

# PROJECT PROPOSAL:

## Development and Implementation of a Fisheries Data Management System

JS

November 5, 2012

### Contents

1	Introduction	1
2	Background	2
2.1	General Context . . . . .	2
3	Project Objectives	24
3.1	Data Input . . . . .	24
3.2	Data Validation . . . . .	26
3.3	Data Transformation . . . . .	27
3.4	Data Output . . . . .	28
4	Aknoweledgments	28
	References	43

## 1 Introduction

The project's longer-term development objective is to contribute to the sustainable management of the fisheries exploiting the living resources of the Ocean in general, in accordance with the Code of Conduct for Sustainable Fisheries and the Ecosystem Approach to Fisheries. The immediate objective is to improve the technological (software/hardware) component of the small-scale fisheries monitoring system. For that purpose, we are going to review the existent fisheries management systems, analysing their strengths and their limitations, and propose a set of solutions, based on the latest technologies on databases and Geographical Information Systems.

## 2 Background

According to [22], "decision making for fisheries policy-making, planning and management relies largely on processed information, not raw data". Often the datasets are so large, that the only way to deal with them is through a Database Management System (DBMS). The DBMS in this context may have many functions, but the most relevant ones are: to store the data and ensure its validity and compatibility; it should also be capable of processing it and outputting it, in different ways.

When designing a complete system, many components should be taken into account; namely, the User Interface (UI), the data entry, the data processing, the reporting and the Geographic Information Systems (GIS) functionalities (that can be involved in many of the previous steps); the documentation is also something to consider, as it ensures the usability of the system, and some of the reviewed systems also tackle the component of knowledge transfer, through tutorials and workshops. It was not found any system that implemented all of these components, but they seem to focus on some of them instead (for instance, data processing). As long as the system focuses on interoperability, by using open standards and free and open source software (FOSS), this is not a problem since we can combine different systems for our purposes, and avoid in this way, the redundancy on development.

In terms of architecture, there are many options available for building such system; ranging from using complete commercial solutions, by tying together different components, to creating a tailor-made application; all of them have different advantages and inconveniences. In this review, we tried to focus mostly on systems that are already in use, rather than in proposals or assessments of "how" systems should be. This is because to be put in place, fisheries management systems may face many obstacles, such as: bureaucracy, costs, etc; their ability to cope with them is also part of their success.

### 2.1 General Context

In this section we describe the results of reviewing 25 fisheries management systems. The criteria for selecting the systems that integrate this review, was their visibility; either because they were referred by some colleague, or because they appeared in the internet - as a result of search queries, or within the site of some well-known institution/organisation.

These were some key aspects that were taken into account, when reviewing the systems:

- security
- engaging and involvement of stakeholders
- validation
- ability to export information
- modular architecture and scalability

- licensing
- ability to work with incomplete data
- status (completed, draft, etc)

Whenever possible the software was tested, but in some cases, where this option was not available, the review is based on user manuals/articles or webpages. Finally it is important to note that often, the systems do not present any information about the type of license, even if sometimes they are distributed at no charge.

The results of this review, are summarized in a table, on a SQLite database [23]. In the next few paragraphs, we are going to analyse the content of this table, reviewing specific characteristics of these systems.

Most of the systems analysed include some sort of database system, to store and manage information (see Figure 1a). The DBMS found are very diverse, ranging from free systems such as MySQL, to expensive proprietary solutions such as Oracle or MS SQL Server. Nevertheless, the majority of the systems use MS Access DB, which although commercial, is low-cost solution (and there are even some free versions).

The same conclusion can be drawn about GIS (see Figure 1b). The majority of systems uses a GIS component, or has a direct connection with GIS. However, the number of systems without GIS is close to the number of systems with GIS.

As we can see in see figure 2a, the vast majority of the reviewed systems does not require an internet connection, in order to work. Also because some of the systems, are low in technology and make little (or no) use of the computer. However, it has to be said, that in modern systems there is an increase use of the server-client architecture, where they centralize the information (database, maps), and serve it through "thin" clients (normally web browsers). The use of web-mapping services such as WMS, triggered a "boom" in the adoption of this kind of model. It is important to note, that although the big advantage of this model is its distributed use, it may also possible to setup it locally on a LAN, or even in a single computer,

Most of the reviewed systems, whether they use software or not, do not present any information about the type of license in place (if any). This can be a consequence of the fact that most of the systems are not distributed, since they are nowhere available for download.

Moreover, most systems do not require a high knoweledge of GIS, databases, statistics, etc, in order to be used (see figure 2b). This is either because: costum made solutions were implemented ("tailor-made" software), or widely-known software is used (like for instance MS Access). However, in some cases, specific tools are used that require a bit more literacy (for instance R); in this cases, there is a trade-off since there is something to gain in terms of potentialities of the software allowing to perform more complex tasks, but it is also narrowed the number of possible adopters.

