This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

Deliverable No. 6.1

Project acronym:
FarFish

Project title:
Responsive Results-Based Management and capacity building for EU Sustainable Fisheries Partnership Agreement- and international waters

Grant agreement No: **727891**
Project co-funded by the European Commission within the Horizon2020 Research and innovation programme

Start date of project: **1st June 2017**
Duration: **48 months**

Due date of deliverable:	30/09/2017
Submission date:	30/09/2017
File Name:	FarFish D6.1_Data input protocol for the FFDB
Revision number:	01
Document status:	Final[1]
Dissemination Level:	PU[2]

Revision Control

Role	Name	Organisation	Date	File suffix[3]
Author/ Task Leaders	Jamie Lentin	STL	29/09/2017	JL
Author	Þormóður Dagsson	MATIS	29/09/2017	ÞD
Authors	Margarita María Rincón Hidalgo Javier Ruiz	CSIC	29/09/2017	MMRH
WP leader	Javier Ruiz	CSIC	29/09/2017	JR
Coordinator	Jónas R. Viðarsson	MATIS	30/09/2017	JRV
Administrative Coordinator	Oddur M. Gunnarsson	MATIS	30/09/2017	OMG

[1] Document will be a draft until it was approved by the coordinator
[2] PU: Public, PP: Restricted to other programme participants (including the Commission Services), RE: Restricted to a group specified by the consortium (including the Commission Services), CO: Confidential, only for members of the consortium (including the Commission Services)
[3] The initials of the revising individual in capital letters

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

Deliverable D6.1

Data Input Protocol for the FFDB

30/09/2017

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

Executive Summary

This document contains the initial data input protocol for the FarFish Database (FFDB) where much of the data collected within the project will be stored and made available for the project partners, as well as beyond the project. The first section re-states and clarifies the aims of Task 6.1 (Development of the FarFish database); that it is a collection of components whose aim is to collate data from partners and integrate FarFish outputs with the FarFish website. Secondly, we detail the likely nature of these components, and finally we discuss the infrastructure and architecture required to support these tools.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

1 Introduction

Task 6.1, the FarFish DataBase, doesn't have any one particular problem to solve or services to provide; instead there are multiple areas in which data-handling expertise are required. As a result, there will be a number of outputs as part of FFDB, which will focus on one of the following themes:

- Data-handling; in particular;
  - Data collection, archival (for manually collected data, for example)
  - Data transformation (for modelling, for example)
- Making FarFish outputs (visualisation tools, results) available on the web.
- Integrating all FFDB outputs with the main project website at http://farfish.eu

There will, however, be some commonality across all the tools developed. Infrastructure will be shared between all tools and with the FarFish website. This means that all FarFish output will be gathered in one place, which makes it easier for interested parties to find, and also means that it will be available for the lifetime of the FarFish website (as opposed to relying on third party systems that may well disappear before the end of the site).
All bespoke development will be open sourced and available on GitHub at https://github.com/farfish. During FarFish this will provide a place to collaborate on work with other partners, so all can work on the latest versions of the code and/or feedback on problems. It will also disseminate FarFish activities to a wider audience, and make code developed in FarFish accessible for reproducing results and/or continuing development.

The following goes into more detail on likely components developed under the FFDB banner. At this early stage of the project the nature of tools required is quite abstract. The precise components developed may well change as the collaboration continues, to adjust to the needs of FarFish.

2 Website data input form / storage

For situations which involve manually collected data (versus being downloaded from an institutional database, for example), we can offer a Javascript web application that allows a user to:

- Drag and drop a spreadsheet file
- The page will make sure appropriate columns are defined
- The page will also highlight any likely errors, e.g. values outside bounds
- Only if all data is valid, data will be ingested into a Postgresql Database.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

This will be used to collect data from each case study to then be able to apply analytical methods. They will be provided an excel template to fill in, like the sample below; which can then be uploaded into the FFDB system:

	A	B	C	D	E	F
1	Name	Test				
2	Year	1998	1999	2000	2001	2002
3	Catch	43457	33899	15.28	4941	48827
4	Abundance index	2211	1911	1955	1725	1.56
5	Duration t	13				
6	Average catch over time t	23.03074512				
7	Depletion over time t	0.180710371				
8	M	0.13				
9	FMSY/M	0.5				
10	BMSY/B0	0.35				
11	MSY	NA				
12	BMSY	NA				
13	Length at 50% maturity	84.13				
14	Lenght at 95% maturity	110.03				
15	Length at first capture	15.21				
16	Length at full selection	60.21				
17	CAA 2008	0	1	20	74	70
18	CAA 2009	0	0	19	74	66
19	CAA 2010	0	0	23	67	59
20	Current stock depletion	297				
21	Current stock abundance	66465				

Then an R package will make this data available for other parts of WP6 for analysis, and use in visualisation tools on the FarFish website. This means the visualisations can be updated without having to change the code.

