THE ATLANTIC-WIDE RESEARCH PROGRAMME FOR BLUEFIN TUNA (GBYP Phase 11)

SHORT TERM CONTRACT FOR THE SUPPORT TO THE DEVELOPMENT OF THE ICCAT ELECTRONIC TAGS MANAGEMENT SYSTEM "ETAGS"

(ICCAT GBYP 02/2022)

Final Report

Hong Kong, 25 August 2022

Big Fish Intelligence Company Limited

Dr. Chi Hin Lam, contractor

This project is co-funded by the European Union

Background

Over the years ICCAT has released in the Atlantic Ocean and adjacent Seas many electronic tags on tuna, tuna-like and various shark species, to record information on the behaviour and migrations of those species. A large portion of those electronic tags are related to ICCAT research programmes (www.iccat.int/en/ResProgs.html) in which GBYP (about 59 internal archival tags and 446 satellite pup-up tags released since 2008) and AOTTP (about 430 internal archival tags and 169 satellite pop-up tags) are two major contributors. Overall, these tagging activities have behind a considerable investment. Yet, the associated information was never properly stored in a centralised relational database held in ICCAT, which would greatly improve its potential use in scientific studies, allowing for example the development (or improvement of existing ones) of more efficient and complex analytical tools.

Currently, the ICCAT electronic tagging information has various weaknesses:

a) Lacks a complete and efficient inventory (metadata: scientific programmes, deployment activities, tag event characteristics, raw data availability, manufacturer's raw binary files, resulting scientific work associated, etc.),

b) The raw binary files are spread across various laboratories and/or scientists (only a small portion is held in ICCAT, mostly associated to GBYP and AOTTP programmes),

c) The raw data files are archived using various structures and formats (in many cases reflecting the changes made manufacturers over the years),

d) The different models of electronic tags (internal archival, satellite pop-up) used over time and the output formats used by each manufacturer could be inconsistent in time (changes in software, data field policies, etc.)

The ICCAT existing and future electronic tagging information will be much more valuable to the ICCAT scientific community, if all the information is validated and stored in a centralised relational database, together with all the associated metadata. Therefore this "ETAGS" project is conceived to build the necessary software infrastructure to handle the enormous electronic data needs of ICCAT, and provide a means for future development and integration of its other data systems within the organization. This is achieved mainly through:

1. A simple, flexible flat file format "eTUFF" that serves as an intermediary exchange format to facilitate the consolidation of data products from various tag manufacturers
2. The electronic tag data in eTUFF format can then be uploaded to the a specialized database management system, Tagbase-server (github.com/tagbase/tagbase-server). Tagbase-server is a Flask (flask.palletsprojects.com/en/2.0.x) application which provides OpenAPI REST endpoints for ingestion of various files into the Tagbase SQL database (PostgreSQL engine).
3. Data can be managed through Tagbase-server, which allows connections to various visualization and analytical endpoints

Given its complexity (combination of various technologies, large amounts of data to inventory, recover and treat, etc.) this project was envisioned to be completed over at least three phases. This report described the activities and work carried out during Phase 1 between January 15 and June 30, 2022.

Progress to date

A. Improvements to the relational data model

The original data model assumed a 1-to-1 relationship between a particular tag deployment and its associated geo-positioned track. However, due to the limitations and updates to light-based geolocation algorithms, it is common to have multiple track solutions that either represents different processing options or improvements on estimation over time and/ computation capabilities. After consultation with ICCAT, we had updated the relational data model that maintains the original design concepts while accommodating multiple (potentially unlimited) track solutions. The upgraded data model was vetted by ICCAT staff and became formally incorporated into Tagbase-server. Other minor improvements were also adapted on the suggestions of the Secretariat. A schematic is represented below:

