Metadata description of the AMAZON FISH database

Céline Jézéquel, Pablo A. Tedesco, Javier A. Maldonado-Ocampo, Saúl Prada-Pedreros, Hernán Ortega, Max Hildalgo, Koen Martens, Aaike De Wever, Jansen Zuanon, Gislene Torrente-Vilara, Leandro M. Sousa, Tommaso Guiarrizzo, Jonathan Valdiviezo, Jaime Sarmiento, Mabel Maldonado, Fernando M. Carvajal-Vallejos, Takayuki Yunoki, Carlos Donascimiento, Edwin Agudelo, José Iván Mojica, Raúl Ríos Herrera, Francisco Villa & Thierry Oberdorff
Metadata description of the AMAZON FISH database

Céline Jézéquel 1, Pablo A. Tedesco 1, Javier A. Maldonado-Ocampo 2, Saúl Prada-Pedreros 2, Hernán Ortega 3, Max Hildalgo 3, Koen Martens 4, Aaike De Wever 4, Jansen Zuanon 5, Gislene Torrente-Vilara 6, Leandro M. Sousa 7, Tommaso Guiarrizzo 7, Jonathan Valdiviezo 17, Jaime Sarmiento 9, Mabel Maldonado 10, Fernando M. Carvajal-Vallejos 17, Takayuki Yunoki 11, Carlos Donascimiento 12, Edwin Agudelo 13, José Iván Mojica 14, Raúl Ríos Herrera 15, Francisco Villa 16 & Thierry Oberdorff 1

1 Institut de Recherche pour le Développement (IRD), UMR EDB, Toulouse, France; corresponding author: celine.jezequel@ird.fr
2 Pontificia Universidad Javeriana, Bogota, Colombia
3 Universidad Nacional Mayor de San Marcos, Universidad del Perú, Decana de América, Museo de Historia, Lima, Peru
4 Royal Belgian Institute of Natural Sciences (RBINS), Brussels, Belgium
5 Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brasil
6 Universidade Federal de São Paulo (UNIFESP), São Paulo, Brasil
7 Universidade Federal do Pará (UFPa), Pará, Brasil
8 Museo Ecuatoriano de Ciencias Naturales (MECN), Quito, Equador
9 Museo Nacional de Historia Natural, Instituto de Ecología, La Paz, Bolivia
10 Universidad Mayor de San Simón, Cochabamba, Bolivia
11 Universidad Autónoma del Beni, Centro de Investigación de Recursos Acuáticos, Trinidad, Bolivia
12 Instituto Alexander von Humboldt (IAvHP), Bogota, Colombia
13 Instituto Amazónico de Investigaciones Científicas SINCMI, Leticia, Colombia
14 Instituto de Ciencias Naturales, Universidad Nacional de Colombia (ICNHN), Bogota, Colombia
15 Instituto para la Investigación y la Preservación del Patrimonio Cultural y Natural (INCIVA), Cali, Colombia
16 Grupo de Investigación en Zoología, Facultad de Ciencias, Universidad del Tolima, Ibagué, Colombia
17 Instituto Nacional De Biodiversidad (INABIO), Quito, Ecuador

Please cite this paper as follows: Jézéquel, C., Tedesco, P. A., Maldonado-Ocampo, J. A., Prada-Pedreros, S., Ortega, H., Hildalgo, M., Martens, K., De Wever, A., Zuanon, J., Torrente-Vilara, G., Sousa, L. M., Guiarrizzo, T., Valdiviezo, J., Sarmiento, J., Maldonado, M., Carvajal-Vallejos, F. M., Yunoki, T., Donascimiento, C., Agudelo, E., Mojica, J. I., Herrera, R. R., Villa, F. & Oberdorff, T., 2019. Metadata description of the AMAZON FISH database. Freshwater Metadata Journal 43: 1-9. https://doi.org/10.15504/fmj.2019.43

Received: 2017-12-21 / Published: 2019-03-18

Keywords
fish, database, Amazon basin, Amazon river, database description

Short description of the dataset/summary

The title "AMAZON FISH database" is a working title for the Amazonian Freshwater Fish Biodiversity database that was constructed in the framework of the AMAZON FISH (ERANetLAC/DCC-0210) project.

In order to construct the AMAZON FISH database, which aims to become a high quality freshwater fish biodiversity
database covering the entire Amazon drainage basin, data from a wide range of sources were compiled in a central database. The content of this database can be considered as a collection of curated datasets or dataset exports (as the database only covers the Amazon basin, while source datasets may cover larger areas).