As we can see in figure 3, most of the reviewed systems, whether they use software or not, do not present any information about the type of license in place (if any). This may be a consequence of the fact that they are not widely distributed (?).

From the systems that specify licenses, there is the same amount of proprietary and open-source type licenses, which clearly indicates the existence of the two segments. Looking at the timeline, there has been an increased adoption of "free and open-source-type" licences, and the expression of interest in adopting these licenses in systems that still rely on proprietary components. This is a very positive trend with great social impacts, specially regarding the adoption of these systems in developing countries.

The table bellow presents the 25 reviewed systems.

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
Fishframe	ICES	Web-based data warehouse application. It is a platform for data compilation and tabulating, that allows: upload of information, data quality control, data integration and raising.	Modularized software architecture; server client architecture; browser independent; well specified format;
InterCatch	ICES	Web-based system for handling fish stock assessment data, focusing on documenting characteristics of the catches.	Client-server architecture (runs on a browser); it uses ASP and SQL Server;
ICES Spatial Facility	ICES	Free interface for searching, viewing and downloading spatial data.	WMS service (Geoserver) and client (OpenLayers); it was developed in PHP, with an SQL Server database and makes use of the Geonetwork;
<i>continued on next page</i>			

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
ICES EcoSystem Data	ICES	Spatial database; it uses the standards adopted by the user community for the dissemination, visualization and sharing of ICES raw spatial data.	WFS server and client(Geoserver, OpenLayers). MS SQL Server database.
FMS	Schooner Solutions	Software that allows fishermen to track and manage information on daily trap fishing activities; at the same time, it produces reports of catch and effort for the DO.	Stand-alone software that runs on Windows (98 to XP).
The Canadian Albacore Tuna Catch And Effort Relational Database	Oceanic Fisheries Program (SPC)	System that provides annual catch estimates for the Canadian albacore troll fishery based on sales slips, logbooks, and hail Information.	MS Access
TUFMAN	Oceanic Fisheries Program (SPC)	Database fishery management tool. It provides solutions for: data entry, data management, data quality control, administration, and reporting.	Server-client architecture; server in SQL Server and client in MS access (supports the free versions); the mapping is provided through a loose couple with Mapinfo;

*continued on next page*

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
Catch and Effort Query System (CES)	Oceanic Fisheries Program (SPC)	Menu-driven system, which allows member countries to extract summaries of operational logsheet data, aggregate public-domain catch and effort data, and annual catch estimates.	It uses Mapinfo for mapping.
Fishery Analyst Online version	MappaMondoGIS	Web-based application, to query fishery data, analyze and visualize temporal and spatial patterns of fishery dynamics.	It uses the Google and ArcGIS JavaScript APIs and the Dojo framework.
Fishery Analyst for ArcGIS 9.x and 10	MappaMondoGIS	Software for the quantitative estimation and visualization of catch and effort and their variation in space and time, analysis of fishing vessel utilization, data quality control, and deriving information on the location of important economic and threatened species.	Developed in VBA (arcobjects), inside ESRI ArcGIS. The data is stored in DBF files.

*continued on next page*

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
FiRST	Collaborative effort across eight South and South-east (World Fish Center) Asian Countries.	Data management system for scientific trawl survey data. It includes data summary and reporting functionality; outputs: catch, CPUE and biomass estimates. It includes visualization tools, an analytical routine to estimate biomass, and data import/export modules.	Server-client architecture. Server in MS access/ SQL Server (Windows); client in a browser.
PESCART	Instituto Nacional de Investigação Pesqueira (IIP)	Database system for storing data from artisanal fisheries and efectuate some calculus and reports (catch, effort, CPUE). It includes an interface for data acquisition, analysis and reporting.	Developed in Microsoft Access 97 (Windows 98); the system is split in two databases: one containins the data and the other one the app (UI+processing); it uses some MS specific ActiveX (Microsoft Flex-Grid Control 5.0) and Libraries (.e.g DAO)

*continued on next page*

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
Geocrust	University of Algarve (UALG)	System that uses Satellite data (VMS) to map effort and landings;	Modular stand-alone application with embebed GIS. It was developed using: Visual Basic 6.0., ADODB and MapObjects 2.0 Pro. It uses a MS Access Database (2000).
Geocrust 2.0 (Geopescas)	University of Algarve (UALG)	Improvement of the Geocrust application, as a consequence of extending the methodology to the ground fish trawl flee, a larger period, and a larger geographical area.	The db was migrated to MySQL running on a Linux server (higher security). It were introduced Artificial Intelligence methods (automatic classification methods) for trawl detection, in addition to the manual procedures.

