DEPARTMENT OF
SUSTAINABILITY AND ENVIRONMENT

Murray-Darling Freshwater Fish Information System

MD1282 (Phase 2.1)
Report to the Murray-Darling Basin Commission,
Native Fish Strategy

Sharley, D.J., Kitchingman, A. and Cant, B.
2009

Arthur Rylah Institute for Environmental Research

David J Sharley, Adrian Kitchingman and Belinda Cant

Arthur Rylah Institute for Environmental Research
123 Brown Street, Heidelberg, Victoria 3084

June 2009

In partnership with:

Arthur Rylah Institute for Environmental Research
Department of Sustainability and Environment
Heidelberg, Victoria

Report produced by: Arthur Rylah Institute for Environmental Research
Department of Sustainability and Environment
PO Box 137
Heidelberg, Victoria 3084
Phone (03) 9450 8600
Website: www.dse.vic.gov.au/ari

Citation: Sharley, David J., Kitchingman, Adrian and Cant, Belinda. (2009) Murray-Darling Freshwater Fish Information System, Supplement Report to the Murray-Darling Basin Commission, Native Fish Strategy MD1282 (Phase 2.1), Arthur Rylah Institute for Environmental Research, Department of Sustainability and Environment, Heidelberg, Victoria.

Authorised by: Victorian Government, Melbourne

Printed by: Arthur Rylah Institute

Table of Contents

Table of Figures

List of Tables

Murray-Darling Freshwater Fish Information System

Executive Summary

The MDFFIS was initiated because of the need to collate the myriad of datasets on native freshwater fish in the Murray Darling Basin. Phase 2 of the MDFFIS has seen significant progress into the development of the MDFFIS model. Both fish and biophysical meta-data from numerous fish ecology projects conducted by various organisations across the Murray Darling Basin have been added to the data warehouse. Spatial information from some of these projects has also contributed to the spatial data system which will eventually link with the warehouse to form a comprehensive spatial information system. The benefits of research synthesis are only just beginning to be realised and will become more apparent to researchers, managers and policy makers who work in the Murray Darling Basin once the MDFFIS is fully integrated in Phase 3.

The advent of the MDFFIS will significantly support the implementation of the Native Fish Strategy (NFS), and allow targets set out in (MDBA 2004) to be measured. It will greatly improve NFS accountability and help meet reporting obligations to the MDBA and the Australian Government Minister for Climate Change and Water. The inclusion of NFS objectives and recommendations as MDFFIS data fields, will now act to preserve 'corporate memory', a key aim of any organisation, and help both NFS and MDBA gain knowledge from the projects it funds.

1 Introduction

In the environmental sciences, one of the most common methods of collecting data is through the initiation of an environmental monitoring programme within a defined research project. A defined set of variables and processes are measured across space and or time, to answer specific questions (Spellerberg 2005). Within each monitoring programme, it then becomes essential that the data collected is maintained in a manner conducive to real-time re-analysis in response to changes in ecosystems and management context. For instance, formats for data storage, personnel involved with data collection, or technological advances in methodologies may all change over the life of a project, and without active stewardship, the data may become either lost or uninterpretable (Schroder *et al.* 2006). In general, it has been found that information content of data sets tend to decrease over time (Michener and Brunt 2000), therefore, the challenge facing researchers and data managers is how to preserve the value of the data over extended periods of time?

It is also important for individual researchers, managers and policy makers to appreciate the importance of data synthesis, and the benefits that lie within research synthesis (Ford and Ishii 2001; Henry *et al.* 2008; Pickett 1999). Although data collected from individual experiments remain at the core of ecological research, the value of the data increases considerably when it is integrated and synthesized to reveal important patterns not seen in individual datasets (Jones *et al.* 2006). In addition, data synthesis can integrate studies across many disciplines and allows for the data to be used for purposes not initially intended and address questions that were not able to be answered when the data was collected (Jones *et al.* 2006).

