address areas that were identified as gaps in spatial coverage, diadromous species distributions were added for 63 small-sized catchments located in Russia, Norway, France, Greece, Spain, Italy, Sicily, Sardinia, and Morocco. Species distributions in the previous version of EuroDiad (2.0) were only recorded for two time periods: 1851-1950 and 1951-2010. Temporal coverage of the database was improved by updating the historic distribution of each species within the 63 newly entered catchments to also include 1751-1850. This was not done for all catchments within the database. For the 196 catchments originally entered into the database, the time period of 1751-1850 was only updated for the distribution of four emblematic species: *Alosa alosa*, *Acipenser sturio*, *Anguilla anguilla* and *Salmo salar*.

For the 2019-2020 update resulting in EuroDiad 4.0, effort was focused on catchments located in the Atlantic Area for use in DiadES projects. As a consequence, several small-sized catchments were added because they were part of a case study (WP 6), or because information was provided through the validation process (described below). EuroDiad 4.0 now stores data for 350 catchments, though the precision of information varies (Table 3 and Table 4). Distribution information (i.e. presence/absence and population functionality categories) was divided into three time periods (1751-1850, 1851-1950, and 1951-2010), though not every species had information for each of these time periods as explained above.

Table 3. "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 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 4. Comparison of EuroDiad 4.0 catchments with distribution information to the Catchment Characterisation Model (CCM2) database (https://ccm.jrc.ec.europa.eu/). EuroDiad catchments outside of the CCM coverage were excluded here for the analysis.

Drainage area (km²)	Number of catchments in the CCM	Number of catchments in EuroDiad 4.0	Cumulative drainage area in the CCM (km²)	Cumulative drainage area in EuroDiad (km²)
<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
< 10000	56	29	373 723.6	194 537.3
< 50000	74	38	1 508 510.4	811 291.8
>=50000	45	37	8 205 620.2	6 635 297.5

