Catalog of the Fishes of Colombia

Programme: BID
Project ID: BID-CA2020-030-USE
Project lead organization: Universidad Nacional de Colombia
Project implementation period: 1/7/2021 - 30/6/2023
Report approved: 27/1/2022

Narrative Early Progress report

Executive Summary

The Catalog of the Fishes of Colombia has gathered over 75 specialists to complete the taxonomic and ecological information of marine and freshwater species distributed in Colombia.
We have made progress towards updating the checklist of the freshwater fishes, which is the backbone for the freshwater species. We have started communications with several icthyological collections in Colombia to publish additional specimen datasets in GBIF and SiB Colombia, including an important dataset from the Universidad del Quindio. These data will be incorporated in the Catalog via API and will help confirm identification, validity, and distribution of species. We started conversations to incorporate existing information about endemic species currently listed and completed in the Catalogue of Colombian Biodiversity, and are in preliminary talks with FishBase about bidirectional data sharing. We started mapping the database fields and this will be completed before the end of the year. We have recruited two interns to assist with marine species checklist and with data cleaning and quality control. A Master's student was also recruited to collaborate on aspects of Catalogue development and interoperability. The project progress has been monitored via email, and in meetings with the editors, authors and project partners. We recently implemented a beta version for a ticket system that will allow a more accurate assignment of tasks and follow up of duties.

Planning phase - User(s) consultation outcome(s)

Defining Data User Needs

The Catalog is intended for several audiences (e.g. researchers, decision makers) and will inform government institutions that legislate the fish and fisheries resources in Colombia and non-governmental organizations that support conservation efforts. Among those entities we have the Ministry of the Environment and Sustainable Development (MADS), National Natural Parks of Colombia (PNN), National Authority of Aquaculture and Fisheries (AUNAP), WWF Colombia, and Invemar. We are aiming to answer several scientific questions: What species are present in Colombia and what areas have they been reported? How can the species be used sustainably? What areas of Colombia need to be reviewed based on the presence of species with problematic conservation status? How many invasive species are currently reported in Colombia? Which entities have governance over specific species?
The partners and future users of the Catalog are looking for a platform that is easy to use and intuitive and that will help all audiences answer specific questions about the species.
The Catalog will be a go-to resource for the partner entities and will aid in the management of fisheries and hydrobiological resources by providing information about conservation status. The government and non-government institutions have mentioned that having the information aggregated in a single platform will allow better decision making regarding the use of the resource a priori and in situ (e.g. fishing ports).
We will ensure user input into the development process via communication and training workshops. An API is being designed in consultation with our partners specifically to meet their needs. Evaluation criteria we can implement include the number of queries made, citations in government documents and peer-reviewed publications, and updates to fish-related legislation.

Defining Data Needs

1. The primary function of the Catalog is to provide a specimen-based, expert verified checklist of all

the fish species in Colombia. This will enable decision makers and stakeholders to identify which commercial and hydrobiological resources we have and where in Colombia they are distributed. We are currently working with our partners in government and collaborating NGOs to define how the taxonomic backbone and geographic information from the Catalog can be made interoperable with their systems.

2.The most suitable dataset types for addressing our users' needs are the checklist and occurrence data, as well as the species pages.

3.We will mobilize (and update) existing data about endemic species in Colombia, this data is currently hosted in the biodiversity catalog. We will also mobilize data relevant to models of distribution for species (Biomodelos), this data is hosted by the Humboldt Institute. We will mobilize and update the data from the freshwater fishes checklist and the images from the Visual Catalog of Fishes (CavFish) by the Universidad de Ibagué. We are also in conversations with FishBase about bidirectional sharing of information (FishBase provides data to the Catalogue of Life).

4. Some of our collaborating authors are revising GBIF records to help investigate doubtful species or outlying distributions. The specimen holding institutions can be queried for photographs or in-person visits arranged to corroborate specimen identification.

5. To make the data accessible for a broad audience, we are creating an online portal similar to the Catalog of the Plants and Lichens of Colombia, using a minimalistic web design with simple searches by common or scientific name and highly customizable advanced searches with different types of filters (i.e by region, taxonomic level, etc.). With our partners we are developing an API to share information that will fill their specific needs.

