

# SCIENTIFIC DATA

OPEN

## Data Descriptor: One millennium of historical freshwater fish occurrence data for Portuguese rivers and streams

Gonçalo Duarte<sup>1</sup>, Miguel Moreira<sup>2</sup>, Paulo Branco<sup>1</sup>, Luís da Costa<sup>1,3</sup>, Maria Teresa Ferreira<sup>1</sup> & Pedro Segurado<sup>1</sup>

Received: 4 May 2018

Accepted: 21 June 2018

Published: 14 August 2018

The insights that historical evidence of human presence and man-made documents provide are unique. For example, using historical data may be critical to adequately understand the ecological requirements of species. However, historical information about freshwater species distribution remains largely a knowledge gap. In this Data Descriptor, we present the Portuguese Historical Fish Database (PHish-DB), a compilation of 2214 records (557 at the basin scale, 184 at the sub-basin scale and 1473 at the segment scale) resulting from a survey of 194 historical documents. The database was developed using a three-scale approach that maximises the inclusion of information by allowing different degrees of spatial acuity. PHish database contains records of 25 taxonomical groups and covers a time span of one millennium, from the 11th until the 20th century. This database has already proven useful for two scientific studies, and PHish further use will contribute to correctly assess the full range of conditions tolerated by species, by establishing adequate benchmark conditions, and/or to improve existing knowledge of the species distribution limits.

Design Type(s)	data integration objective • time series design • biodiversity assessment objective
Measurement Type(s)	Historical Data Analysis
Technology Type(s)	digital curation
Factor Type(s)	temporal_interval
Sample Characteristic(s)	Portuguese Republic • freshwater river biome

<sup>1</sup>University of Lisbon, School of Agriculture, Forest Research Centre, Tapada da Ajuda, 1349-017 Lisbon, Portugal. <sup>2</sup>University of Lisbon, Instituto Superior Técnico, CERIS-Civil Engineering Research and Innovation for Sustainability, Avenida Rovisco Pais, 1049-001 Lisbon, Portugal. <sup>3</sup>Centre for Ecology, Evolution and Environmental Changes (cE3c), Museu Nacional de História Natural e da Ciência, Rua da Escola Politécnica 56/58, 1250-102 Lisbon, Portugal. Correspondence and requests for materials should be addressed to G.D. (email: goncalo.f.duarte@gmail.com).

## Background & Summary

Collecting historical data on species diversity and occurrence from time periods earlier than the major impactful human activities taking place (e.g., damming, the Industrial Revolution, modern fishing, and river channelisation), may lead to an improvement of knowledge about species ecology. However, historical documents have limitations that need to be understood to avoid incorrect interpretations<sup>1</sup> and/or extrapolations<sup>2</sup>. There is cultural filtering that affects not only the spatial and temporal availability, completeness, and reliability of documentary records but also their quantity and quality<sup>2</sup>. Most historical records rely factually on questionnaires/interviews, hence if there are erroneous answers, the inventories or scientific surveys will present incorrect data<sup>3</sup>. Nevertheless, the utility of historical insight cannot be underestimated and using historical data for ecological studies is valid<sup>2</sup>. Using an interdisciplinary approach<sup>4</sup>, combining information from several independent spatial and temporal sources<sup>1,2,4,5</sup>, cross-checking lines of evidence with independent datasets<sup>6</sup> and blending different methods<sup>2</sup> can help mitigate the limitations and lead to more accurate knowledge about past ecosystem conditions.