3.2. Data tables

Data was divided into multiple tables (Figure 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, variables in the columns of these tables were recorded as a numeric index rather than as a character vector. This numeric index was associated with an explicit wording in a separate table. As explained above, when combining these tables, it is possible to permanently join multiple tables into a view. In a view, the numeric index can be replaced with its explicit description so that the user does not have to refer back to the associated table. Two views have been created in EuroDiad 4.0, one to highlight distribution information and one to highlight species life history trait information (Figure 2).
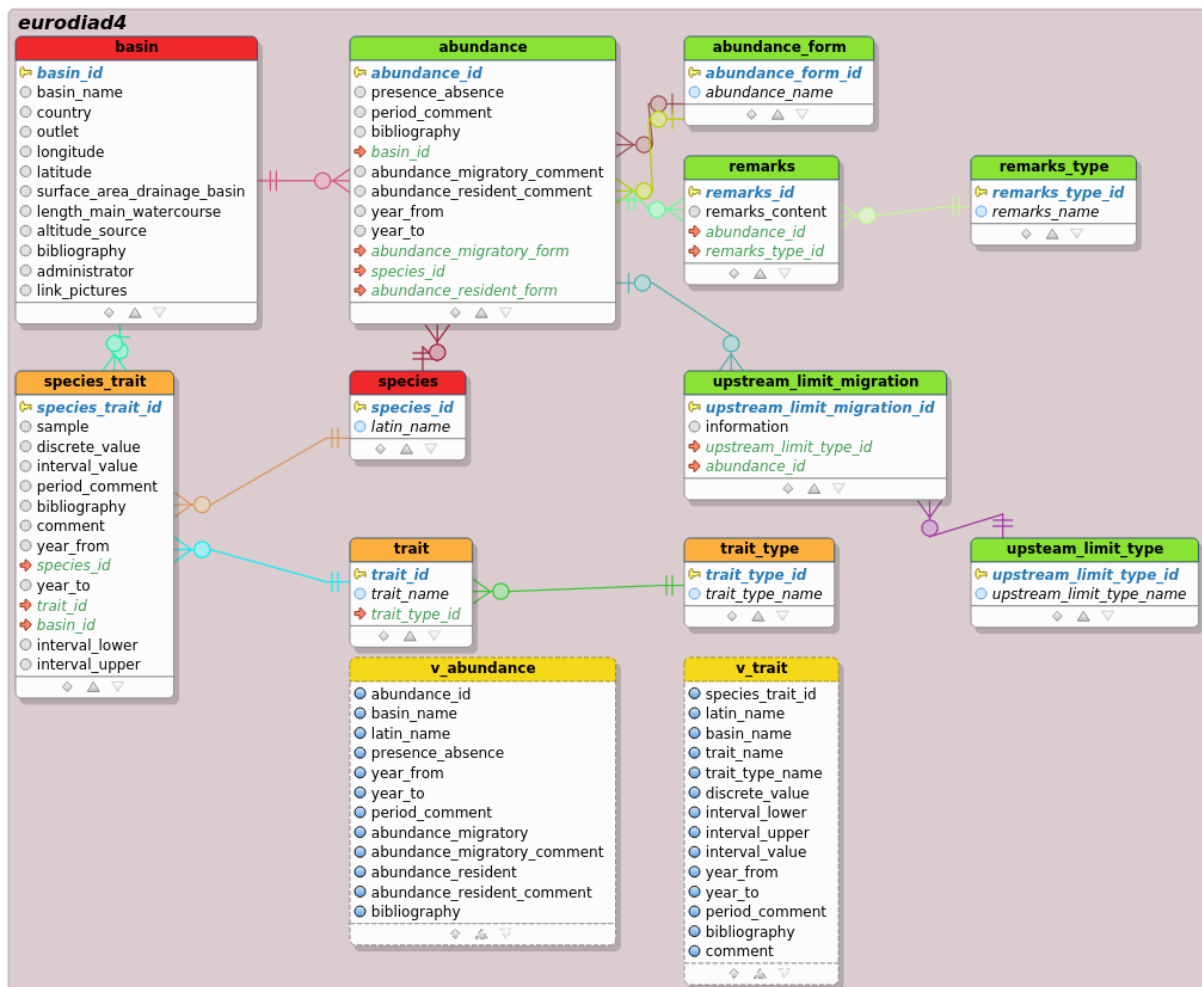
Figure 2. EuroDiad 4.0 relational schema. Arrows indicate connections/relationships between tables. Views are represented with yellow, and are designated with a "v" before the table name.

3.2.1. Catchment information

Information associated with each catchment was stored in the table "basin" (Table 5). Values for the surface area of the drainage basin, length of the main watercourse, and altitude of the source were compiled using primary literature, specialized websites, and expert validation. Data for all catchments in the Atlantic Area were considered for validation by DiadES experts as described below in section 3. For catchments outside of the Atlantic Area, data were not updated in 2019 and were directly used from the 2009-2010 validation. Ecoregion was determined using the Marine Ecoregions of the World shapefile (Spalding et al. (2007); https://www.arcgis.com/home/item.html?id=b0ca60c9472a432f9d659b86864f3764; Table 6), which classifies coastal and continental shelf waters. A catchment's ecoregion was defined as the marine region where its outlet was located. For the majority of catchments, the longitude and latitude of the outlet was defined as the point that aligned with the coastline, rather than further upstream at the head of tide. 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 in accordance with CCM designation. 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").

Table 5. 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 km²
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 6. Marine Ecoregions of the World (MEOW) included in database.

