Data sharing standard for archival and satellite tags

Version 1.2

Jan. 13, 2016

Xavier Hoenner1, Rob Harcourt2

1 Integrated Marine Observing System, University of Tasmania, Private Bag 110, Hobart, TAS 7001, Australia.

2 Department of Biological Sciences, Faculty of Science and Engineering, Macquarie University, Sydney, NSW 2109, Australia.

Table of contents

1. Introduction 2

2. Database schema design 2

3. Discussion 3

4. Database schema: tables and relationships 5

Schema diagram 5

Device 6

Tag deployment 8

Tag recovery 9

Animal 10

Species 12

Animal measurement 14

Project 15

Data center 17

Users 19

Project role 20

Organisation 21

5. References 22

# 1. Introduction

Following the Animal Telemetry Data & Metadata Meeting held in Halifax on July 14 2015 one action item was to develop a metadata template for archival and satellite telemetry projects that could be used to share metadata and eventually data easily across organizations. The need for such a standard has become obvious due to a growing number of projects (*e.g.* TOPP, POST, AATAMS) collecting important volumes of archival and satellite telemetry data (Kranstauber et al. 2011, Campbell et al. 2015, Dwyer et al. 2015, Campbell et al. 2016), which are more likely to result in the publication of articles in high impact factor journals (Block et al. 2011, Hussey et al. 2015, Kays et al. 2015).

The standard proposed hereafter is primarily based on the metadata convention developed for acoustic telemetry projects (Payne et al. 2013) and the structure of the AATAMS database (AATAMS 2015), with additional metadata information gathered from various projects (*e.g.* OTN, TOPP, Movebank) (Kranstauber et al. 2011). The present convention follows the nomenclature used in Payne et al. (2013) with database fields flagged as ‘required’ or ‘optional’ so that data users will be able to get a minimum amount of viable information for each dataset. Constraints and relationships between database tables (Figure 1) are indicated to highlight how information from different tables can be joined together. For additional database related logic, refer to Payne et al. (2013). The proposed tables were designed using the PostgreSQL database language and the PostGIS extension to convert longitude/latitude coordinates into geometries; this PostgreSQL code is available for download at: <https://goo.gl/KEFwJT>.

# 2. Database schema design

We propose that all metadata should be contained within 11 different tables presented below. Detailed field names and meanings for each database table are provided in section 4. Each table has (1) a primary key which constrains some fields to have a unique combination of values, and (2) a set of foreign keys which allow joins between tables so that information from different tables can be extracted in a flat CSV table format.

1. [**Device**](#_Device_1): information about each tag that has been deployed including manufacturer name, model, and onboard sensors.
2. [**Tag deployment**](#Tag_deployment): information about which tag was deployed, on which animal, by whom, and how.
3. [**Tag recovery**](#Tag_recovery): information about which tag was recovered, when and where.
4. [**Animal**](#Animal): information about each animal equipped with tags including sex, species, age, capture and release dates and locations.
5. [**Species**](#_Species_1): Read only table providing a simplified list of species. This list should eventually follow a comprehensive standard approved upon by the tagging community (*e.g.* WoRMS for marine species, <http://www.marinespecies.org/>).
6. [**Animal measurement**](#Animal_measurement): lists all the morphological measurements taken for each tagged animal.
7. [**Project**](#Project): information on tagging projects, which data center hosts their data, who is the principal investigator.
8. [**Data center**](#Data_center): name and attributes of each data center.
9. [**User**](#_User_1)**s**: names and contact details of people involved in tagging projects or data centers.
10. [**Project role**](#Project_role): mapping table between the project and users tables.
11. [**Organisation**](#_Organisation_1): information about institutions users belong to.

# 3. Discussion

While we believe all critical metadata information have been included in the list of database tables (section 4), the present model remains simplistic and its robustness could be improved by adding constraints on various fields to avoid content inconsistencies. This preliminary model now requires feedbacks from the archival and satellite tagging community before more effort be devoted to improving this metadata template, in particular:

* Need of controlled vocabularies (see ‘Constraints’ column in section 4) to standardize the content of some database fields.
* Need to agree upon a taxonomical standard for the list of species.