6. The completion of the marine checklist and updating of the freshwater list are being undertaken by taxonomic specialists. The creation of the online resource is being undertaken by the Biodiversity Informatics Program at the Instituto de Ciencias Naturales, which has extensive experience in data publishing and creation of online databases. Processing of data consumed by our partners via the API will be done by professionals within their respective agencies.

7. The Catalog will be hosted on Universidad Nacional de Colombia infrastructure, which is housed in a data center administered by the National Directorate of Digital Strategies, which provides virtual servers under the NetApp system implementing VMware solutions. In the Biodiversity Informatics Program of the ICN, directed by the PI, has a several machines in that data center and for the project we expect to have a new server under linux operating system for provide a robust solution for the web application, also the databases are managing in different machines (specific servers for databases), and the images of the museum specimens will be shared via a Loris image server hosted on a dedicated virtual machine using the University's infrastructure. For live specimen images we are partnering with the CavFish project, and the images are stored on servers at the Universidad de Ibagué. The Catalog will be hosted in a subdomain under unal.edu.co main domain (i.e catalogopecesdecolombia.unal.edu.co).

9. Within the timeframe of the project we expect to complete all of the planned activities.

Planning Phase Outcomes

We are not making substancial changes to the project. Some of our partners have wish lists of the types of data they would like to see included in the Catalog, especially regarding commercial use of species. Our commitment was to share expert verified taxonomic and geographic information with them. We did not commit to including datasets from government agencies in the Catalog but if they are willing to do so, we will try to accommodate them, however we need to use caution here because our experts cannot verify that taxonomic identification in information about fish catches for example (which almost never have voucher specimens) is correct. We do expect to mobilize at least one additional dataset that was not originally planned (Icthyology collection of the Universidad del Quindío) A provisional workflow diagram is attached.

Report on Activities

Activity progress summary

Some marine species are still awaiting assignment to specialists. If none can be found, the editorial committee will review them. The updating of the freshwater checklist is slightly behind schedule due to one of the editors getting sick from Covid (he is now recovered). The list is now up to 1616 species from 1432 in the original publication. We expect to complete it in early 2022 if not before. The list of Caribbean marine species is also behind schedule due to some authors needing additional time, but will also be finished in the first half of 2022 and the delay will not impact the overall project goals. Integration of information on endemic species from the Catalogue of Colombian Biodiversity (SiB Colombia) is underway. Completion of the backend database and interface for authors is delayed due to a family tragedy but should be completed in November and is not expected to impact the project timeline significantly. A first planning workshop was held with partners in September. Other activities in the timeline are scheduled to begin in 2022.

Activity: Assignment of species

Description: Some marine taxa have not yet been assigned to specialists. If no specialist exists or are available to treat these groups, then members of the editorial committee will assume responsibility for orphan taxa.
Start Date - End Date: 10/1/2020 - 17/12/2021
Verification Sources: Here is the current list of orphan species:
Ilisha amazonica
Pellona castelnaeana
Pellona flavipinnis
Pristigaster cayana
Amazonsprattus scintilla
Anchoviella guianensis
Anchoviella jamesi
Jurengraulis juruensis
Lycengraulis batesii
Pterengraulis atherinoides

Activity: Data completion of freshwater species

Description: This activity was initially projected to for completion by November 30, 2021. We are behind schedule due to one of the primary editors being sick with Covid, but progress has been made. The original list had 1432 species. The current revised list has 1616 species. Once the modifications to the backend of the database and author user interface are complete the authors can make additional edits directly in the Catalog.
Start Date - End Date: 1/7/2021 - 31/3/2022
Verification Sources: revised list of freshwater species is in development

Activity: Digital Platform Design

Description: The key person working on this aspect of the project suffered a family tragedy and needed to take a few weeks off work. The backend of the Catalog and interface for authors to edit content were scheduled for completion at the end of September and is now pushed to mid-November.
Start Date - End Date: 1/7/2021 - 15/11/2021
Verification Sources: We are including some screenshots of the database in the attachments

Activity: Planning Workshop with Partner Institutions

Description: Workshop to present the project, discuss partner needs and expectations. We have since been in contact with our partners about their specific data needs and technical requirements for API design and implementation.
Start Date - End Date: 18/8/2021 - 18/8/2021
Verification Sources: Meeting notes:
https://docs.google.com/document/d/1vr9II0j_9NB_82SgKMt4gcZtzSb3LtIt80Dkon4mfIE/edit