Ecoregion code	Ecoregion name
452	Caspian Sea
20002	North and East Iceland
20004	West Greenland Shelf
20018	North and East Barents Sea
20019	White Sea
20020	South and West Iceland
20022	Southern Norway
20023	Northern Norway and Finnmark
20024	Baltic Sea
20025	North Sea
20026	Celtic Sea
20027	South European Atlantic Shelf
20028	Saharan Upwelling
20030	Adriatic Sea
20044	Black Sea
25031	Aegean Sea
25032	Levantine Sea
25034	Ionian Sea
25035	Western Mediterranean
25036	Alboran Sea

3.2.2. Distribution information

Presence/absence and population functionality categories were stored in the table "abundance" (Table 7). For each record, the species was designated as either being present (1) or absent (0) for a particular time period (1751-1850, 1851-1950, 1951-2010) 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 1951-2010). 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 most recent validation, as well as any other comments pertinent to the designation of population functionality category. Many records had additional longer comments associated with the data, and so another table 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.

Table 7. 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 5 (Basin)
abundance_migratory_form	Population functionality category for migratory population associated with each record (1-4; see Table 9 for description of categories)
species_id	Identifies the species associated with the record (See Table 8)
abundance_resident_form	Population functionality category for resident population associated with each record (1-4; see Table 9 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

Four of the columns in the distribution information table contained data that were categorized using numbers instead of words. These included basin_id, species_id, abundance_migratory_form, and abundance_resident_form. Each of these four columns had their own table that associated the numeric category with its description. Basin_id was linked to the "basin" table, which provided the name and other information for each catchment. Species_id was linked to the "species" table, where the scientific name for 28 diadromous species was listed (Table 8). The migratory_form colums 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 9). 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. These categories (and some of the issues in defining them), are discussed in more detail in Section 3.

Table 8. 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	*Liza 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 9. Population functionality categories used to describe both the migratory and resident populations.

Code	Description of data
1	The species was never recorded in the basin
2	Occasional vagrants were recorded in the basin
3	Functional populations were present in the basin
4	Functional populations were present and numerically dominant in the freshwater community

15

3.2.3. Other relevant tables

The table "species_trait" lists information, where available from the primary literature, about the life history, morphology, and phenology of four species (*Alosa alosa*, *Petromyzon marinus*, *Salmo salar*, and *Anguilla Anguilla*). When multiple samples were available for a particular trait, each value was listed using a unique species trait id. Trait type was designated using a numeric code (Table 10) that was linked to each species id and basin id (Table 11). The general category of trait type (life history, morphological, or phenological) was also listed using a numeric code (Table 12). The table "upstream_limit_migration" lists for each catchment the furthest upstream point that a diadromous population could successfully access, given the presence of barriers to passage.

Table 10. Numeric codes to designate different types of species traits. Trait type is numeric code linked to Table 11.

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

16

Table 10 Continued

Code	Description of species trait	Trait type
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
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 11. 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 10)
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 12. Trait type associated with description of traits.

Code	Trait type
1	Life history
2	Morphological
3	Phenological

4. 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 (IFI), the UK (Cefas), Spain (AZTI and USC-EHEC), Portugal (CMVNC and MARE-UÉ), and France (INRAE and MNHN). In many cases, local experts aided in the validation process, including staff from AFBI, EA, and NRW (see Annex 1 for the full list of partners and experts involved in the validation process). In the UK, local experts were available to validate data in all regions except Scotland.

4.1. Steps in validation

Partners 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. Partners were also asked to provide data for any other catchments for which they had access to information on fish population status.

4.2. Discussions related to validation process

Several questions were raised and addressed over the course of the validation process. Many of these discussions were scenario-specific and so consensus was reached between partners involved and did not include the entire group. There were, however, several major discussion points that were addressed during the WP6 meeting in Norwich (18th to 20th of November 2019). Some of these discussion points are outlined briefly below.

4.2.1. Uncertainty and false presences

The question of how to handle uncertainty in the accuracy of presence or absence data was brought up, which led to discussions about the consequences of false presences or false absences in the output of habitat suitability modelling. This problem was initially put forward in relation to European shads on the Cantabrian coast of northern Spain where a lack of reliable information was detected. From these discussions, the decision was made to leave the presence/absence records in the abundance table blank when there was reasonable doubt about the reliability of presence/absence data. This was done 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 easily excluded from queries.