For some organisms and specific purposes, historical data might be essential to model the potential species distribution (e.g., Lassalle and Rochard<sup>7</sup>, Clavero and Hermoso<sup>8</sup>) because current distributions are often highly constrained by anthropogenic pressures that alter the natural realised ecological niche. A typical example is the case of diadromous fish species with their inland progression being gradually constrained by the presence of artificial barriers<sup>8,9</sup>. Consequently, current occurrence data will only cover a restricted range of the full conditions tolerated by species. To create the Portuguese Historical Fish Database (PHish-DB) we scouted 194 historical documents, resulting in 2214 records from 30 basins, 280 sub-basins and 490 segments. Data collection started in 2007 and was performed by researchers in history and ecology. Despite some underrepresentation of coastal areas, the spatial distribution of the historical records is homogeneous throughout the country and covers all the major river basins (Fig. 1). Three international river basins stand out (Douro, Minho and Guadiana) with a high number of records (Fig. 1a). The sub-basins with the highest number of records are from the River Tâmega (Douro) and River Zêzere (Tagus) (Fig. 1b). Spatial acuity of the records depended on the information present in the historical source. Thus, we opted for a three-scale approach to maximise the collected information. This has resulted in 557 records limited to the basin scale, 184 reaching the sub-basin scale, and 1473 records identified down to the highest accurate spatial scale, the river segment. PHish database covers a time span of one millennium, from the 11<sup>th</sup> until the 20<sup>th</sup> century, having a larger number of records for the second half of the millennium and particularly for the 18<sup>th</sup> and 19<sup>th</sup> centuries (Fig. 1c). The Interpretation of historical data can be very subjective<sup>2</sup>, and matching ancient fish common names with current taxonomy was challenging. To minimise uncertainty in the taxonomical classification of a fish record, a conservative approach was followed to establish the adequate taxonomical groups. Of the 25 group names defined from the records gathered, three stood out: Petromizontidae, *Chondrostoma* sp. and *Salmo trutta* (Fig. 1d).

The information present in the database has been partially used in the work of Segurado, *et al.*<sup>10</sup>, and also incorporated in a relevant European project, the European Fish Index-Plus (EFI+) (<http://efi-plus.boku.ac.at/>). This database can nevertheless be useful to: improve existing scientific knowledge in Iberian context (e.g., Clavero and Hermoso<sup>8</sup>, Clavero, *et al.*<sup>11</sup>); expand scientific knowledge in European context via an Iberian occurrence scenario of a species with broad-European distribution (e.g., Filipe, *et al.*<sup>12</sup>); be used for research where historical interactions between human activities and riverine fish communities and population are relevant.

## Methods

These methods are expanded and updated versions of descriptions in our related work Segurado, *et al.*<sup>10</sup>.

### Historical sources

The present historical records compilation of riverine fish distribution was based on geographical dictionaries and other published information for Portugal, dated between the 11<sup>th</sup> and early 20<sup>th</sup> centuries. Portugal is the most south-western part of the European continent, representing 15% of the Iberian Peninsula. There are four major international rivers (Douro, Guadiana, Minho, and Tagus) and numerous other relevant national rivers. The available historical information on fish populations for this period was almost exclusively based on qualitative data of species occurrence. Available sources dated before the 16<sup>th</sup> century included charters, inquiries, donations, and monastic chronicles. From the 16<sup>th</sup> century onwards, more thorough recordings of the patrimony of the Portuguese kingdom were available, with the emergence of chorographies, historical-geographical memos, parish inquiries and dictionaries that recorded historically and geographically the Portuguese landscape. In addition to these sources, information from private libraries was also included. A total of 194 documents were consulted (Table 1 (available online only)). These historical sources contain information varying from aspects of the Portuguese physical territory, records about the natural resources of rivers, or cultural context of fisheries exploitation. Most of this data were compiled in the context of the EU projects EFI+ (<http://efi-plus.boku.ac.at/>) and DURERO (Douro River Basin: Water Resources, Water Accounts and Target Sustainability Indices; [http://138.100.137.130/durero\\_project\\_2014/](http://138.100.137.130/durero_project_2014/)), with the main purpose of providing data on reference conditions to compute biotic indicators based on diadromous species. Many regions of Europe have been shaped for centuries by human activities, leading to an absence of natural reference conditions for many water body types<sup>13</sup>. Hence, the definition of benchmark conditions may depend on the