*continued on next page*



*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
Geopescas Website	University of Algarve (UALG)	Website to disseminate geo-referenced information on fishing effort, landings, catches and catch rates, through fisheries researchers, fishery administration and fishing industry. It shows some of the maps generated by the Geocrust software and it offers dynamic mapping, using Geotools.	It uses the Geoserver mapping component, customised through JAVA programming.
<i>continued on next page</i>			

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
COST	Project financed by the European Commission under the call FISH/2006/15 – lot 2	Open Source Tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation.	set of R libraries, with the following functionalities: import and handle fisheries data, explore the data, estimate the parameters and related precision and do simulation. It relies on data in a well specified format.
Catch Assessment Survey and Statistical frame survey of the artisanal fisheries of Kontagora Reservoir	?	Catch Assessment Survey and Statistical frame survey, that aims to sustain the formulating of management plan and policies for the Kontagora reservoir.	Count and weighting of the fish.
Artisanal Fisheries Data Collection And Management In Tanzania	UNU Fisheries Training Program Fisheries Division - Statistical Unit	Simple sampling procedure, community based data collection model and specifying of types of data to be collected. This system also includes analysis methods and a "tax motivation" strategy.	MS excel, MS Access Database.

*continued on next page*

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
IFMP	Lake Victoria Fisheries Organisation (LVFO)	Catch Assessment Surveys, two-stage stratified sampling design and estimation of CAS-based Indicators.	data forms (paper), MS Excel.
Barahona Data Collection System	Barahona Data Col PROPESCAR-SUR project	National Data Collection System	it relies on paper records and the calculations are done" by hand".
Samaná Fisheries Development Project (CEDEP)	Japan?	Logbook recording system, for a small number of vessels.	?
FISBOAT	IFREMER	Fisheries Independent Survey-Based Operational Assessment Tools. The tools tackle the major issue of unreliable fish stock assessment, through the use of new methods based exclusively on research vessel survey data.	R, FLR simulation framework (R based)

*continued on next page*

*continued from previous page*

<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>
Probability-based survey techniques in Mozambique	Instituto Nacional de Investigação Pesqueira (IIP)	Probability-based survey techniques for monitoring catch and effort in the coastal small-scale fisheries.	onsite survey, database (?)
Programme for refining estimates of local fishery production in Western Samoa	Peace Corps Marine Biologist Fisheries Division, Western Samoa	A programme for refining estimates of local fishery production, that consists in data collection(interviews) and data analysis;	It uses written down interviews to collect data and a hand-held programmable calculator for analysis; as a database, it is used DBASE III.
Electronic Fish Catch Logbook Project (EFCL)	Northwest Fisheries Centre (NWFSC)	Ship-based software and shore-based web application to collect and report logbook, catch "(fish tickets)", biological sample and observer data. There is also a report system, a central database and web based app to upload data.	Client server architecture, Oracle database, Arcinfo GIS;

*continued on next page*

<i>continued from previous page</i>			
<b>Name</b>	<b>Institution</b>	<b>Description</b>	<b>Technical Specifications</b>

Table 1: Table-resume of the reviewed Fisheries Management Systems

In the following table, the reviewed systems are organized according to their technical characteristics:

- *internet*: indicates if the software needs an internet connection to work (binary value);
- *GIS*: indicates if the software uses GIS (binary value);
- *database*: indicate if the software uses a database (binary value);
- *license*: indicates the type of license;
- *portability*: comments about the cross-platform ability;
- *expertise*: indicates if there is the need of expert knoweledge to run the system (binary value);
- *released*: indicates if the system is released and already in use;

<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
Fishframe	1	1	BSD	1	Since the clients are browser based, they are cross platform; the server has to run on Windows, since it uses Microsoft specific active technologies (ASP)	0	yes
<i>continued on next page</i>							

<i>continued from previous page</i>							
<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
InterCatch	1	?	?	1	Since the clients are browser based, they are cross platform; the server has to run on Windows, since it uses Microsoft specific active technologies (ASP)	?	?
ICES Spatial Facility	1	1	?	1	Server and client are cross-platform.	0	yes
ICES EcoSystem Data	1	1	?	1	Server and client are cross-platform. Database is Windows specific.	0	yes
FMS	0	?	proprietary	?	Windows only (98 to XP)	?	under final modifications
The Canadian Albacore Tuna Catch And Effort Relational Database	0	1	?	1	Windows only	0	yes
<i>continued on next page</i>							

