Want to Know About Fish Aggregations Globally? We Have a Web Database!

¿Quieres Saber sobre las Agregaciones de Peces a Nivel Mundial? Tenemos Una Base de Datos Web!

Vous Voulez Connaître les Agrégations de Poissons au Niveau Mondial? Nous Avons Une Base de Données Web!

MARTIN RUSSELL[1] and LEATH MULLER[2]
[1]*Science and Conservation of Fish Aggregations (SCRFA), 215/1000 Ann St, Brisbane, Australia, 4006.* martinrussell99@gmail.org. [2]*Townsville, Queensland, Australia.* leathal73@gmail.com.

ABSTRACT

The status of most of the world's fish aggregations is unknown, with many aggregations decreasing, and some are actually gone. A global web-based database on fish aggregations has been developed by *Science and Conservation of Fish Aggregations* (SCRFA) as a comprehensive tool to assist in understanding fish aggregations. This is the first database of its kind, containing information on fish aggregations, including biological and habitat characteristics, current status, management, traditional knowledge and exploitation history. Information is entered using parameter menus developed to keep a consistent dataset for analysis. The database currently has 888 records from 52 countries, and includes 209 species of fish from 44 families.

KEY WORDS: Aggregation, database, spawning, fish, reproduction

INTRODUCTION

Many fishes form aggregations for reproduction or feeding, and for most fishes, spawning aggregations are unique opportunities to reproduce (Sadovy de Mitcheson and Colin 2012). Unmanaged fishing on aggregations, especially spawning aggregations, can rapidly deplete fish populations, impacting on the ecosystem and the livelihoods of those who depend on the fish.

In 2001, the then *Society for the Conservation of Reef Fish Aggregations* (SCRFA) identified the need to compile and provide to the global marine scientific and management community, information on the locations and attributes of reef fish spawning aggregations throughout the tropical reefs of the world. SCRFA developed a comprehensive web-based database that is user-friendly, accessible to the general public, and allows for entry of new data and information retrieval for fine-grained analysis. A recent upgrade of the database design has increased the number of records and the functionality of the system. MySQL is used as the backend database with all information retrieved and updated in the database through web services that are implemented on the server using PHP scripting language.

SCRFA designed the database to be comprehensive with standard parameter terminology, easily added to online, and importantly, provide for ease of data retrieval. Several design iterations took place over its development, and further updates will occur as more information is added and utility is recognised.

To facilitate the key functional specifications, the database was designed using modern development paradigms to support easy data entry and retrieval. The first step in designing the database was to produce a schema that ensured all key information to be recorded allowed for comprehensive data retrieval for fine grained analysis. For example, the database output needed to at least answer questions such as:

i) How many species from which families aggregate to spawn, and what is the evidence that they form spawning aggregations?
ii) What are the relative distributions of known spawning aggregations in the Atlantic, Caribbean and Indo-Pacific regions?
iii) What is the timing in the lunar cycle that aggregations occur?
iv) What proportion of aggregations are managed?

There were several important factors considered in the design specifications of the database including ease of use, data security, restrictions to information stored in the database based on permissions, and standardisation and the validation of all data entry. For this, a two-phase approach was taken. Phase one was the critical stage of the project, delivering the building blocks for further developments, and Phase two involved fine tuning, date entry and further web-based integration.

The database was initially prototyped and developed in *Microsoft Access*. This phase involved:

i) Development of the overall user interface layout and design,
ii) Conversion and consolidation of existing data,
iii) Implementation of a tiered permission system with four levels of user access allowing a range of data entry and retrieval options,

iv) Design and implementation of the data entry forms and standardised reports,
v) Testing and debugged the system with sample data, and
vi) Production of database documentation including a User Manual and Developer documentation.

MySQL became the chosen database backend, utilising PHP for all server-side scripting functionality including form development and database access. Initial versions of the online database used web forms managed on the server for all data entry, validation, and updates to the database. The online system also supported updates to the database that could be imported from the Microsoft Access based front end, facilitating a centralised repository for consolidation of all entered data.

Between Phase one and Phase two, an analysis of the database was done to provide initial statistics of aggregations globally and to help determine functionality needs, (Cornish 2005). In Phase two, we converted the database systems to a larger scale; and also converted forms and reports to a web-based interface. This phase facilitated data entry directly into, and retrieval from a central repository. During and after this phase, over 300 more records were entered, increasing the database to the now 888 records.

RESULTS

At the time of writing, 888 records of fish aggregations have been entered into the database. Each record is for a single species of fish at a single aggregation site. Therefore, an aggregation site may be represented more than once in the database if there is more than one species using the site.