We also recommend that, in parallel to this reviewing process, be established a central repository (*e.g.* Dropbox folder) for the design of an archival and satellite tagging **data** template. Setting this up would allow researchers to upload sample data files from different tag manufacturers and models so that database managers can develop standard format to deliver different data products (*e.g.* CTD profiles, location data, TDR data).

# 4. Database schema: tables and relationships

## Schema diagram

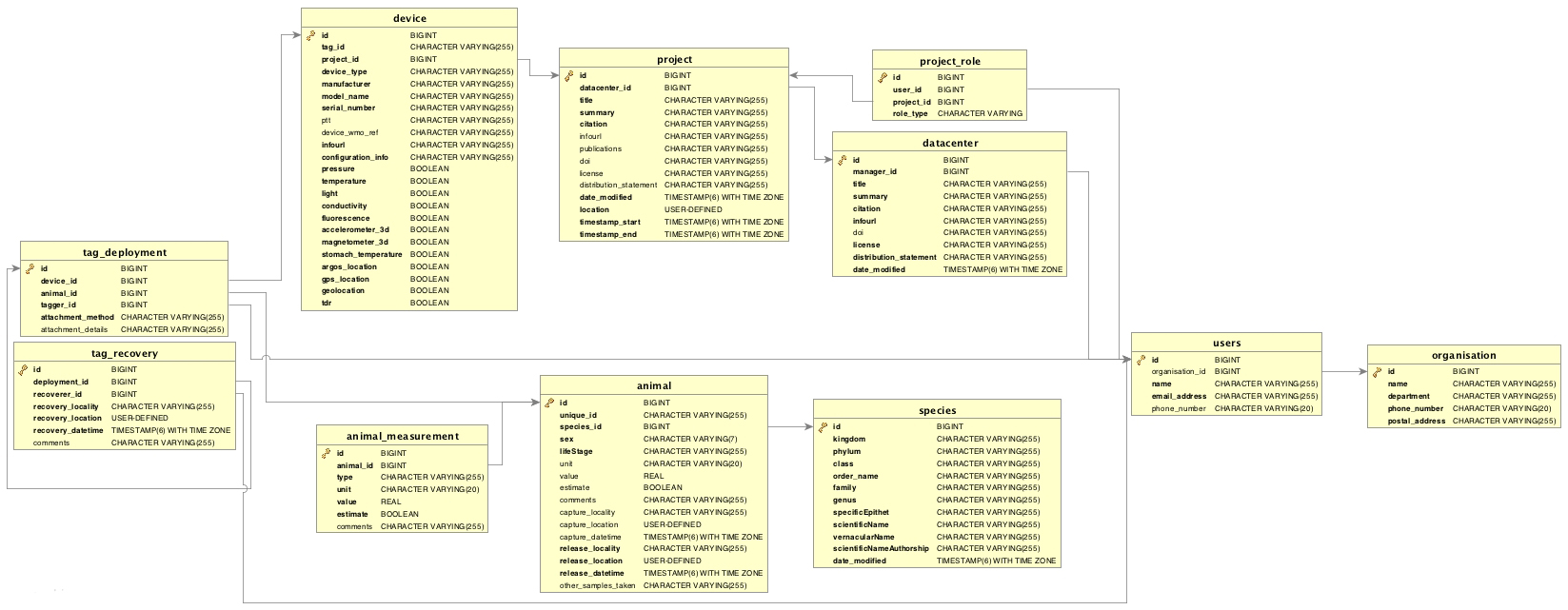


Figure 1: Database schema showing the content and relationships between the 11 different tables.