One solution is the development of a robust environmental meta-database. The Native Fish Strategy, fish ecologists and river managers, recognised the need to collect, collate and manage freshwater fish research data in the Murray-Darling Basin (Cant, 2007, Kitchingman 2008). In response, the Murray-Darling Freshwater Fish Information System (MDFFIS) was designed to manage both meta-data and data. The MDFFIS is not only designed to catalogue ancillary information, such as format, owner and origin of the data, but also the information required to understand and interpret the data, such as sampling methodologies, spatial and temporal information as well as fish species targeted. This will in the first instance highlight genuine research gaps and will also provide the opportunity to use past research to refine questions and aid the design of future projects.

1.1 MDFFIS development

We proposed to roll out the MDFFIS over three phases of development. Phase 1 involved the scoping and outlining of logistical issues of the system. Phase 2 will be the main development stage where the system is constructed and the data/meta-data collection regime put in place. Phase 3 is ongoing and involves the continued maintenance of the MDFFIS and further development of any new feature required over time.

This report marks the beginning of Phase 2, the development of the MDFFIS. Phase 2.1 covered in this report deals with the development of a 'seed' database of meta-data. The seed database will both act as a pilot to test the process of data collection, while also providing meta-data to begin populating the MDFFIS.

The collection of meta-data has a number of uses and advantages. First, meta-data is easier to obtain than raw data. It is generally not constrained by security, intellectual property and quality control surrounding the transfer of raw data. Second, collecting meta-data allows rapport and trust between the

MDFFIS project team and data custodians to be enhanced, and allowing custodians of the data to take an active role in the development of the database without the issues of data sharing. Being active participants in the MDFFIS development will hopefully lead to a fuller understanding of the benefits of opening "data gates" to other researchers in the future; under the proviso that access is restricted or permission-based where necessary.

2 Methods

To store the collected meta-data, a preliminary meta-data database was developed. The dual aims of the preliminary database were first to store data in a simplistic format to allow a further refined design to be implemented in later stages of phase 2, and secondly, to gather information on the issues that arise during the collection of the meta-data, which is crucial for phase 3 of the MDFFIS.

2.1 Database development

The preliminary meta-data database is a simplified meta-data storage system allowing collected data to be reorganised into a more refined version later in the MDFFIS development. Some elements such as the meta-data descriptors will constitute core components of the final MDFFIS. A number of ecological meta-data descriptors have been proposed as a common platform to categorise ecological information (Michener 2006). A recent and increasingly popular ecological meta-data standard is the Ecological Meta-data Language (EML). Ecological Meta Language is a meta-data specification designed for ecological data being developed through the Knowledge Network for Biocomplexity (KNB) project, based at the University of California. The current version of EML (2.0.1) is extensive, and now relatively stable. It was decided that due to the ecological focus and its current placement as the defacto ecological meta-data standard, EML should form the basis of the meta-data descriptors for this stage of the MDFFIS.

A number of ecologists were consulted on the descriptors that were available in EML and a descriptor set finalised. Importantly, the decision to base the MDFFIS on EML descriptors will allow any new design of the MDFFIS system, using an EML or similar base, to easily import the meta-data held in the preliminary database.

A number of descriptors have been classed as mandatory and appear in red in the MDFFIS meta-database. These descriptors are aimed to capture a minimum, yet useful amount of meta-data, should there be problems collecting all data for some projects. The mandatory descriptors also contain enough information for a third party to complete the full suite of meta-data

for each project or allow further contact with the data custodian to gather the missing information.

The preliminary MDFFIS was built using Microsoft Access, with a form-based user interface to ease data entry and management. A simple search function enables meta-data records to be filtered and retrieved. The user interface consists of two parts. First, there is a list of all records in the database with a search function located above the list (Figure 1). When searching for text across fields, the 'AND' button is selected. If the search text is within one field then the 'OR' button is selected. The second part allows the user to either edit existing records or enter a new record (Figure 2). Mandatory descriptors are coloured red and must be completed to be able to save the record.

2.2 Data collection

Two methods were used to collate information: literature searching and project questionnaires.

1. Literature searching

Literature searching was conducted using primarily journal articles and government agency reports (internal and external). All useful information was extracted and incorporated into the database. Priority was given to information regarding fish research within the Murray-Darling Basin; however fish research that did not primarily reside within the basin were also considered on a case by case basis.