Meeting recording:
https://drexel.zoom.us/rec/share/AoOyazuwAtVR3-
6xbFOf0Rght7Ns7YtKUezDK6mmnGPHWxLPW98F5QZS-ldqKh64.HiTwlH_3xUJ7sc6j?
startTime=1629299753000

access code for the recording is
+ZJq3+lc

Activity: Caribbean fishes checklist publication

Description: The completion of the list is behind schedule and thus the publication will be delayed until the first half of 2022. Some of the authors needed more time.
Start Date - End Date: 1/7/2021 - 30/6/2022
Verification Sources: Once the list is published we will provide the link

Activity: Data completion of marine species

Description: Underway
Start Date - End Date: 1/7/2021 - 31/5/2022
Verification Sources: Slightly delayed to due to delays in production of interface for authors to edit content, but should not affect projected end date. Will give update in the midterm report and provide link to published dataset once it is available.

Activity: Data integration for endemic freshwater species

Description: Inclusion in database of descriptive data for 400 endemic species from Catalogue of

Colombian Biodiversity. This activity is underway and projected to be completed on time.
Start Date - End Date: 1/10/2021 - 31/1/2022
Verification Sources: once the data have been integrated into the database we can provide screen shots

Report on Deliverables

Deliverables progress summary

All originally planned deliverables are in development. The freshwater fishes list and the list of Caribbean marine fishes will be finished in the first half of 2022 instead of November 2021 (see activities section, above) but will not impact the overall project outcomes. The production of the backend of the Catalog is also slightly behind schedule and will be completed in mid November 2021 (see activities section above). This delay will also not impact the overall project outcomes. A new deliverable is planned: publication of a specimen occurrence dataset from the University of Quindio.

Progress towards deliverables

Deliverables - Project planning phase

Checklist of freshwater species containing 1494 species of fishes distributed in Colombia
Description: A corrected checklist of Colombian freshwater fishes in which the species occurrence have been confirmed using data from collections, this list includes all records published for the country. The current version of the list can be found here: DOI:10.15472/numrso.
% complete: 80%
Status update: List at time of project submission had 1492 species; it is now up to 1616 species. We are waiting for specialist review of some taxa.
Sources of verification: The current version of the list can be found here: DOI:10.15472/numrso. Checklist data will be uploaded to Catalogue and published in SiB Colombia and GBIF

Planning Workshop
Description: Initial workshop to consult with partners on development of platform and APIs
% complete: 100%
Status update: We are following up with individual partners via emails and phone calls
Sources of verification: see links to meeting notes and recording under related project activity (above) and in Events section (below)

Digital platform to host the Catalog of the Fishes of Colombia
Description: Structured database and interfaces for including taxonomic and ecological information for the fishes of Colombia (marine and freshwater) as well as images, geographic distribution, and information relevant to Colombian collections with material of each species.
% complete: 90%
Status update: some delays due to a tragedy in the family of the person leading this task
Sources of verification: screen shots in Attachments

Checklist of Colombian Pacific marine fish containing 800-900 species
Description: Verified checklist of Pacific marine fish species including both continental and insular areas. We aim to cover all marine fishes in this region which includes three classes Class Myxini, Class Chondrichthyes, and Class Osteichthyes
% complete: 80%
Status update: Data need to be imported into Catalog; orphan taxa still need to be assigned.
Sources of verification: No links yet. A datapaper will be published in peer reviewed journal; checklist data uploaded to Catalogue and published in SiB Colombia and GBIF

Checklist of Colombian Caribbean marine fish containing 1400-1500 species
Description: Verified checklist of Caribbean marine fish species including both continental and insular areas. We aim to cover all marine fishes in our region which includes three classes Class Myxini, Class Chondrichthyes, and Class Osteichthyes
% complete: 80%
Status update: Data need to be imported into Catalog; orphan taxa still need to be assigned
Sources of verification: No links yet. A datapaper will be published in a peer reviewed journal; checklist data uploaded to Catalogue and published in SiB Colombia and GBIF

Integration of data on endemic freshwater fishes
Description: SiB Colombia currently publishes a resource on endemic freshwater fish with approximately 400 species. These will be shared with the Catalog so as not to duplicate efforts and updates and corrections will be communicated to SiB Colombia.