## Device

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information about each tag that has been deployed including manufacturer name, model, and onboard sensors.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Device ID (unique). | required | numeric | Primary key |  |
| tag\_id | Tag ID under each project (*e.g.* ct111-030-13) | required | text |  |  |
| project\_id | Project ID. | required | numeric | Foreign key to project table |  |
| device\_type | Type of device (*e.g.* archival, pop-up, satellite). | required | text | Need controlled vocabulary |  |
| manufacturer | Name of manufacturer. | required | text | Need controlled vocabulary |  |
| model\_name | Model name. | required | text | Need controlled vocabulary |  |
| serial\_number | Device serial number. | required | text |  |  |
| ptt | Platform Transmitting Terminal (PTT) number for Argos transmission. | optional | text |  |  |
| device\_wmo\_ref | World Meteorological Organization (WMO) number allocated to this device. | optional | text |  |  |
| infourl | URL to tag model specifications. | required | text |  | ERDDAP |
| configuration\_info | Describe how the tag was configured prior to deployment (*e.g.* at what depth a dive starts, how many times the tag should try transmitting data). | required | text |  |  |
| pressure | Was pressure data recorded? | required | boolean |  |  |
| temperature | Was temperature data recorded? | required | boolean |  |  |
| light | Was light data recorded? | required | boolean |  |  |
| conductivity | Was conductivity data recorded? | required | boolean |  |  |
| fluorescence | Was fluorescence data recorded? | required | boolean |  |  |
| accelerometer\_3d | Was speed data recorded? | required | boolean |  |  |
| magnetometer\_3d | Was orientation data recorded? | required | boolean |  |  |
| stomach\_temperature | Was stomach temperature data recorded? | required | boolean |  |  |
| argos\_location | Which type(s) of location estimates was provided by the tag? | required | boolean |  |  |
| geolocation | required | boolean |  |  |
| gps\_location | required | boolean |  |  |

## Tag deployment

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information about which tag was deployed, on which animal, by whom, and how.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Tag deployment ID (unique). | required | numeric | Primary key |  |
| device\_id | Device ID. | required | numeric | Foreign key to device table |  |
| animal\_id | Animal ID. | required | numeric | Foreign key to animal table |  |
| tagger\_id | ID of the person who deployed the device on the animal. | required | numeric | Foreign key to users table. |  |
| attachment\_method | Describe how the tag was attached (*e.g.* glued, implant). | required | text | Need controlled vocabulary |  |
| attachment\_details | Describe tag attachment (*e.g.* exact location of tag on animal, use of anaesthetics). | optional | text |  |  |

## Tag recovery

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information about which tag was recovered, when and where.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Tag recovery ID (unique). | required | numeric | Primary key. |  |
| deployment\_id | Tag deployment ID. | required | numeric | Foreign key to tag\_deployment table. |  |
| recoverer\_id | ID of the person who recovered the device on the animal | required | numeric | Foreign key to users table |  |
| recovery\_locality | Locality, State/Territory, Country of recovery | required | text |  |  |
| recovery\_location | Longitude and latitude of recovery, transformed into a geometry (point). | required | geometry |  |  |
| recovery\_datetime | Date and time (UTC) of tag recovery. | required | timestamp |  |  |
| comments | Describe the recovery event (*e.g.* damage on tag, biofouling, tag sent back to manufacturer for refurbishing). | optional | text |  |  |

## Animal

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information about each animal equipped with tags including sex, species, age, capture and release dates and locations.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Animal ID (unique). | required | numeric | Primary key |  |
| unique\_id | Unique ID for each animal *(e.g*. numbered tag, band, transponder*)* | required | text |  |  |
| species\_id | Species ID. | required | numeric | Foreign key to species table |  |
| sex | Sex of animal. | required | text | Can only be female, male, or unknown. | Darwin Core |
| lifeStage | Animal life stage (*e.g.* adult, juvenile, subadult, weaner) | required | text | Need controlled vocabulary | Darwin Core |
| unit | Unit of age (*e.g.* days, months, years) | optional | text | Need controlled vocabulary |  |
| value | Age of animal | optional | numeric |  |  |
| estimate | Is the age value an estimate? | optional | boolean |  |  |
| comments | Additional information on animal (*e.g.* animal origin – wild vs. hatchery, stock, damages) | optional | text |  |  |
| capture\_locality | Locality, State/Territory, Country of capture | optional | text |  |  |
| capture\_location | Longitude and latitude of capture, transformed into a geometry (point). | optional | geometry |  |  |
| capture\_datetime | Date and time (UTC) of capture. | optional | timestamp |  |  |
| release\_locality | Locality, State/Territory, Country of animal release | required | text |  |  |
| release\_location | Longitude and latitude of animal release, transformed into a geometry (point). | required | geometry |  |  |
| release\_datetime | Date and time (UTC) of animal release. | required | timestamp |  |  |
| other\_samples\_taken | Describe whether any samples were taken during the tagging experiment (*e.g.* tissue, blubber, blood). | optional | text |  |  |