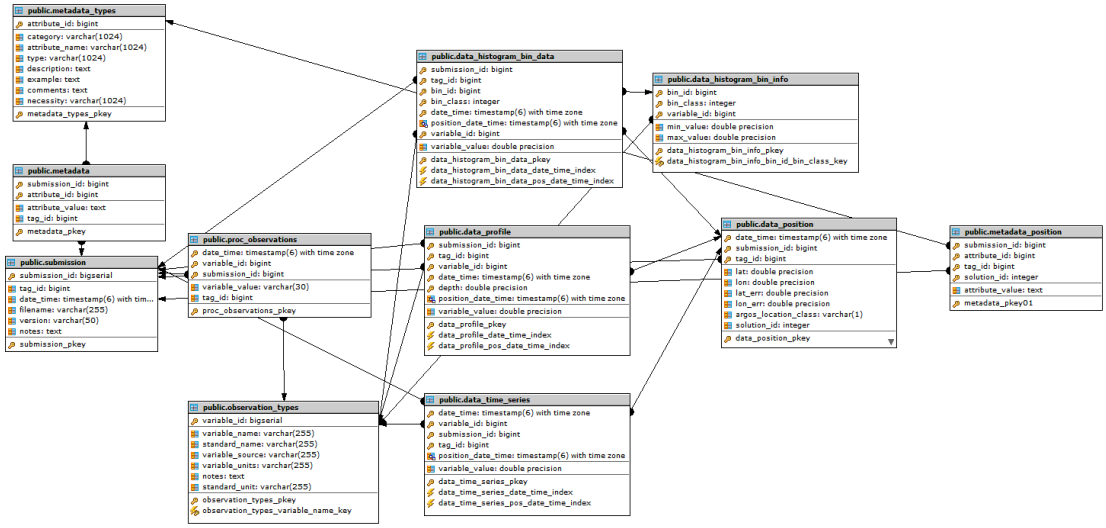

Figure 1. Updated relationship-entity model of Tagbase-server

B. Adaptation of supported metadata attributes to ICCAT needs

A major advantage of Tagbase-server is the integration of detailed metadata that informs an instance of a tag deployment and its data recovery. Such metadata can cover various details from hardware and field operations, environmental conditions to biological parameters and other experimental treatments of the tagged animal. Given the vast amount of possible metadata attributes, we conducted five rounds of collaborative discussions to finalize the list of attributes useful to ICCAT and their operations. This iteration process also prompted internal discussions within the Secretariat to formulate data sharing and usage privilege policies, which resulted in a new category of metadata, termed as “share”. Within this “share” category, we streamlined the number of attributes down to five (Table 1) to assist with ICCAT’s management of data usage for the eventual, integrated system. An internal, draft document on tag data usage has been developed by the Secretariat to describe sharing scenarios and provides the basis for future discussions and updates.

Category	AttributeI	AttributeName	Description
share	81	data_dissemination_level	Permission to distribute instrument data to personnel other than the owner(s)
share	82	results_dissemination_level	Permission to show/ publish instrument data in any setting (e.g., journal research article, technical report, general audience magazine feature or web-based GIS)
share	83	restriction_expiration_date	If a restriction is placed by the data owner(s) on sharing or publishing, when will this expire?
share	84	citation	DOI of a published article/ SCRS document number/ url/ website where the data set has been published
share	85	data_repository	DOI or URL of the dataset in a public repository, e.g., www.pangaea.de

Table 1. New metadata attributes incorporated into the metadata types supported by Tagbase-server

C. Technological upgrades to Tagbase-server

In order to support the use of Tagbase-server in ICCAT, we conducted a comprehensive review on database integrity and security at the beginning of the project. Since then, major upgrades have been conducted to strengthen the system for operational deployment that resulted in 5 major beta-releases within this Phase 1 period. All the releases are fully documented and changes tracked, and code traceable in the Github repository (github.com/tagbase/tagbase-server/releases). A listing of the notable improvements is summarized below:

1. Modernized and upgraded components
2. Improved code quality
3. Review and eliminate vulnerabilities
4. Ensured maximum code traceability through Github issue tracking, release history & documentation
5. Code expansion developed in mind for long-term expansion/ capable deployment, e.g., Amazon Web Services & CloudFormation
6. Comprehensive logging
7. Real-time operational notifications

Continuity tasks

A number of developmental areas are initiated in Phase 1 and will continue into Phase 2, and progress on each of them is briefly highlighted.

1. Support of multiple tag manufacturers and track solutions

Throughout the project, we have been updating and modifying R routines (github.com/camrinbraun/tags2etuff) that assist with the preparation of tag manufacturer data products into a simple common flat file format, eTUFF, to facilitate import into Tagbase-server. Various upgrades were made to ensure support to Wildlife Computers output files generated from their web-based cloud portal. We also extended support to some legacy files generated from older, desktop-based decoding software. These efforts were conducted mainly because the bulk of ICCAT data holdings are based on tags manufactured by Wildlife Computers. We are now in the process to extend the same support to the tag outputs obtained from other manufacturers that ICCAT had used, namely Lotek Wireless and Microwave Telemetry. Debugging and continuous improvements are expected in Phase 2 as more tag files are being processed.

2. Implementation of Tagbase-server at ICCAT for test runs

A first beta test version of Tagbase-server was delivered on July 13, 2022. We are working with the Secretariat on setting up a local test environment, due to a delay in obtaining and setting up a Rackspace server. More testing and tuning are planned to ensure ICCAT staff can test drive Tagbase-server and provide necessary feedback. This activity will be featured prominently in Phase 2.

3. Customization of components based on data access patterns

We have developed a few modes of file ingestion pipelines in Phase 1, providing support to file transfer from FTP, HTTPS, bulk upload via zip file etc., given ICCAT's various number of potential data contributors from around the world. A follow-up step in Phase 2 is to identify preferred data access patterns and streamline an ingestion routine that is best set up for the Secretariat. The same customizations will be applied to other tools we have built for system notifications and logging. The goal will be to make a responsive system that helps the Secretariat to acquire and import data into Tagbase-server.

Acknowledgements

This work has been carried out under the ICCAT Atlantic-Wide Research Programme for Bluefin Tuna (GBYP), which is funded by the European Union, several ICCAT CPCs, the ICCAT Secretariat, and other entities (see https://www.iccat.int/gbyp/en/overview.asp). The content of this paper does not necessarily reflect ICCAT's point of view or that of any of the other sponsors, who carry no responsibility. In addition, it does not indicate the Commission's future policy in this area.