% complete: 20%
Status update: Discusiones underway with editors of the original list to define protocols for updates and planning is underway with SiB Colombia for us to import the data.
Sources of verification: No links yet. Indicators will be endemic freshwater fish data imported to Catalog; updates, corrections transmitted to SiB Colombia

Deliverables - Project data mobilization phase

Ichthyology collection University of Quindio
Dataset type: Occurrences
Dataset scope: Specimen dataset from University of Quindio representing freshwater species occurrences in the Central Cordillera. Not sure yet how many records we can expect. An estimate is given below.
Number of records: 100
Data holder: Universidad del Quindío
Data host institution: UNAL
% complete: 0%
Status update: This is a new dataset that we did not include in the original proposal. The curator is very interested in publishing the data in GBIF and SiB Colombia. The records are in Excel and we need to review them and bring them into Darwin Core compliance. It is a well curated collection with valuable information for this part of the country.
DOI:
Expected date of publication: 2023-03-31

Checklist of Fishes from Colombia
Dataset type: Checklist
Dataset scope: records of museum specimens to support distribution of marine fish species included in the Catalog. Records for freshwater species already exist. The marine data will be published here for the first time.
Number of records: 2,260
Data holder: ACICTIOS and Universidad Nacional de Colombia
Data host institution: UNAL and SiB Colombia
% complete: 10%
Status update: specimen digitization is underway at ICN (UNAL)
DOI:
Expected date of publication: 2023-03-31

Checklist of Fishes from Colombia
Dataset type: Checklist
Dataset scope: a list of all fish species (marine and freshwater) occurring in Colombia including their taxonomic status, ecological data, and conservation status in the country. At time of proposal submission 3794 species were included. This number is expected to change and and estimated number of records is provided below
Number of records: 4,000
Data holder: ACICIOS, UNAL
Data host institution: UNAL and SiB Colombia
% complete: 80%
Status update: Orphan taxa still need to be assigned; updates are nearly complete for Caribbean species. Marine species data will be added in the coming months.
DOI:
Expected date of publication: 2022-06-30

Deliverables - Project evaluation phase

Photographs of living and preserved specimens
Description: Compilation of photographs of living and preserved specimens sourced from collaborators to represent as many of the species as possible.
% complete: 10%
Status update: We have secured a commitment from the curators of the CavFish database of the Universidad de Ibagué to share photos of live specimens. Discussions are underway to explore technical options for sharing and displaying the images
Sources of verification: N/A, but you can review the CavFish portal here: https://cavfish.unibague.edu.co/

Training Workshops with partner and non-partner institutions
Description: Training workshop to present the final version of the platform and teach users how to search, generate reports, and provide user feedback.
% complete: 0%
Status update: this is planned for the end of the project

Sources of verification: Not aplicable yet

Documentation for Catalog of the Fishes of Colombia
Description: Documentation for the Catalogue of the Fishes of Colombia online platform and corresponding API will be provided.
% complete: 0%
Status update: this process has not yet begun
Sources of verification: N/A

Catalog of the Fishes of Colombia
Description: Searchable online platform fully populated with all content (data and images) running on servers at Universidad Nacional de Colombia
% complete: 10%
Status update: we have some preliminary designs for the website
Sources of verification: N/A (this is for the end of the project)

APIs
Description: Application programming interfaces implemented for sharing taxonomic, distribution and other Catalog data with PNN, AUNAP, MADS, and others.
% complete: 10%
Status update: discusiones underway and technical requirements are being evaluated
Sources of verification: N/A

Events

Planning Workshop

Dates: 2021-08-23 - 2021-08-23
Organizing institution: UNAL
Country: Colombia
Number of participants: 12
Comments: Links to the notes from the meeting and the recording of the meeting are also posted in the activities section.
Website or sources of verification: https://drexel.zoom.us/rec/share/AoOyazuwAtVR3-6xbFOf0Rght7Ns7YtKUezDK6mmnGPHWxLPW98F5QZS-ldqKh64.HiTwlH_3xUJ7sc6j?startTime=1629299753000 (passcode +ZJq3+lc)

GBIF