4.2.2. Designation of population functionality categories

Population functionality categories (Table 9) were designated during the development of the original versions of EuroDiad, and in order to keep the data consistent these definitions were not changed. These categories included:

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

When the validation process started, these four categories were associated with descriptive words (Missing, Rare, Common, and Abundant, respectively). However, as these descriptive words can have different definitions depending on the context and a researcher's background, it was decided to instead use a code for the four population functionality categories in order to alleviate any confusion. 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.

4.2.3. Changes in species presence within a time period category

For many species, there was a mismatch between the rate of population change and the time periods used in EuroDiad 4.0. This meant that for some species, a change in population functionality category was seen within a time period instead of between time periods. In these instances, the condition at the end of the time period should be used to define the population status. For example, if a species was present in 1970 but absent by 2010, that species was recorded as being absent for the whole time period. 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.

4.2.4. Decisions for individual catchments

During the course of updating the database, several problems came up related to particular catchments that had to be 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.

5. Conclusion

EuroDiad 4.0 provides morphological catchment information and diadromous species distribution data that could be used for many population modelling projects. Validation by local experts in 2019-2020 strengthened the usefulness of this data. This schema is now updated and available for use by project partners on request to Betsy Barber, Patrick Lambert and Géraldine Lassalle.

6. Annex

Annex 1. List of individuals involved in the 2019 validation process and their affiliations.

Validator	Affiliation
DiadES Project Partners	
Jimmy King	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	USC/EHEC
Rufino Vieira	USC/EHEC
Carlos Antunes	CMVNC
Catarina Mateus	UE/MARE
Ana Filipa Belo	UE/MARE
Pedro Almeida	UE/MARE
Carlos Alexandre	UE/MARE
Esmeralda Pereira	UE/MARE
Local Experts	
Robert Rosell	AFBI
Trevor Harrison	NIEA
Richard Kennedy	AFBI
Andy Gowans	EA
Charles Crundwell	EA
Darryl Clifton-Dey	EA
Phillip Rippon	EA
John Foster	EA

Validator	Affiliation
Local Experts Continued	
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 Agirre	ANBIOTEK (Basque Country)
Ainhize Uriarte	AZTI (Basque Country)
Ignacio Ferrando	Vaersa (Valencia)
Francesc Jesús Gómez	Government of Catalunya
Jerónimo de la Hoz	Government of Asturias
Marc Ordeix	CERM (Catalunya)
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
Gema Campillo	Government of Cantabria
Pablo Caballero	Xunta de Galicia

7. References

Béguer, M. (2005) 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. p. 33.

Béguer, M., Beaulaton, L. & Rochard, E. (2007) Distribution and richness of diadromous fish assemblages in Western Europe: large scale explanatory factors. Ecology of Freshwater Fish 16, 221-237.

Lassalle, G. (2008) Impacts des changements globaux sur la distribution des poissons migrateurs amphihalins - Une approche par modélisation à l'échelle continentale. Thèse de doctorat, Université Bordeaux I, Bordeaux, France, p. 244.

Lassalle, G., Béguer, M., Beaulaton, L. & Rochard, E. (2008). Diadromous fish conservation plans need to consider global warming issues: an approach using biogeographical models. Biological Conservation 141, 1105-1118.

Lassalle, G. & Rochard, E. (2009) Impact of twenty-first century climate change on diadromous fish spread over Europe, North Africa and the Middle East. Global Change Biology 15, 1072-1089.

Lassalle, G., Crouzet, Ph., Gessner, J. & Rochard, E. (2010) Global warming impacts and conservation responses for the critically endangered European sturgeon. Biological Conservation 143, 2441-2452.

Spalding, M.D., Fox, H.E., Allen, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halpern, B.S., Jorge, M.A., Lombana, A., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A. & Robertson, J., (2007) Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. BioScience 57(7):573–583.