## Species

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Simplified list of species (read only). This list should eventually follow a comprehensive standard approved upon by the tagging community (*e.g.* WoRMS for marine species,** [**http://www.marinespecies.org/**](http://www.marinespecies.org/)**)** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Species ID (unique). | required | numeric | Primary key |  |
| kingdom | Taxonomical details for each species. | required | text |  | Darwin Core |
| phylum | required | text |  | Darwin Core |
| class | required | text |  | Darwin Core |
| order\_name | required | text |  |  |
| family | required | text |  | Darwin Core |
| genus | required | text |  | Darwin Core |
| specificEpithet | required | text |  | Darwin Core |
| scientificName | required | text |  | Darwin Core |
| vernacularName | required | text |  | Darwin Core |
| scientificNameAuthorship | Recognised authority (author and date) for each species name, and source of this information. | required | text |  | Darwin Core |
| date\_modified | Date on which each species entry was last modified. | required | timestamp |  | NACDD |

## Animal measurement

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Lists all the morphological measurements taken for each tagged animal.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Animal measurement ID (unique). | required | numeric | Primary key |  |
| animal\_id | Animal ID. | required | numeric | Foreign key to animal table |  |
| type | Type of measurement (*e.g.* length, weight, total length, carapace length, carapace width, fork length, width). | required | text | Need controlled vocabulary |  |
| unit | Unit of measurement (*e.g.* mm, cm, m, g, kg). | required | text | Need controlled vocabulary |  |
| value | Measurement value. | required | text |  |  |
| estimate | Is the measurement value an estimate? | required | boolean |  |  |

## Project

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information on tagging projects, which data center hosts data, who is the principal investigator.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Project ID (unique). | required | numeric | Primary key |  |
| datacenter\_id | ID of data center. | required | numeric | Foreign key to datacenter table |  |
| title | Project name | required | text |  | NACDD |
| summary | Description of project including background, methods, and objectives. | required | text |  | NACDD |
| citation | Citation to be used in publications using the data from the project should follow the following format: “ProjectName. [year-of-data-download], [Title], [Data access URL], accessed [date-of-access]”. | required | text |  |  |
| infourl | URL to project information website or metadata record. | optional | text |  | ERDDAP |
| publications | Published or web-based references that describe the data or methods used to produce the data from the project. Multiple references should be separated with a semicolon. If available DOIs should be given. | optional | text |  |  |
| doi | Digital Object Identifier (DOI) for the project. | optional | text |  | IDF |
| license | Describe the project restrictions to data access and distribution. | optional | text |  | NACDD |
| distribution\_statement | Statement describing data distribution policy (*e.g.* ‘You accept all risks and responsibility for losses, damages, costs and other consequences resulting directly or indirectly from using the data from this project’). | optional | text |  |  |
| date\_modified | Date on which the project data was last modified. | required | timestamp |  | NACDD |
| location | Min and max longitude and latitude of the project data, transformed into a geometry (polygon). | required | geometry |  |  |
| timestamp\_start | Start date and time (UTC) of the project data. | required | timestamp |  |  |
| timestamp\_end | End date and time (UTC) of the project data. | required | timestamp |  |  |

## Data center

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides name and attributes of each data center.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Data center ID (unique). | required | numeric | Primary key |  |
| manager\_id | ID of manager of data center. | required | numeric | Foreign key to users table |  |
| title | Data center name | required | text |  | NACDD |
| summary | Description of data center including background, methods, and objectives. | required | text |  | NACDD |
| citation | Citation to be used in publications using the data from the data center should follow the following format: “DataCenterName. [year-of-data-download], [Title], [Data access URL], accessed [date-of-access]”. | required | text |  |  |
| infourl | URL to data center information website or metadata record. | required | text |  | ERDDAP |
| doi | Digital Object Identifier (DOI) for the data center. | optional | text |  | IDF |
| license | Describe the data center restrictions to data access and distribution. | optional | text |  | NACDD |
| distribution\_statement | Statement describing data distribution policy (*e.g.* ‘You accept all risks and responsibility for losses, damages, costs and other consequences resulting directly or indirectly from using the data from this project’). | optional | text |  |  |
| date\_modified | Date on which the data center data was last modified. | required | timestamp |  | NACDD |