These datasets are of different nature and include:
A. Information extracted from the literature (published articles, books, gray literature).
B. Data available from museums/collections and universities.
C. Data held or compiled by project partners (e.g. at country level). Some of these have previously been published through the Global Biodiversity Information Facility (GBIF) network, while others will be released in the framework of AMAZON FISH.
D. Data extracted from online databases such as GBIF, FishNet, SpeciesLink and IABIN.
E. New data obtained during (a) sampling campaign(s) in the framework of the project.

At this stage (December 2018), the database already provides a comprehensive overview of the fish diversity and distribution in the Amazon basin, with over 305,000 occurrence records from 123 sources (mostly covering B, C & D). After the project formally ends, it will still be updated by including new data from the project partners (C), from literature references (A) and potentially by including data from new sampling campaigns (E) to areas for which we identified a gap in the available data.

Through the integration of these different data sources, the AMAZON FISH database constitutes a unique resource on the Amazon basin, which concentrations the highest freshwater biodiversity on earth. By compiling the knowledge on the spatial distribution of freshwater fishes and addressing the taxonomic and sampling gaps, this database should enable comprehensive analysis and modeling. As such it will answer the essential need for understanding this unique ecosystem that has produced such high freshwater fish diversity.

**General information**

- **dataset entry ID:** AF_1
- **name of the dataset:** AMAZON FISH database
- **type of dataset:** species distribution data
- **science keywords according to GCMD:** Biosphere, Biological Classification
- **ISO topic category according to ISO 19115:** Biota, Environment, Inland Waters
- **INSPIRE keywords according to GEMET:** Species distribution
- **related project:** AMAZON FISH (ERANet-LAC/DCC-0210)
- **funding:** ERANet-LAC: Network of the European Union, Latin America and the Caribbean Countries on Joint Innovation and Research Activities

**Technical and administrative specifications**

- **data format:** others/specify
- **others/details:** PostgreSQL
operating system: all operating systems
data language: English
current access level: restricted access
currently available through GBIF: no
exchange planned: no
data in data repository: no

Do you plan to publish the data on the Freshwater Biodiversity Data Portal:
yes
update level: update planned
documentation:
type: others/specify
others/details: Data Management Plan (https://doi.org/10.6084/m9.figshare.5150716.v1)
language: English

contact details:
metadata contact person:
  first, last name: Céline Jézéquel
  email: celine.jezequel@ird.fr
  institution: Institut de Recherche pour le Développement (IRD)
  address: UMR EDB, 118 route de Narbonne
  postal code, city: 31000 Toulouse
  country: France

technical contact person:
  first, last name: Céline Jézéquel
  email: celine.jezequel@ird.fr
  institution: Institut de Recherche pour le Développement (IRD)
  address: UMR EDB, 118 route de Narbonne
  postal code, city: 31000 Toulouse
  country: France

scientific contact person:
  first, last name: Thierry Oberdorff
  email: thierry.oberdorff@ird.fr

Intellectual property rights and citation
dataset creator (data compiler):
  contact name: Thierry Oberdorff
  contact email: thierry.oberdorff@ird.fr
  contact institution: Institut de Recherche pour le Développement (IRD), UMR EDB 5174
data contributors to/owners of this dataset:
  number: multiple
  data contributor/owner 1:
    contact name: Thierry Oberdorff
    contact email: thierry.oberdorff@ird.fr
    contact institute: Institut de Recherche pour le Développement (IRD), UMR
    criteria for using this part of the dataset:
    The dataset needs to be requested from dataset creator with specific conditions of use.
data contributor/owner 2:
- contact name: Javier A. Maldonado-Ocampo
- contact email: maldonadoj@javeriana.edu.co
- contact institute: Pontificia Universidad Javeriana, Unidad de Ecología y Sistemática (UNESIS)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 3:
- contact name: Hernán Ortega
- contact email: hortega@gmail.com
- contact institute: Museo de Historia Natural, Universidad Nacional Mayor de San Marcos
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 4:
- contact name: Aaike De Wever
- contact email: aaike.dewever@naturalsciences.be
- contact institute: Royal Belgian Institute of Natural Sciences (RBINS)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.
- comments: No direct data contributions included. Contribution mainly consisted of technical advise during database construction.

