General information

name of the dataset: AMAZON FISH database
  full name of the dataset: AMAZON FISH database
  dataset short name: AMAZON FISH database

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 concentrates 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
AMAZON FISH database 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.

keywords according to GCMD:
- topic: Biosphere, Biological Classification

ISO topic category according to ISO 19115:
- Biota, Environment, Inland Waters

INSPIRE keywords according to GEMET:
- Species distribution

own science keywords:
- freshwater fish, Amazon river drainage, Amazon basin, occurrence database, 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: Céline Jézéquel
  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: Céline Jézéquel
  first, last name: Céline Jézéquel
  email: celine.jezequel@ird.fr

scientific contact person: Thierry Oberdorff
  first, last name: Thierry Oberdorff
  email: thierry.oberdorff@ird.fr
Intellectual property rights and citation

(if the dataset is already published):

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
  provider 1:
    provider institute: Institut de Recherche pour le Développement (IRD), UMR
    contact name: Thierry Oberdorff
    contact email: thierry.oberdorff@ird.fr
    criteria for using the data in a publication/scientific analysis: The dataset needs to be requested from dataset creator with specific conditions of use.
  provider 2:
    provider institute: Pontificia Universidad Javeriana, Unidad de Ecología y Sistemática (UNESIS)
    contact name: Javier A. Maldonado-Ocampo
    contact email: maldonadoj@javeriana.edu.co
    criteria for using the data in a publication/scientific analysis: The dataset needs to be requested from dataset creator with specific conditions of use.
  provider 3:
    provider institute: Museo de Historia Natural, Universidad Nacional Mayor de San Marcos
    contact name: Hernán Ortega
    contact email: hortega@gmail.com
    criteria for using the data in a publication/scientific analysis: The dataset needs to be requested from dataset creator with specific conditions of use.
  provider 4:
    provider institute: Royal Belgian Institute of Natural Sciences (RBINS)
    contact name: Aaike De Wever
    contact email: aaike.dewever@naturalsciences.be
    criteria for using the data in a publication/scientific analysis: 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.
  provider 5:
    provider institute: Instituto Nacional de Pesquisas da Amazônia (INPA)
    contact name: Jansen Zuanon
    contact email: jzuanon3@gmail.com
    criteria for using the data in a publication/scientific analysis: The dataset needs to be requested from dataset creator with specific conditions of use.
  provider 6:
    provider institute: Universidade Federal de São Paulo (UNIFESP)
    contact name: Gislene Torrente-Vilara
contact email: gtvilara@gmail.com

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 7:**

provider institute: Universidade Federal do Pará (UFPA)

contact name: Leandro M. Sousa

contact email: m.sousa@gmail.com

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 8:**

provider institute: Instituto Nacional De Biodiversidad (INABIO) - Ecuador

contact name: Jonathan Valdiviezo

contact email: jonathan.valdiviezo@biodiversidad.gob.ec

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 9:**

provider institute: Museo Nacional de Historia Natural, Instituto de Ecología

contact name: Jaime Sarmiento

contact email: jsarmientotavel@gmail.com

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 10:**

provider institute: Universidad Mayor de San Simón, Unidad de Limnología y Recursos Acuáticos (ULRA)

contact name: Mabel Maldonado

contact email: mabelmaldonado.m@fcyt.umss.edu.bo

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 11:**

provider institute: Universidad Autónoma del Beni, Centro de Investigación de Recursos Acuáticos

contact name: Takayuki Yunoki

contact email: takayukiyunoki@yahoo.com

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 12:**

provider institute: Instituto Alexander von Humboldt (IAvHP)

contact name: Carlos Donascimiento

contact email: cdonascimiento@humboldt.org.co

criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

**provider 13:**

provider institute: Instituto Amazónico de Investigaciones Científicas SINCHI (CIACOL)
contact name: Edwin Agudelo  
contact email: eagudelo@sinchi.org.co  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 14:  
provider institute: Instituto de Ciencias Naturales, Universidad Nacional de Colombia (ICNMHN)  
contact name: José Iván Mojica  
contact email: jimojicac@unal.edu.co  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 15:  
provider institute: Instituto para la Investigación y la Preservación del Patrimonio Cultural y Natu  
contact name: Raúl Ríos Herrera  
contact email: raulriosh@gmail.com  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 16:  
provider institute: Universidad del Tolima (CZUT-IC)  
contact name: Francisco A. Villa-Navarro  
contact email: favilla@ut.edu.co  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 17:  
provider institute: UNMSM, Peru  
contact name: Max Hidalgo  
contact email: mhidalgod@unmsm.edu.pe  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 18:  
provider institute: FAUNAGUA  
contact name: Fernando M. Carvajal-Vallejos  
contact email: fernando.carvajal@faunagua.org  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 19:  
provider institute: IRD  
contact name: Céline Jézéquel  
contact email: celine.jezequel@ird.fr  
criteria for using the data in a publication/scientific analysis: 
The dataset needs to be requested from dataset creator with specific conditions of use.

provider 20:
criteria for using the data in a publication/scientific analysis:

The dataset needs to be requested from dataset creator with specific conditions of use.

citation of this dataset:

author(s):

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):

title and journal (name, number, pages):
Metadata description of the AMAZON FISH database. Freshwater Metadata Journal 0: 0-0

year: 0000
doi (if applicable):
https://doi.org/10.15504/fmj.0000.0

comments:
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:
spatial extent of the dataset: catchment
continents: 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 Brazilian Shield, Mamore - Madre de Dios Piedmont, Rio Negro, Tapajos - Juruena, Ucayali - Urubamba Piedmont, Western Amazon Piedmont, Xingu

ecosystem type:
rivers, wetlands

covered timeframe:
year from: 1865
year to: 2017
Site specifications

coordinate system/grid data:  latitude/longitude
                                  projected

datum (e.g. WGS84):  WGS84

grid data available:  no

number of sites:  >1000

exact number of sites:  22000
**Biological data**

**biological data origin:** from sampling,

**specify project:** AMAZON FISH project ERANetLAC/DCC-0210

**specify method:**

**comments:**

*Timeframe of the sampled sites: 1865 to 2017.*

**organism group addressed:** fish
Sample resolution

fish:
taxonomic resolution:
  percentage of species level data: 100
taxonomic coding:
  taxalist according to: FishBase & California Academy of Sciences-Catalogue of Fishes
  http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp

sample specifications:
Other specifications

GIS layers, shape files related to the dataset: catchments, river-sub-basins

availability of photos: no
availability of maps: no

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.