2. Project questionnaires

We thought it likely that many unpublished details or reports not publicly available remained with fish ecologists and research consultants. An attempt to engage with agency professionals in order to gather available information on fish research within the basin was made primarily using phone and email communication. Two methods were used to gather information in this way

1. A questionnaire with the required information was sent to each recipient to fill in and send back.
2. Each recipient was asked to email or send a PDF of all relevant publications corresponding to their research. Information was then extracted using the same methods as the literature search.

Table 1 MDFFIS descriptors (*denotes NFS specific descriptors – bold indicates mandatory descriptors)

Groupings	Descriptors	Description
General	**Title**	Title of project
	Alternate Title	Secondary project identifier
	Abstract Recommendations and Objectives	Brief summary of dataset
	Keywords	Likely search words
	Series	Series, Volume, Sub Project of, etc
	Institution	Institution conducting research
	External ID	NFS or institutional ID number
	Project Type*	Basis of project (i.e. scoping, research)
	NFS Actions*	6 driving actions of the NFS
Contact Info	**Contact**	Contact person
	Address	Custodian address
	Telephone	Custodian telephone number
	Email	Custodian email
	Associated Parties	Other people, organisations associated
History	Progress	Status of the dataset
	Maintenance	Frequency of changes
	Lineage	Source of data
	Published Date	Date resource results published
Control	**Access**	Access restrictions on data
Format	**Format**	Current format of data
	Formats	Other data formats available
	Medium	How to transfer data
	Record No	Estimated number of records in dataset
	URL	storage file path of dataset
Geographic	**Location Description**	Name or description of study location
	Altitude Zone	General altitude coverage
	Coordinate System	Coordinate System in which spatial data recorded
	Accuracy	Positional accuracy
	N-Lat	North latitude
	S-Lat	South latitude
	E-Lon	East latitude
	W-Lon	West latitude
Temporal	**Begin Date**	Start of data collection
	End Date	End of data collection
	Frequency	Frequency of collection
Attributes	**Attributes**	Types of information collected
Taxonomic	**Species**	Species or type of animals sampled
Other	**Comments**	Any other relevant information

Figure 1 Record Listing and Basic Search (first part)