Group Name	Sub-group Name
<i>Acipenser sturio</i>	NA
<i>Alosa alosa</i>	NA
<i>Alosa fallax</i>	NA
<i>Alosa</i> sp.	<i>Alosa alosa</i>
<i>Alosa</i> sp.	<i>Alosa fallax</i>
<i>Alosa</i> sp.	NA
<i>Anaocypris hispanica</i>	NA
<i>Anguilla anguilla</i>	NA
<i>Carassius auratus</i>	NA
<i>Chondrostoma</i> sp.	<i>Achondrostoma occidentale</i>
<i>Chondrostoma</i> sp.	<i>Achondrostoma oligolepis</i>
<i>Chondrostoma</i> sp.	<i>Iberochondrostoma lemmingii</i>
<i>Chondrostoma</i> sp.	<i>Iberochondrostoma lusitanicum</i>
<i>Chondrostoma</i> sp.	<i>Iberochondrostoma</i> sp.
<i>Chondrostoma</i> sp.	NA
<i>Chondrostoma</i> sp.	<i>Pseudochondrostoma duriense</i>
<i>Chondrostoma</i> sp.	<i>Pseudochondrostoma polylepis</i>
<i>Chondrostoma</i> sp.	<i>Pseudochondrostoma willkommii</i>
<i>Cyprinus carpio</i>	NA
<i>Iberochondrostoma almaicai</i>	NA
<i>Lampetra</i> sp.	<i>Lampetra fluviatilis</i>
<i>Luciobarbus bocagei</i>	NA
<i>Luciobarbus sclateri</i>	NA
<i>Luciobarbus</i> sp.	<i>Luciobarbus comizo</i>
<i>Luciobarbus</i> sp.	NA
<i>Mugilidae</i>	NA
NA	<i>Alosa</i> sp.
NA	<i>Anguilla anguilla</i>
NA	<i>Cyprinus carpio</i>
NA	<i>Luciobarbus bocagei</i>
NA	<i>Luciobarbus</i> sp.
NA	NA
NA	<i>Salmo</i> sp.
<i>Petromyzontidae</i>	NA
<i>Petromyzontidae</i>	<i>Petromyzon marinus</i>
<i>Pleuronectiformes</i>	NA
<i>Salmo salar</i>	NA
<i>Salmo</i> sp.	<i>Salmo salar</i>
<i>Salmo trutta</i>	NA
<i>Salmo trutta</i>	<i>Salmo trutta fario</i>
<i>Salmo trutta</i>	<i>Salmo trutta trutta</i>
<i>Salmo trutta fario</i>	NA
<i>Salmo trutta trutta</i>	NA
<i>Squalius</i> sp.	<i>Squalius alburnoides</i>
<i>Squalius</i> sp.	<i>Squalius aradensis</i>
<i>Squalius</i> sp.	<i>Squalius carolitertii</i>
<i>Squalius</i> sp.	<i>Squalius pyrenaicus</i>

**Table 2.** Combination of “Group Name” field and “Sub-group Name” field occurring in the historical records table of the PHish database. NA–Non-applicable.

corresponding upper taxonomic level under the column “Group Name” (information that we recommend to use without any uncertainty). Whenever there were plausible reasons to believe that the record belonged to a given species, the full scientific binomial nomenclature was attributed to the column “Sub-group Name”. This has some associated uncertainty as the decision was made by expert judgement based on the available information. Whenever no plausibility existed, the higher taxonomic group (genus or family) was maintained without the attribution of a “Sub-group Name”. If there existed a possibility of confusion between species that did not fit the higher taxonomical groups defined, NA was attributed to “Group Name”. If plausible, an educated guess, for a species or a taxonomical group, was made into the “Sub-group Name”, based on the interpretation of the historical text extract. All the species and species groups considered are detailed in Table 2. To add value to the database, whenever available, information about the phenology and conservation status (national and international) was included.

### Georeferencing

To create a spatial representation of the historical data we have used the Catchment Characterisation and Modelling– River and Catchment database v2.1 (CCM2) (<http://data.europa.eu/89h/fe1878e8-7541-4c66-8453-afdae7469221>). An advantage of this pan-European database is its hierarchical structure, besides representing a fully integrated system between rivers and drainage catchments<sup>17</sup>. Using three spatial scales (basin, sub-basin and segment) allowed storing historical records with distinct spatial accuracy. Even though finer scales are more informative, historical data at a coarser scale is not irrelevant. For the basin scale, we used the identification code that CCM2 gives for each basin (WSO\_ID) to link an historical record to this scale level. The same procedure was established at the segment scale, using the ID code that CCM2 assigns for each segment (WSO1\_ID). Since CCM2 does not have any identification or spatial representation of the sub-basins within each sea outlet basin, we used a free software to create this information, the River Network Toolkit (RivTool). This software (available at [www.rivtoolkit.com](http://www.rivtoolkit.com)) uses integrated data about river networks and landscape/environmental datasets to produce new or aggregated data via calculations that consider the directional hierarchical network nature of rivers. The set of natural sub-basins of all sea outlet basins of the study area was created using the “sub-basin ID” function of RivTool.