The records are from 52 countries, with the majority (over 75%) of records from 12 countries (Figure 1). The first published report on reef fish spawning aggregations, of which we are aware, is from Cuba (Russell et al. 2012). Forty-four families are represented, with over half (69%) of records from two families, Serranidae and Lutjanidae, (Figure 2). This is a substantial increase since reported in Cornish (2005), when we had 18 families represented. 209 species of fish are represented, with 11 species representing half the records (Figure 3), of these two species, *Epinephelus striatus* and *Plectropomus areolatus,* are the most common in the database. Of those records that have information on habitat type, solar and lunar phases, most aggregations occur at reef promontories, reef passes, outer reef slopes or drop-offs, on the full moon at dusk.

For each record in the database, we have entered an indication of status of the aggregation. Four categories are provided as options in a drop down menu: unknown, decreasing, same, gone, and increasing. The status of most of the 888 records is unknown (52%), with many aggregation sites decreasing (25%), and some are actually gone (4%). There are also a few aggregation sites increasing (4%), (Figure 4).

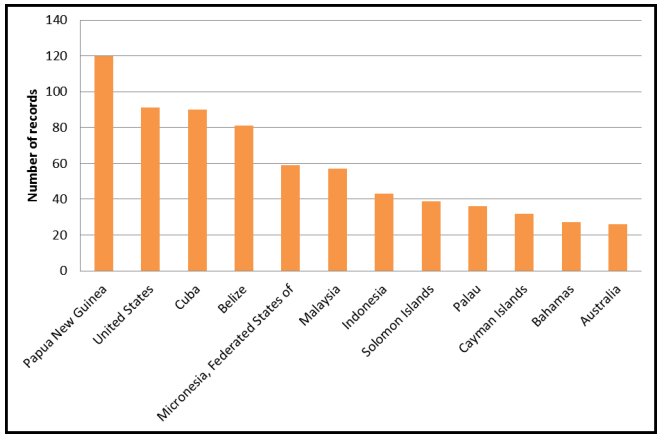

Figure 1. Countries recorded in the database. 75% of the records in the database are from 12 countries.

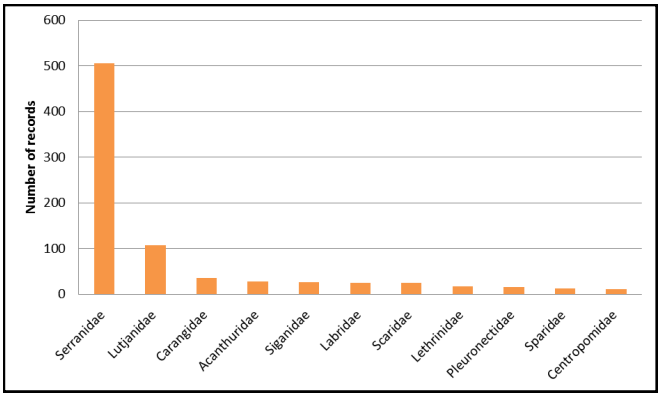

Figure 2. Families recorded in the database. 50% of the records in the database are for two families, Serranidae and Lutjanidae.

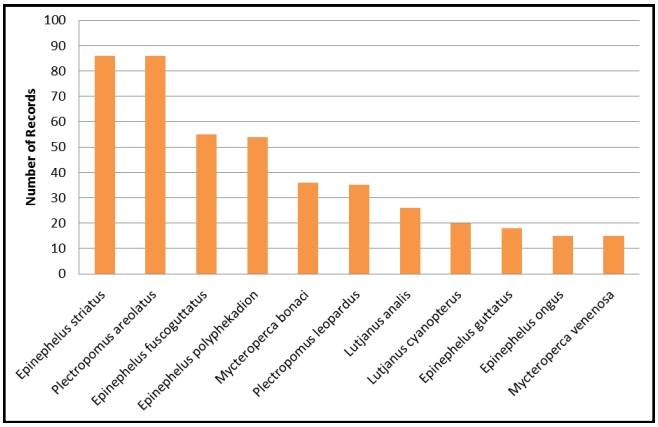

Figure 3. Species recorded in the database. 50% of the records in the database are for 11 species, with *Epinephelus striatus* and *Plectropomus areolatus* being the most common species.