data contributor/owner 5:
- contact name: Jansen Zuanon
- contact email: jzuanon3@gmail.com
- contact institute: Instituto Nacional de Pesquisas da Amazônia (INPA)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 6:
- contact name: Gislene Torrente-Vilara
- contact email: gtvilara@gmail.com
- contact institute: Universidade Federal de São Paulo (UNIFESP)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 7:
- contact name: Leandro M. Sousa
- contact email: m.sousa@gmail.com
- contact institute: Universidade Federal do Pará (UFPA)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 8:
- contact name: Jonathan Valdiviezo
- contact email: jonathan.valdiviezo@biodiversidad.gob.ec
- contact institute: Instituto Nacional De Biodiversidad (INABIO) - Ecuador
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.
**data contributor/owner 9:**
- contact name: Jaime Sarmiento
- contact email: jsarmientotavel@gmail.com
- contact institute: Museo Nacional de Historia Natural, Instituto de Ecología
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 10:**
- contact name: Mabel Maldonado
- contact email: mabelmaldonado.m@fcyt.umss.edu.bo
- contact institute: Universidad Mayor de San Simón, Unidad de Limnología y Recursos Acuáticos (ULRA)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 11:**
- contact name: Takayuki Yunoki
- contact email: takayukiyunoki@yahoo.com
- contact institute: Universidad Autónoma del Beni, Centro de Investigación de Recursos Acuáticos
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 12:**
- contact name: Carlos Donascimiento
- contact email: cdonascimiento@humboldt.org.co
- contact institute: Instituto Alexander von Humboldt (IAvHP)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 13:**
- contact name: Edwin Agudelo
- contact email: eagudelo@sinchi.org.co
- contact institute: Instituto Amazónico de Investigaciones Científicas SINCHI (CIACOL)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 14:**
- contact name: José Iván Mojica
- contact email: jimojicac@unal.edu.co
- contact institute: Instituto de Ciencias Naturales, Universidad Nacional de Colombia (ICNMHN)
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.

**data contributor/owner 15:**
- contact name: Raúl Ríos Herrera
- contact email: raulriosh@gmail.com
- contact institute: Instituto para la Investigación y la Preservación del Patrimonio Cultural y Natu
- criteria for using this part of the dataset: The dataset needs to be requested from dataset creator with specific conditions of use.
data contributor/owner 16:
contact name: Francisco A. Villa-Navarro
contact email: favilla@ut.edu.co
contact institute: Universidad del Tolima (CZUT-IC)
criteria for using this part of the dataset:
The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 17:
contact name: Max Hidalgo
contact email: mhidalgod@unmsm.edu.pe
contact institute: UNMSM, Peru
criteria for using this part of the dataset:
The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 18:
contact name: Fernando M. Carvajal-Vallejos
contact email: fernando.carvajal@faunagua.org
contact institute: FAUNAGUA
criteria for using this part of the dataset:
The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 19:
contact name: Céline Jézéquel
contact email: celine.jezequel@ird.fr
contact institute: IRD
criteria for using this part of the dataset:
The dataset needs to be requested from dataset creator with specific conditions of use.

data contributor/owner 20:
contact name: Pablo A. Tedesco
contact email: pablo.tedesco@ird.fr
contact institute: IRD
criteria for using this part of the dataset:
The dataset needs to be requested from dataset creator with specific conditions of use.

citation of this dataset:
author(s): Jézéquel, C., Tedesco, P. A., Maldonado-Ocampo, J. A., Prada-Pedreros, S., Ortega, H., Hildalgo, M., Martens, K., De Wever, A., Zuanon, J., Torrente-Vilara, G., Sousa, L. M., Guiarrizzo, T., Valdiviezo, J., Sarmiento, J., Maldonado, M., Carvajal-Vallejos, F. M., Yunoki, T., Donascimiento, C., Agudelo, E., Mojica, J. I., Herrera, R. R., Villa, F. & Oberdorff, T.
title and journal (name, number, pages):
The Amazon Fish database: freshwater fish distribution for the entire Amazon drainage
year: 2019
citation of the metadata:
author(s): Jézéquel C., Tedesco P. A., Maldonado-Ocampo J. A., Prada-Pedreros S., Ortega H., Hildago M., Martens K., De Wever A., Zuanon J., Torrente-Vilara G., Sousa L. M., Guiarrizzo T., Valdiviezo J., Sarmiento J., Maldonado M., Carvajal-Vallejos F. M., Yunoki T., Donascimiento C., Agudelo E., Mojica J. I., Herrera R. R., Villa F. & Oberdorff T.
The AMAZON FISH database integrates a wide range of datasets from different public and project partners sources, and results from a three years international team effort.