The descriptions found in the historical sources varied greatly in their geographical precision. Most presence records referred to a given river or stream within a restricted region, usually described as being near a given village, township or city. When the geographic location was available, the record was georeferenced in a Geographical Information System (GIS) using CCM2. These were the most spatially precise records, the segment scale, where a segment corresponds to a river reach between two consecutive tributaries. In some cases, regions or town names were obsolete and further investigation was needed to clarify the current location and/or designation associated with that former nomination. Nevertheless, for some historical records, the former names did not have any information or relation with the current designations or did not have enough precision to be linked to a river segment. In those cases, the record was attributed to a higher spatial scale (Sub-basin or Basin). Data entries that could only be related to a watercourse that is a major river or stream that flows to the Atlantic Ocean, coastal lagoons or estuaries, were considered as low precision records and spatially defined at the basin scale. When the watercourse was identified as a tributary (i.e., smaller river or stream not flowing to the Ocean, coastal lagoon or estuary), the precision was considered higher and the record was spatially defined at the Sub-basin scale.

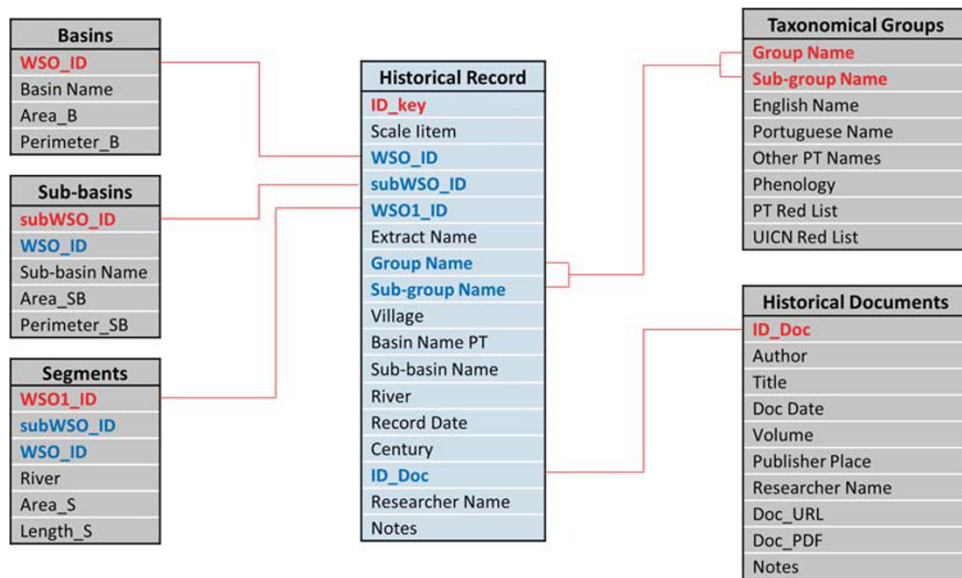
### Data Records

A relational database structure (Fig. 2) was created in Microsoft Access<sup>®</sup> (available in the .accdb file extension) to adequately organise and store the historical data collected with their spatial and temporal dependencies, and also to maintain their link to the historical sources (Table 1 (available online only)). The PHish database is publicly available at the Open Science Framework (Data Citation 1) and at the University of Lisbon, School of Agriculture repository <http://www.isa.ulisboa.pt/proj/PHish/>. The database contains six tables: three related with spatial organisation, “Basins”, “Sub-basins”, “Segments”; one with taxonomical identification, “Taxonomical Groups”; one establishing the details of historical sources, “Historical Documents”; and finally, one aggregating historical record information with respective spatial, taxonomical and historical source information, “Historical Records” (Table 3 (available online only)). The latter table is the core of the relational database structure, relating to all other tables (Fig. 2) and where resulting historical data are stored (Table 3 (available online only)).

### Technical Validation

Interpretation of historical data can be very subjective and historical science is mostly inductive<sup>2</sup>. To increase objectivity and guarantee a correct assessment of historical information it is necessary to perform a critical evaluation of sources<sup>3</sup>, while comparing and combining multiple and independent sources and methods<sup>18</sup>. Special attention was taken to verify if authors were not just replicating information from other sources, and by that leading to duplication of results in the database. This was done not only while researchers were reading and surveying the historical documents and sources, but also by analysing, comparing and searching within the final set of historical records for similarities. For example, combined similarities in taxonomical groups and spatial references, similarities in paragraphs, sentences or parts of sentences were normally an indication that the author was just citing text from another document without acknowledging it explicitly. Whenever there was reasonable doubt about the originality of the





**Figure 2.** The relational structure of the Portuguese Historical Fish Database. Each box represents one table, with the header of a box indicating the name of the table, followed by the list of fields included in the table. The red lines indicate the primary relationship between tables; red fields are the primary key of each table; fields in blue indicate secondary relationships between tables.

information present in the historical source, or of the historical record, only the oldest one was included in the database.