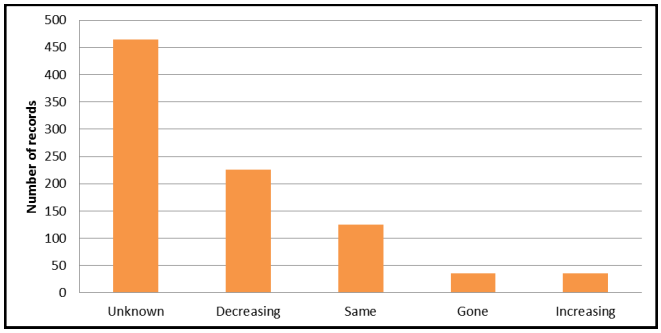

Figure 4. Status of the aggregations recorded in the database.

DISCUSSION

The *SCRFA Fish Aggregation Database* is a comprehensive online database which contains information on biological and habitat characteristics, current status, management, traditional knowledge and exploitation history of fish aggregations. It has been developed by SCRFA since 2002 to assist in understanding fish aggregations globally and is a key initiative undertaken to document known information about aggregations (Sadovy de Mitcheson et al. 2008).

Designing the database was an adaptive process, building a new concept reference tool not available in any other similar context or functionality. The initial design was born through discussions by SCRFA scientists about the need to better understand aggregations on a global scale and compare species, habitats, timings, etc., and noting ideas on a flip chart to outline the need and basic outputs required of such a database.

The database now has a standardised data entry process where information is entered via drop down and multiple selection menus wherever possible to reduce data entry errors. These menu parameters have been carefully developed to ensure data consistency that allows comparable analysis. Some of these parameters are mandatory, and a record will not be accepted for inclusion in the database unless certain information is provided. Some information is also entered using free text. Newly entered records or updated records undergo quality assurance checks (authentication) before being made available on the publicly available search.

The database can be searched anonymously without having a login; however a login is required to add or amend data. This ensures the integrity of the existing records and sets up a clear process for update. Another important design feature is that the precise locations of aggregation sites are not made publically available to protect them from further exploitation. The SCRFA Database Administrators considers requests on a case-by-case basis, and arrange for a data sharing agreement should such information be needed for specific conservation or management initiatives.

The database comprises four levels of access permissions. The first level of access is "*Public Access*", where a limited subset of information can be viewed in table format in a web browser, downloaded for viewing in a spreadsheet, or viewed with Google Earth, and does not include specific details of the aggregation locations, and precise timing and habitat characteristics. The next three access levels require log-in permission. The "*User*" access level includes data entry rights, and the ability to view some additional information not available in the *Public Access*. The "*Super User*" access level includes data entry rights, and allows access to all information. The "*Administrator"* access level manages all aspects of the database, including modifying parameters, validation of new or updated records, and managing user access. This tiered user access approach allows for controlled use of the information, and ensures integrity.

The database utilises the MySQL database engine to store information. To ensure accuracy and integrity of information entered, the database primarily uses lookup tables. For example, each aggregation is associated to a single species that is selected through a pull-down menu on the data entry form. In addition to one-to-one correlations, the database uses a variety of many-to-many relationships to capture data, such as information about direct and indirect spawning evidence for each aggregation. All entries in the database are checked by an Administrator before being accepted in the *verified* list of aggregations. This ensures that SCRFA can maintain data integrity in the database and that it has a high level of accuracy when generating reports and statistics. The database stores information for each aggregation in three core tables:

i) Spawning details, including the species, physical information (such current flow and tidal phases), and temporal information including the time of year and moon phase.
ii) Site information, including the geographical location and site area.
iii) Exploitation and protection information, such as when exploitation first occurred at the site, and which bodies monitor and protect the site.

To ensure high quality and accuracy of data, records are initially entered by a registered user and flagged as "unverified". These unverified entries must be verified by an Administrator before being included in the online search facility.

One of the key functional specifications of the database design was to allow editing of previously verified information in circumstances where additionally refined information with greater accuracy became available. However, the original verified aggregation needed to be maintained for existing outputs such as the online search facility, until the changes to the record were approved by an Administrator. The database subsequently needed to account for potential cancellation of changes or non-acceptance of proposed changes to the unverified record.

To support the requirement of having records in two possible states, "verified" and "unverified", the database uses copies of all tables that store key information in both the verified and unverified states. For example, the "spawning table" stores all information that is verified, while a temporary version of the table stores all unverified entries. This ensures data integrity of verified information.