A list of the currently required rows can be found in Appendix A; however, these will be subject to revision once it is decided which methodologies will be utilised.

3 Data transformation for modelling

Some of the modelling frameworks have large data requirements, for example GADGET. If modelling activity is required, then given sources of data (e.g. institutional databases), FFDB can use tools to collate and transform data to produce model data input, for example the MFDB tools developed as part of MareFrame.

4 Hosting / back end for visualisation tools

Several tasks in WP6 will be producing visualisations, for example by using the R Shiny toolkit. If the partners creating these tools upload them to the FarFish GitHub site, the code can be taken and then integrated with the FarFish website. This will mean they are easy to find, both for other partners and outside FarFish.

The data input form could also be used to allow others to then update the data that the visualisation tool is using automatically.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

5 Website data archival to meet DMP requirements

Data sets, e.g. spreadsheets, will need to be published on the website as part of the data management plan (DMP). Wordpress will be used to ensure these are available for other users and search-able through the website.

6 Infrastructure

The FarFish website and all FFDB components will be developed to fit into the following infrastructure: Either we are using off-the-shelf open source tools, for example NGINX, Wordpress, MySQL, PostgreSQL, or we are using programming languages with big open source communities, for example Javascript, Python and R. This allows us to develop quickly by re-using existing tools and components whilst still being able to distribute project output freely without being restricted by licencing agreements from software companies.

Which parts will be relevant depends on the application. For example, the back-end for visualisation tools will be provided by the Python data API, the visualisation tools themselves will be hosted as Shiny applications.

Regardless, everything will be accessible via. https://www.farfish.eu

7 Conclusion

At this stage, we have restated aims for this task, and made requirements more concrete. The next step will be to start making prototypes of the components suggested in this document, so they can be trialled by FarFish partners and feedback received.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

Appendix A: Indicative case study data requirements

The spreadsheet template given to case-studies will contain the following headings. The precise list will vary as the project progresses.

- **Name**:
- **Year**:
- **Catch**: Annual catches in weight (landings plus dead discards)
- **Abundance index**: Relative abundance index (e.g. standardized Catch Per Unit Effort (CPUE), acoustic survey)
- **Duration t**: Length of catch time series (in years)
- **Average catch over time t**: Average catch over time t
- **Depletion over time t**: Depletion over time t
- **M**: Natural mortality estimate
- **FMSY/M**: The ratio of FMSY to natural mortality rate (typically in the range 0.3 - 1.5)
- **BMSY/B0**: The depletion level corresponding to the most productive stock size (BMSY)
- **MSY**: Maximum Sustainable Yield estimation
- **BMSY**: Biomass at MSY
- **Length at 50% maturity**:
- **Lenght at 95% maturity**:
- **Length at first capture**:
- **Length at full selection**:
- **CAA 2008**: Catch at age in a given year (for this example 2008)
- **Current stock depletion**:
- **Current stock abundance**:
- **Von Bertalanffy K parameter**:
- **Von Bertalanffy Linf parameter**:
- **Von Bertalanffy t0 parameter**:
- **Length-weight parameter a**:
- **Length-weight parameter b**:
- **Steepness**:
- **Maximum age**:
- **CV Catch**: Imprecision in historical annual catches (CV refers to Coefficient of variation)
- **CV Depletion over time t**:
- **CV Average catch over time t**:
- **CV Abundance index**:
- **CV M**:
- **CV FMSY/M**:
- **CV BMSY/B0**:
- **CV current stock depletion**:
- **CV current stock abundance**:
- **CV von B. K parameter**:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727891

- **CV von B. Linf parameter**:
- **CV von B. t0 parameter**:
- **CV Length at 50% maturity**:
- **CV Length at first capture**:
- **CV Length at full selection**:
- **CV Length-weight parameter a**:
- **CV Length-weight parameter b**:
- **CV Steepness**:
- **Sigma length composition**: standard deviation for length composition data
- **Units**: Catch time series units (usually tonnes)
- **CAL_bins**: The definition (break points) of the length classes
- **CAL 2008**: Catch-at-length data in a given year (frequency of catches in each length class. For this example year=2008)
- **MPrec**: A previous recommendation of a management procedure (e.g. a catch limit in tonnes).
- **LHYear**: Last historical year of the simulation (prior to projections)