Despite the numerous hurdles, taxonomical identification of ancient species common names followed a conservative approach that guaranteed no uncertainty for the “Group Name” field. Concerning the “Sub-group Name” field, the integration of information between spatial location and taxonomical identification of a record, the reliable considerations based on literature and the knowledge of experienced ecologists assured low levels of uncertainty. Moreover, when there was reasonable doubt or lack of plausibility, no consideration was made.

Spatial information for the records location was primarily accessed based on three Portuguese official water management and administrative sources at GIS environment: 1) Rivers map (Shapefile from [www.hidrografico.pt](http://www.hidrografico.pt)); 2) Administrative regions and municipalities map (Shapefile from [www.dgterritorio.pt](http://www.dgterritorio.pt)); and 3) Online orthophotomaps (WMS link from [www.igeo.pt](http://www.igeo.pt)). When the record city/council name or region was not easily connected to the information available in maps, numerous municipalities and parish websites were consulted, along with other websites from relevant local or regional associations, to help identify the more site-specific or out-dated spatial references. The connection with the CCM2 database was performed only after this thorough process. Records for places or historical locations which were not geographically identifiable were conservatively handled, either by discarding or including them in upper spatial scales (sub-basin or basin scale) when the river name was objectively identified.

### Usage Notes

Just like every database of historical records, the PHish database is neither definitive nor complete. All reasonable and possible updates will be held, though nevertheless dependent on resources and future funding. Methods will be maintained to avoid usage biases and/or interpretation issues. Future surveys for historical data should focus on rectifying spatial and temporal data heterogeneity. Obtaining information for older times (backward from the 16<sup>th</sup> century) and focusing on the coastal areas of Portugal should bridge the spatial and temporal knowledge disparity.

Despite our best efforts, and even considering future updates to this database, true species occurrence will inevitably be broader than what historians and chroniclers may have reported. The cultural filtering<sup>2</sup>, accidental or intentional destruction of documents, doubtful sources of the historians/chroniclers and bias towards certain species<sup>3</sup> affect both spatial and temporal availability, completeness, and reliability of documentary records<sup>2</sup>. In England, copyright law was limited to a special group of people until the 18<sup>th</sup> century; indeed documents availability was still censored and limited to printers and publishers rights rather than to authors properties<sup>19</sup>. The concept of author’s intellectual property over its work only proliferated during the 19<sup>th</sup> century, particularly in culturally developed countries such as France and Great Britain<sup>20</sup>. This means that at least until the 18<sup>th</sup> century, and in Portugal most likely until the 19<sup>th</sup> century, authors could quote other works without acknowledging them. Thus, data duplication is a possibility within this database, though we consider it in a very low probability given our cautious

approach to this conundrum. Also, users must be aware that PHish database is a presence-only compilation of historical records. Without a wary systematic sampling, absence data is inevitably prone to a high degree of uncertainty, and to our best knowledge, no modern-day systematic survey of fish assemblages across the country was undertaken in Portugal until the end of the 20<sup>th</sup> century.

The lack of data for coastal areas, and more specifically for the southern coastal regions of Portugal may result from several particular circumstances. Smaller basins are composed of smaller rivers and inevitably with less human settlements. Adding to this, southern Iberian rivers are Mediterranean-type freshwater ecosystems strongly shaped by autumn/winter flooding and summer drought events<sup>21</sup>. This seasonal instability has implications in the structure of freshwater communities<sup>22</sup>, meaning that these rivers will probably tend to support less interesting fishing areas and species. The database is also temporally unbalanced, i.e., although covering a vast time-scale it does not represent a consistent time-series due to lack or reduced number of records for the first half of the millennium. Future updates to this database will probably not fully overcome this as it is also the result of temporal filtering<sup>3</sup> of historical sources. Another relevant issue is the heterogeneity of the established taxonomical groups in the user recommended field (Group Name). Our conservative approach followed herein avoids uncertainty in this classification, translating into correct and objective taxonomical information. However, for example, it may be thwarting to use some taxonomical groups (e.g., “Mugilidae” or “*Chondrostoma*”) when the objective is to perform species environmental niche modelling.