The initial version of the database used server side forms for all data entry, validation and updates. The current version of the database performs all information retrieval and updates to and from the database through custom web services that are implemented server side using the latest version of the PHP scripting language. Using web services has advantages including:

i) Lower cost of communication overhead with the server, thus supporting low bandwidth and unreliable connections to the Internet,
ii) Increased security for database access through abstraction of the database backend to client access requests and protection from script based injection and other malicious forms of attacks,
iii) Support for other client applications such as dedicated smart phone or tablet apps, and
iv) Easier maintainability of the system that allows changes to the backend database while having minimal impact on clients.

Architecturally, backend access to the database is based on the model view, view model (MVVM) architectural pattern, implemented as PHP classes that query and update the database, and provide backend business logic where required. This allows the web services to provide an abstraction layer to the database's information, retrieving information using the appropriate MVVM class when required. Web services are provided through three main web service interfaces: retrieval, update, and verification. The information returned from the retrieval web services are restricted based on the current user's access level. *Public Access* users can retrieve a limited subset of data, while users that are correctly authenticated and logged in can retrieve additional information such as aggregation details they have entered. *Super Users* and *Administrators* can generally retrieve all information from the database, while *Administrators* can perform all database maintenance such as adding and removing parameters to data entry forms.

The update services allow posting of information to the database when entering and modifying entries to the database, and requires a minimum level of authorisation, while verification services require the highest level of authorisation by an *Administrator*. The verification services allow *Administrators* to move unverified aggregations to the verified aggregations area, deleting existing database entries (both unverified and verified), or allow verification of other lookup information such as new species or literature reference details.

All data entry and retrieval is performed through a web-based form on a user's client computer, using any modern web browser. The client side application relies heavily on JavaScript manipulation of data returned from web services, and utilises jQuery extensively to perform asynchronous access to the database. The client side application makes calls to the database through the web services, and updates the form based on data returned through those services. All data returned from the web services is either formatted as HTML (such as option statements for pull downs), or as JavaScript Object Notation (JSON) objects that are formatted and processed by the client.

By utilising asynchronous calls to web services, we have provided a level of robustness for data entry that would otherwise be difficult to implement. For example, entry often occurs on network connections that are quite unreliable and slow, such as over older GSM phone networks in remote locations. To amortise possible data loss, we send small updates at predefined intervals to the server to save any changes to a record being edited, while not impacting on the user interface (the user does not notice the background save occurring). Additionally, if information being retrieved for the client is interrupted (such as filling the species pull down menu), the client can ask the server for the information again for that individual pull down, rather than requesting all form data from the server. This reduces the load on the network and provides additional robustness for users in the field.

Records are entered into the database using a single data entry screen with dropdown or multiple check box menus and free text fields. As more records are being entered, these menus have expanded to accommodate the diversity in aggregation site information, but also kept to a minimum to ensure comparable and meaningful analysis. Appendix A provides a description of the suite of parameters in the database.

Since the recent functionality upgrade, there have been several hundred records entered in the database, with 888 records now available for analysis. The 52 countries represented indicate the broad range of areas fish aggregate, however, given that most of the records are from just 12 countries (Figure 1), there would seem to be scope for further information gathering. Similarly, the forty-four families represented are dominated by just two families, Serranidae and Lutjanidae, (Figure 2), and out of the 209 species of fish, 11 species represent half the records (Figure 3). As more information is gathered these statistics may change, or further confirm species preferences for aggregating.

An important aspect of the data is to be able to look at trends or consistent habits of certain species. Although not all records in the database have information on habitat, solar and lunar phase, of those records that do have this information, most aggregations occur at reef promontories, reef passes, outer reef slopes or drop-offs, on the full moon

at dusk. As we learn more about this, it seems more questions become apparent. Importantly, we are concerned about the status of the aggregations we know of. Given that for over half of the records the status is unknown, and a quarter are decreasing, (Figure 4), there is an urgent need to further investigate aggregations sites, and also hasten to obtain information on sites that have yet to be entered into the database. These sites may be documented in grey literature, or have yet to be described. As exploitation of marine resources expands globally, the remaining aggregation sites we currently don't hold information about might be in grave danger of overexploitation.

CONCLUSION

The *SCRFA Fish Aggregation Database* is unique and provides the only comprehensive data source on fish aggregations on a global scale, and thus a powerful reference tool for marine scientists and managers. The database has been developed with limited resources, relying mostly on voluntary work. Without this input, the cost of developing such a database via normal commercial avenues would have made this project virtually unachievable.

Clear trends have become apparent since compiling information into the database. Many aggregations have a declining status, but also most of the records have an unknown status. This is a concern and effort should be made to fill this information gap, and also attempt to turn the decline around to at least a stable, or preferably, increasing status.