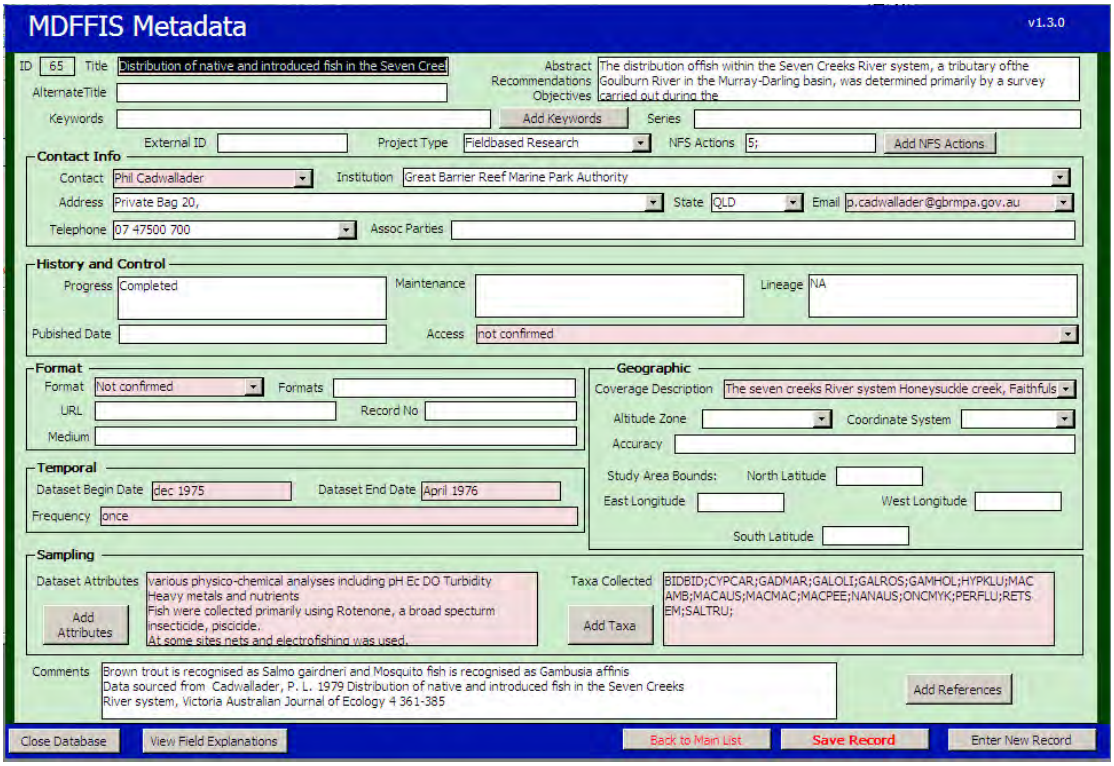

Figure 2 Record edit and entry (second part)

2.3 Spatial Layers

Using ESRI ArcGIS, spatial layers are being developed in parallel to the data warehouse. Using spatial identifiers from each project, a bounding box which gives the spatial extent of each project is projected onto a map. It is then possible to spatially locate projects of interest within the basin. As this will ultimately link to the data-warehouse, a researcher or aquatic manager will be able to search for projects via either the data-warehouse or the spatial database. Links between each database will then allow researchers and aquatic managers to locate projects.

There are currently two methods used to determine the spatial extent of projects:

- Determine the longitude for east and west extents and the latitude for the north and south extents of each project and enter the details into the data warehouse (geographic section) which is then transferred into the spatial database as a feature.
- Using information already known, directly draw bounding box in ArcGIS as a feature.

Each feature which corresponds to the spatial extent of each project is also linked to an attribute table with project information that can be used to search for projects within ArcGIS.

3 Results and Discussion

3.1 Meta-data Collected

3.1.1 Coverage

From the 14th April until 22 May 2009, a total of 93 researchers, database and research managers from 30 organisations across all the Murray Darling Basin States were contacted and asked to contribute to the MDFFIS. Victoria had the highest representation with nine organisations, while South Australia had the lowest with two (Figure 3).