All mentioned issues, biases and unbalances are normal for historical databases, and none of them hampers the usage of the database. To our knowledge, this database is the first public compilation of historical distribution data based on freshwater fish species in Portuguese rivers and South-western Europe. Notwithstanding, users should keep in mind all these features and caveats whenever making any considerations or extrapolation based on this database. The PHish database is geographically limited since it is restricted to inland Portugal. However, the current database contains valuable information by covering data for the Portuguese areas, including sea outlets, of four Iberian international rivers and most Portuguese major watercourses. Moreover, because Portuguese data were not compiled and made available until now, researchers have been using only Spanish records to study fish species distribution for the whole Iberian Peninsula, minimizing its importance as a meaningful biogeographical entity<sup>23</sup>. By using only Spanish data, authors concede to uncertain premises (e.g., Clavero and Villero<sup>24</sup>) and/or extrapolate when predicting for the entire Iberian Peninsula (e.g., Clavero and Hermoso<sup>8</sup>). Hence, this database will fill an important gap in the current knowledge and contribute to the development of new studies covering the whole Iberian Peninsula without being hindered by political borders.

## References

- Rackham, O. in *Conservation Science and Action* (ed. Sutherland William J.) Ch. 8, 152–175 (Blackwell Publishing Company, 1998).
- Swetnam, T. W., Allen, C. D. & Betancourt, J. L. Applied Historical Ecology: using the past to manage for the future. *Ecological Applications* **9**, 1189–1206 (1999).
- Haidvogel, G. *et al.* Typology of historical sources and the reconstruction of long-term historical changes of riverine fish: a case study of the Austrian Danube and northern Russian rivers. *Ecology of Freshwater Fish* **23**, 498–515 (2013).
- Szabó, P. Why history matters in ecology: an interdisciplinary perspective. *Environmental Conservation* **37**, 380–387 (2010).
- Hayashida, F. M. Archaeology, Ecological History, and Conservation. *Annual Review of Anthropology* **34**, 43–65 (2005).
- Crumley, C. in *The World System and the Earth System: Global Socioenvironmental Change and Sustainability Since the Neolithic* Hornborg Alf & Crumley Caroleeds 15–28 (Left Coast Press, 2007).
- Lassalle, G. & Rochard, E. 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 (2009).
- Clavero, M. & Hermoso, V. Historical data to plan the recovery of the European eel. *Journal of Applied Ecology* **52**, 960–968 (2015).
- Lassalle, G., Crouzet, P. & Rochard, E. Modelling the current distribution of European diadromous fishes: an approach integrating regional anthropogenic pressures. *Freshwater Biology* **54**, 587–606 (2009).
- Segurado, P., Branco, P., Avelar, A. & Ferreira, M. T. Historical changes in the functional connectivity of rivers based on spatial network analysis and the past occurrences of diadromous species in Portugal. *Aquatic Sciences* 1–14 (2014).
- Clavero, M. *et al.* Historical citizen science to understand and predict climate-driven trout decline. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20161979 (2017).
- Filipe, A. F. *et al.* Forecasting fish distribution along stream networks: brown trout (*Salmo trutta*) in Europe. *Diversity and Distributions* **19**, 1059–1071 (2013).
- Hohensinner, S. *et al.* Type-specific reference conditions of fluvial landscapes: A search in the past by 3D-reconstruction. *CATENA* **75**, 200–215 (2008).
- Béguer, M., Beaulaton, L. & Rochard, E. Distribution and richness of diadromous fish assemblages in Western Europe: large-scale explanatory factors. *Ecology of Freshwater Fish* **16**, 221–237 (2007).
- European Commission. in *L 327 (European Parliament ed)* 93 (The Official Journal of the European Union, 2000).
- Robalo, J. I., Almada, V. C., Levy, A. & Doadrio, I. Re-examination and phylogeny of the genus *Chondrostoma* based on mitochondrial and nuclear data and the definition of 5 new genera. *Molecular Phylogenetics and Evolution* **42**, 362–372 (2007).
- Vogt, J. *et al.* *A pan-European River and Catchment Database*. 120 (European Commission-Joint Research Centre - Institute for Environment and Sustainability: Luxembourg, 2007).
- White, P. S. & Walker, J. L. Approximating Nature’s Variation: Selecting and Using Reference Information in Restoration Ecology. *Restoration Ecology* **5**, 338–349 (1997).
- Patterson, L. R. *Copyright in historical perspective* (Vanderbilt University Press, 1968).
- Geller, P. E. Copyright history and the future: What’s culture got to do with it. *Journal of the Copyright Society of the USA* **47**, 209 (2000).
- Gasith, A. & Resh, V. H. Streams in Mediterranean Climate Regions: Abiotic Influences and Biotic Responses to Predictable Seasonal Events. *Annual Review of Ecology and Systematics* **30**, 51–81 (1999).