The entire, curated database at the sub-basins scale (modified HydroBASINS dataset) will be documented and released in a data paper that should be submitted during the year 2019. All occurrence records (point scale data) for which we obtained permission for public release will be made available through the AMAZON FISH data online repository or through the Freshwater Biodiversity Portal (http://data.freshwaterbiodiversity.eu/; which is a member of the Freshwater Information Platform). The release of the occurrence records will be effective three years after the official publication of the data paper, a timeframe considered sufficient to publish the analyses suggested in the initial AMAZON FISH projet. During this embargo period, site scale data may be available on request pending agreement of all partners and with authorship negotiation possible.

The sub-basins scale data paper will include an extensive database description along with a discussion on the database content, gap analysis and database statistics, in parallel with the on-line release of the database at the sub-basins scale.

In addition to the publication of the datasets and database itself, key maps produced during the AMAZON FISH project will be disseminated through the Global Freshwater Biodiversity Atlas (http://atlas.freshwaterbiodiversity.eu/) following the publication of scientific paper(s) describing those results.

**General data specifications**

**regional coverage of the dataset:**
- catchment
- South America

**spatial extent (bounding coordinates):**
- southernmost latitude [°]: -25
- northernmost latitude [°]: 5
- westernmost longitude [°]: -85
- easternmost longitude [°]: -45
- minimum altitude: 0 metres
- maximum altitude: 6200 metres
- countries: South America: Bolivia, Brazil, Colombia, Ecuador, Peru, Venezuela

**world climatic regions according to Köppen:**
- Group A: tropical/megathermal climates

**freshwater ecoregions of the world (FEOW) according to WWF:**
- South America: Amazonas Estuary & Coastal Drainages, Amazonas Guiana Shield, Amazonas High Andes, Amazonas Lowlands, Guapore - Itenez, Madeira
ecosystem type: Brazilian Shield, Mamore - Madre de Dios Piedmont, Rio Negro, Tapajos - Juruena, Ucayali - Urubamba Piedmont, Western Amazon Piedmont, Xingu
covered timeframe: 1865 - 2017

Site specifications
coordinate system/grid data: latitude/longitude, format: DD
datum (e.g. WGS84): WGS84
grid data available: no
number of sites: >1000
exact number of sites: 22000

catchments, river-sub-basins

Biological data
biological data origin: from sampling,
AMAZON FISH project ERANetLAC/DCC-0210
comments: Timeframe of the sampled sites: 1865 to 2017.
organism group addressed: fish

Sample resolution
fish: species
level: species
percentage of species level data: 100

taxonomic coding: FishBase & California Academy of Sciences-Catalogue of Fishes
Fricke, R., Eschmeyer, W. & Van der Laan, R. (2018). Catalog of Fishes: Genera, Species, References.
http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.n.asp
Froese, R. & Pauly, D. (2018). FishBase. World Wide Web electronic publication. www.fishbase.org

Other specifications
GIS layers, shape files related to the dataset:
catchments, river-sub-basins
others/details: Modified version of HydroBASINS.
availability of photos: no
availability of maps: no
quality control procedures:
Were any quality control procedures applied to your dataset?
yes
quality control protocols and comments:
To avoid any duplication when integrating data from different sources, all
duplicates are identified and excluded during import by either verifying unique identifiers for occurrences (or fields that should constitute unique identifiers when considered together; e.g. institute + collection + collection number) or based on sampling details in combination with the geographical coordinates of the data.

Where available, details on provenance, citation, along with a modification/"last updated" timestamp (e.g. only available for GBIF and SpeciesLink) are systematically documented for the retained records.

In addition, during import, each record is systematically checked for reliability and consistency, using the FishBase Consortium database, and the California Academy of Science’s Catalog of Fishes, as nomenclature authority files. During the process of building the database, a taxonomic checklist work is achieved to validate all species actually present in the Amazon drainage basin. This important work is done in collaboration with taxonomic experts covering all groups and will be made publicly available through a checklist publication.

Acknowledgements
This project is supported by the ERANet-Lac programme - Grant number: ERANetLAC/DCC-0210.