## Users

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides names and contact details of people involved in tagging projects or data centers.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | User ID. | required | numeric | Primary key |  |
| organisation\_id | ID of organisation each user belongs to. | required | numeric | Foreign key to organisation table. |  |
| name | Name of user | required | text | Should be unique. |  |
| email\_address | Email address of user | required | text |  |
| phone\_number | Phone number of user, including country and area code. | optional | text |  |  |

## Project role

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides mapping between the project and users table.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Project role ID. | required | numeric | Primary key |  |
| user\_id | User ID. | required | numeric | Foreign key to users table. |  |
| project\_id | Project ID. | required | numeric | Foreign key to project table. |  |
| role\_type | Role of user in project (*e.g.* Principal Investigator, Co-Investigator, Administrator, Student). | required | text | Need controlled vocabulary |  |

## Organisation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Provides information about institutions users belong to.** | | | | | |
|  | | | | | |
| *Field name* | *Description* | *Required* | *Data type* | *Constraints* | *Authority* |
| id | Organisation ID. | required | numeric | Primary key |  |
| name | Name of organisation | required | text |  |  |
| department | Department name within organisation | required | text |  |  |
| phone\_number | Phone number of organisation, including country and area code. | required | text |  |  |
| postal\_address | Postal address of organisation | required | text |  |  |

# 5. References

AATAMS (2015) The Australian Animal Tracking and Monitoring System (AATAMS) National Database Web Application. Accessed 11 November 2015. https://aatams.emii.org.au/aatams

Block BA, Jonsen ID, Jorgensen SJ, Winship AJ, Shaffer SA, Bograd SJ, Hazen EL, Foley DG, Breed GA, Harrison AL, Ganong JE, Swithenbank A, Castleton M, Dewar H, Mate BR, Shillinger GL, Schaefer KM, Benson SR, Weise MJ, Henry RW, Costa DP (2011) Tracking apex marine predator movements in a dynamic ocean. Nature 475:86-90

Campbell HA, Beyer HL, Dennis TE, Dwyer RG, Forester JD, Fukuda Y, Lynch C, Hindell MA, Menke N, Morales JM, Richardson C, Rodgers E, Taylor G, Watts ME, Westcott DA (2015) Finding our way: On the sharing and reuse of animal telemetry data in Australasia. Science of The Total Environment 534:79-84

Campbell HA, Urbano F, Davidson S, Dettki H, Cagnacci F (2016) A plea for standards in reporting data collected by animal-borne electronic devices. Animal Biotelemetry 4:1-4

Dwyer RG, Brooking C, Brimblecombe W, Campbell HA, Hunter J, Watts M, Franklin CE (2015) An open Web-based system for the analysis and sharing of animal tracking data. Animal Biotelemetry 3:1-11

Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, Harcourt RG, Holland KN, Iverson SJ, Kocik JF, Mills Flemming JE, Whoriskey FG (2015) Aquatic animal telemetry: A panoramic window into the underwater world. Science 348

Kays R, Crofoot MC, Jetz W, Wikelski M (2015) Terrestrial animal tracking as an eye on life and planet. Science 348

Kranstauber B, Cameron A, Weinzerl R, Fountain T, Tilak S, Wikelski M, Kays R (2011) The Movebank data model for animal tracking. Environmental Modelling & Software 26:834-835

Payne J, Moustahfid H, Mayorga E, Branton R, Mihoff M, Bajona L (2013) A metadata convention for animal acoustic telemetry data. [http://ioostech.googlecode.com/files/AAT Metadata Convention v1.2.pdf](http://ioostech.googlecode.com/files/AAT%20Metadata%20Convention%20v1.2.pdf)