22. Pires, A. M., Cowx, I. G. & Coelho, M. M. Seasonal changes in fish community structure of intermittent streams in the middle reaches of the Guadiana basin, Portugal. *Journal of Fish Biology* **54**, 235–249 (1999).
23. Ribeiro, F., Elvira, B., Collares-Pereira, M. J. & Moyle, P. B. Life-history traits of non-native fishes in Iberian watersheds across several invasion stages: a first approach. *Biological Invasions* **10**, 89–102 (2008).
24. Clavero, M. & Villero, D. Historical Ecology and Invasion Biology: Long-Term Distribution Changes of Introduced Freshwater Species. *BioScience* **64**, 145–153 (2014).

## Data Citation

1. Duarte, G. *et al.* *Open Science Framework* <https://doi.org/10.17605/OSF.IO/TXCYS> (2018).

## Acknowledgements

The authors would like to thank Paula Avelar, Armanda Rodrigues, Inês Vila, José Maria Santos, João Oliveira, Paulo Pinheiro and José Neiva for their participation in this work. Gonçalo Duarte is part of the FLUVIO doctoral program and financed by a grant from the Fundação para a Ciência e Tecnologia (FCT) (SFRH/BD/52514/2014). Miguel Moreira is also part of the FLUVIO doctoral program and financed by a grant from the Fundação para a Ciência e Tecnologia (FCT) (PD/BD/114558/2016). Paulo Branco was financed by a post-doc grant from FCT (SFRH/BPD/94686/2013). Luís da Costa was financed by the MARS Project (Managing Aquatic ecosystems and water Resources under multiple Stress) funded under the 7th EU Framework Programme, Work Package 4 (Multiple stressors at the river basin scale). Pedro Segurado is supported by a contract funded by the Fundação para a Ciência e Tecnologia (FCT) under the IF Researcher Programme (IF/01304/2015). Centro de Estudos Florestais (CEF; “Forest Research Centre”) is a research unit funded by Fundação para a Ciência e a Tecnologia I.P. (FCT), Portugal (UID/AGR/00239/2013). Part of the historical compilation has been supported by the EU project FP6-2005-SSP-5A-044096 (<http://efi-plus.boku.ac.at/>). Part of the historical compilation has been supported by the project DURERO (Douro River Basin: Water Resources, Water Accounts and Target Sustainability Indices; [http://138.100.137.130/durero\\_project\\_2014/](http://138.100.137.130/durero_project_2014/)).

## Author Contributions

G.D. supervised and collaborated in the historical sources analysis and data collection, helped establishing the spatial location for the historical records, helped preparing the data for the relational database, created the access relational database, wrote part of the Data Descriptor, and reviewed the full manuscript. M.M. collaborated in the historical sources analysis and data collection, searched and established the spatial location for the historical records, prepared the data for the relational database, wrote part of the Data Descriptor, and reviewed the full manuscript. P.B. helped establishing the taxonomical groups, helped establishing the structure of the relational database, wrote part of the Data Descriptor, and reviewed the full manuscript. L.C. collaborated in the historical sources analysis and data collection, helped searching and establishing the spatial location for the historical records, helped establishing the taxonomical groups, helped creating the access relational database, and reviewed the full manuscript. M.T.F. coordinated the interaction between institutions, coordinated the work of all the researchers involved, and reviewed the full manuscript. P.S. coordinated the historical sources analysis and data collection, helped establishing the structure of the relational database, wrote part of the Data Descriptor, and reviewed the full manuscript.

## Additional information

Tables 1 and 3 are available only in the online version of this paper.

**Competing interests:** The authors declare no competing interests.

**How to cite this article:** Duarte, G. *et al.* One millennium of historical freshwater fish occurrence data for Portuguese rivers and streams. *Sci. Data* **5**:180163 doi: 10.1038/sdata.2018.163 (2018).

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files made available in this article.

© The Author(s) 2018