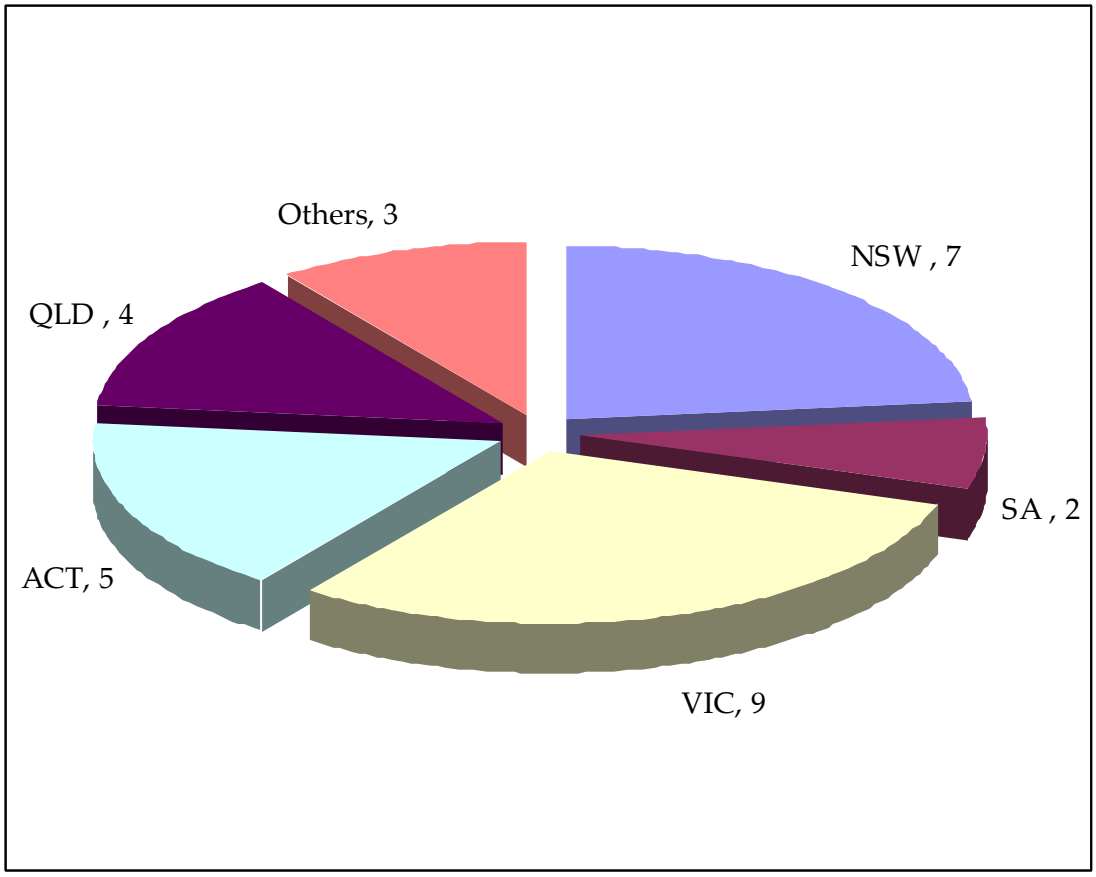

Figure 3 Number of organisations contacted in Phase 2 of MDFFIS development per state and territory

3.1.2 Project entries

There are currently 122 projects entered into the MDFFIS, ranging from field-based research experiments to scoping studies and literature reviews (Figure 4). It must be noted that it was not until partly through Phase 2 that it was decided to include desktop studies and literature reviews into the MDFFIS. Before that decision, only field-based or laboratory based studies with data-sets attached were applicable for the MDFFIS. While this change has increased the breadth of the MDFFIS, for effective data integration, synthesis and meta-analysis, there needs to be a clear delineation between projects that have collected unique datasets and those that are not associated with data. Nineteen organisations are currently represented in the database, which equates to 63% of the organisations initially contacted to supply data. A small number of organisations dominate the project count with the Arthur Rylah Institute (DSE - Vic), the Department of Primary Industries (NSW) and SARDI Aquatic Sciences (SA) contributing more than 79% of the current entries (Figure 5). This result probably reflects individual organisation involvement or ease of finding project information for the MDFFIS, rather than the amount of research conducted by individual organisations. For example, the Murray

Darling Freshwater Research Centre is ranked 8th in respect to project entries, probably not reflecting the output of this active research group.

3.1.3 Overall Completeness

In addition to the 122 projects currently entered in the MDFFIS, there were a further 43 projects that did not have the minimum required information for it to be useful in the database. However the majority of incomplete projects stem from only two organisations, making retrieval of this data relatively easy in the future. It's expected that as more projects are entered into the MDFFIS, the discrepancy of project entries between organisations will narrow.