LITERATURE CITED

Cornish, A.S. 2005. *Development and Summary of Global Spawning Aggregation Database*. Society for the Conservation of Reef Fish Spawning Aggregations. www.scrfa.org. pp 16.

Russell, M.W., B.E. Luckhurst, and K.C. Lindeman. 2012. Management of spawning aggregations. Pages 371-404 in: Y. Sadovy de Mitcheson and P.L. Colin (eds.). *Reef Fish Spawning Aggregations: Biology, Research and Management*. Fish & Fisheries Series 35. Springer Science & Business Media, New York, New York USA.

Sadovy de Mitcheson, Y. and P.L. Colin. (eds.). 2012. *Reef Fish Spawning Aggregations: Biology, Research and Management*. Fish & Fisheries Series 35. Springer Science & Business Media, New York, New York USA.

Sadovy de Mitcheson Y, A. Cornish, M. Domeier, P.L. Colin, M. Russell, and K.C. Lindeman. (2008). Reef fish spawning aggregations: a global baseline. *Conservation Biology* **22**(5):1233-1244.

Appendix A. Database parameter descriptions

Parameter	Description
Country	Select one only menu of the nearest mainland country or island group to the aggregation site being entered.
Ocean Name	Select one only menu of the world's oceans
Sea Name	Select one only menu of the world's seas, gulfs, and bays.
Site Description	Free text box for the name and details of the site location, including reef name or other feature identifiable from maps or charts.
Contact Person	Multiple selection menu of contacts already entered into the database. A new contact can also be added.
GPS Location	The precise location of the site forms a valuable part of the database. Specific locations are not released to the public to avoid misuse of this data. The co-ordinates may be entered as decimal degrees (dd.ddddd), degrees/decimal minutes (dd mm.mm) or degrees/minutes/seconds (dd mm ss). The default format is decimal degrees. The default datum is WGS84.
Species	Select one only menu of family, genus, species and local name. A new species can also be added.
Spawning Type	Select one only menu: Unspecified - if actual spawning was not observed; Pair Spawning – if a male and female pair; or Group Spawning - if one or more females and many males spawn.
Aggregation Type	Select one only menu: Resident - individuals travel a short distance from their usual site of residence (typically only a few hours travel) Transient - individuals migrate long distances over a period of days to weeks to a specific aggregation site. Unknown – where it is yet to be determined what aggregation type the species forms at the site.
Published/Unpublished Sources	Multiple selection menu of published or unpublished references. A new reference can also be added.

Appendix A (continued). Database parameter descriptions

Parameter	Description
Evidence	Multiple selection menu: Direct Evidence – hydrated eggs postovulatory spawning spawning observation Indirect Evidence – (A mandatory parameter with or without direct evidence recorded) Colour changes associated with spawning Courtship GSI High seasonal landings Mature gonads Multiple gravid females Significant increase in numbers observed (>3x)
Current Status	Select one only menu: Decreasing Gone Increasing Same/unchanged
Months of spawning	Multiple selection menu
Solar Phase	Multiple selection menu: Dawn Mid Morning Noon Mid Afternoon Dusk Night
Lunar Phase	Multiple selection menu: New Moon First Quarter Full Moon Third Quarter
Visibility	The water visibility at the time of observation, or what is generally known for the site.
Depth	The depth where the greatest density of fish was observed.
Tidal Movement	Select on only menu: Low Incoming High outgoing
Habitat Type	Multiple selection menu of a range of habitat types such as reef, rubble, estuary, and atoll.
Aggregation Area	Select one only menu of range codes to indicate how large the aggregation site is.
Census Count Code	Select one only menu of range codes to indicate the maximum number of fish observed to be aggregating at the site.
Traditional Knowledge Details	Free text box to enter any information available on historic or traditional knowledge, for example when a local community know when to catch a certain fish that aggregates in large numbers when the leaves on a tree change color.
Year Discovered	The first year the site was known to be fished.
Site Discovery Method	Select one only menu of a range of sources, such as fisher or scientist.
Gear Used	Multiple selection menu of a range of fishing gear types, such as spear, line, or net.
Catch per unit effort trend	Select one only menu: Decreasing Increasing No change or only a single year of data available
Average fish size	The average size of fish observed by either standard (SL) or total (TL) length.
Monitoring information	Multiple selection menu of methods used such as interviews.
Management/Protection	Multiple selection menu of a range of management tool such as closures or gear restrictions.
Enforcement Level	Select one only menu of a range of levels from light to strict.