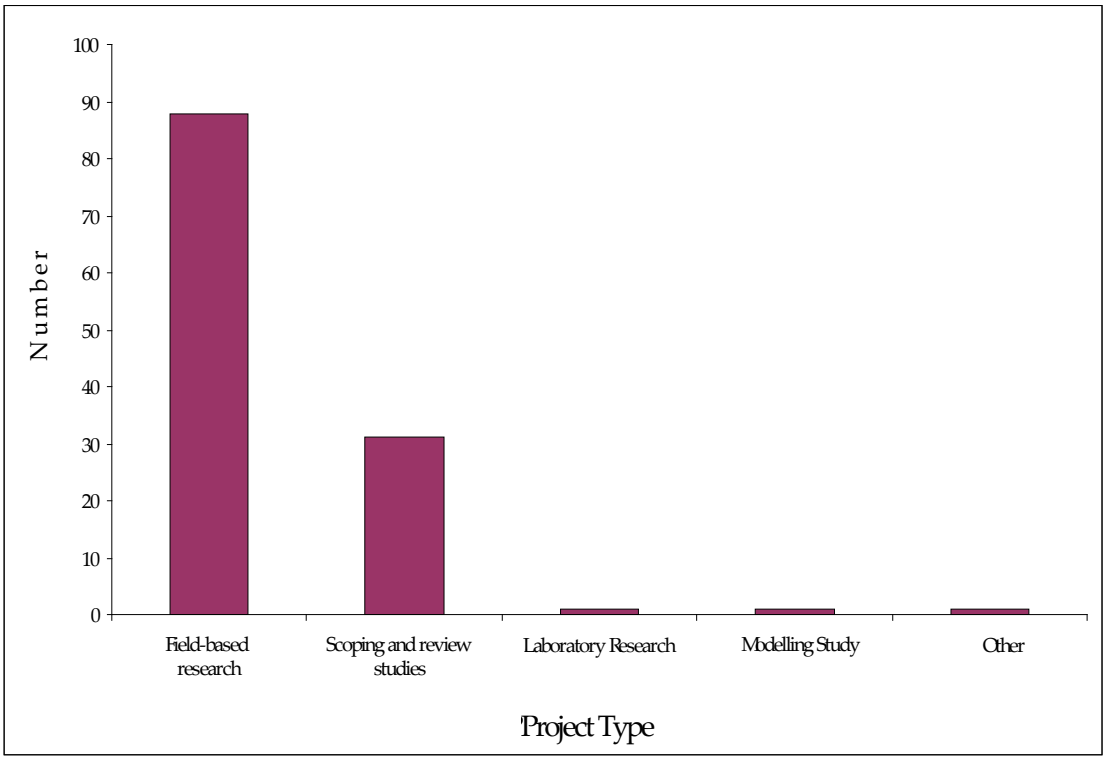

Figure 4 Number of projects in the MDFFIS grouped according to project type.

3.1.4 Spatial analysis

Within the Murray Darling Basin (Figure 6), there are currently 63 projects represented in the spatial layers, which equates to ≅ 52% of projects in the MDFFIS and 72% of field-based research projects. The geographic extents of the projects ranged from small (local river reach or wetland complex) to very large (basin wide) (Figure 7). Although further analysis is necessary, some spatial patterning in project extents was found with a number of projects hotspots spread throughout the basin (Figures 8-10). "Project Hotspots" for the purposes of this report are defined as regions where there is sufficient overlap of projects carried out in within the same spatial coverage.

Hotspot 1 was located within the vicinity of the Lower lakes and Coorong National Park (Figure 8). Projects were primarily carried out by SARDI Aquatic Sciences, with two core areas of research; 1 fishway assessment and 2, fish surveys of the lower lakes and Coorong region. The most common variables measured were; 1) Fish abundance, 2) length/frequency distribution, 3) species composition, and 4) Spawning and recruitment. In relation to the Native Fish Strategies driving actions (MDBA 2004), the majority of projects came under both NFS actions 2 (protecting fish habitat) and 5 (protecting protected native fish species), indicating that the sustainability of viable native fish species into the future is of high concern to aquatic managers within the Lower Lakes and Coorong region.

Hotspot 2 was located in the upper and lower sub-basins of the Murray River (Figure 9). Projects were primarily shared among three research groups, ARI, DPI NSW and SARDI Aquatic Sciences. Factors that were of most interest in this hotspot was what affect flow rate and water manipulation had on populations of all native fish, in particular *Maccullochella peelii peelii* Mitchell (murray cod) and *Macquaria ambigua ambigua* Richardson (golden perch). Research also focused on ability of native fish species to disperse and migrate through fish passages along the Murray River. Variables most measured were length/frequency, species composition, species abundance and recruitment. Native Fish Strategy actions 2 and 5 (MDBA 2004) were the most common.

Hotspot 3 was located within the northern Victorian catchments (Figure 10). The Arthur Rylah Institute conducted the majority of projects with a variety of research areas; affect of flow regimes on native fish, barrier affect on the migration and movement of native fish; the influence of habitat and refugia on native fish populations and the control of alien fish species, mainly *Cyprinus carpio* Linnaeus (common carp), encompassing NFS actions 2, 4 (controlling alien fish species) and 5 (MDBA 2004). Variables most measured included recruitment, movement and dispersal, flow and spawning ability. Although all fish were monitored, three species (*M. peelii peelii*, *Maccullochella macquariensis* Cuvier (trout cod) and *Galaxias olidus* Günther (mountain galaxias) were the most common fish studied.

The linkage between the spatial database and the data warehouse offers a unique opportunity to reveal hidden patterns in fish ecology research and connect project objectives with geographic locations.