<i>continued from previous page</i>							
<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
TUFMAN	0	1	?	1	Windows only	?	yes
Catch and Effort Query System (CES)	0	1	?	1	Windows only	0	yes
Fishery Analyst Online version	1	1	proprietary	?	The client is free and cross platform (runs on a browser)	0	yes
Fishery Analyst for ArcGIS 9.x and 10	0	1	proprietary	?	Windows only	1	yes
FiRST	1	1	?	1	Since the client runs on a browser, it is platform independent. Server runs on Windows.	0	yes
PESCART	0	0	?	1	Windows only	0	yes
Geocrust	0	1	?	1	Windows only	0	yes
Geocrust 2.0 (Geopescas)	0	1	?	1	The application runs on Windows, but the database can now run on Windows or Linux.	0	yes
Geopescas Website	1	1	?	1	?	0	demo version only
<i>continued on next page</i>							

<i>continued from previous page</i>							
<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
COST	0	1	Open-Source	1	since it relies in R, it is cross-platform.	1	yes
Catch Assessment Survey and Statistical frame survey of the artisanal fisheries of Kontagora Reservoir	0	0	?	?	?	0	yes
Artisanal Fisheries Data Collection And Management In Tanzania	0	0	?	1	Windows only	0	yes
<i>continued on next page</i>							



<i>continued from previous page</i>							
<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
IFMP	0	0	?	0	Windows only	0	yes
Barahona Data Collection System	0	0	?	0	n/a	0	yes
Samaná Fisheries Development Project (CEDEP)	0	0	?	0	?	?	yes
FISBOAT	0	0	Open-source	0	multi-platform	1	yes
Probability based survey techniques in Mozambique	0	0	?	?	?	?	yes
Programme for refining estimates of local fishery production in Western Samoa	0	0	?	1	Windows only (DOS)	0	yes
<i>continued on next page</i>							

<i>continued from previous page</i>							
<b>Name</b>	<b>internet</b>	<b>GIS</b>	<b>license</b>	<b>database</b>	<b>portability</b>	<b>expertise</b>	<b>released</b>
Electronic Fish Catch Logbook Project (EFCL)	1	1	?	1	Some components run on a browser, so they may be cross-platform, but the database and GIS are Windows only.	0	

Table 2: Reviewed Fisheries Management Systems, according to their characteristics

The next table, goes a bit more in depth, and it analyses some of the advantages and disadvantages of each system.

<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
Fishframe	- data ownership of the recipients; - deals with incomplete data; - cross platform); - enforces confidentiality; - involvement and engagement of stakeholders; - bottom-up approach	- some components require Windows XP; - development is Windows based; - for complex tasks the browser solution is not efficient;
InterCatch	?	It uses proprietary and platform-specific software
ICES Spatial Facility	- cross platform (browser based); - adopts OGC standards; - friendly UI; - uses Free and Open Source Software( FOSS);	There is no component of data acquiring in this system; it is assumed the raw data is validated.
ICES EcoSystem Data	- cross platform (browser based); - adopts OGC standards; - friendly UI; - uses Free and Open Source Software( FOSS);	- some requests may be slow due to the quantity of data; - it deals only with survey data;
<i>continued on next page</i>		

<i>continued from previous page</i>		
<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
FMS	It does not require a very powerful hardware to run.	- it runs only on windows; - it has a cost;
The Canadian Albacore Tuna Catch And Effort Relational Database	- it deals with unreported catch; - although it is a post-season analysis tool, it can be used as in-season analysis (based only on trip logs);	- the results may be biased due to the use of logbooks without saleslip data; - other sources of bias are the "public sales" and "take homes" that are not registered anywhere (they "escape" from logbook and sales slips);
TUFMAN	- the website provides excellent documentation on how to use the application; - flexible system, able to cater the needs of each country; - it allows to identify gaps in data;	- it uses proprietary components (Mapinfo, Microsoft Office Pro); - it does not feature data acquisition, and therefore it is dependent on the quality of input data; - it does not provide tools to fill the gaps in the data; - it is not found anywhere to download;
Catch and Effort Query System (CES)	- developed reporting facilities, with the use of graphs, tables and maps;	- it uses proprietary components (Mapinfo); - it is not found anywhere to download;
Fishery Analyst Online version	- lightweight and cross-platform application;	- it has a cost;
Fishery Analyst for ArcGIS 9.x and 10	- it takes advantage of GIS to analyse geodata; - it provides a time series animation; - it provides data quality control routines; - it uses transversal information (economical); - confidentiality is addressed;	- it has a cost, although a demo version is provided for free; - it requires ArcGIS 8.3. or 9 with Spatial Analyst (and it needs this specific versions to work); - it requires some knowledge of ArcGIS to use; - it requires some detailed georeferenced data to start with (no estimation provided!)
<i>continued on next page</i>		

<i>continued from previous page</i>		
<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
FiRST	<ul style="list-style-type: none"> <li>- client is cross platform;</li> <li>- no installation required for the client;</li> <li>- a regional database is “compiled” from the national databases;</li> <li>- outputs can be used for ecological modelling (very usefull for management – devise or revise zonation schemes);</li> </ul>	<ul style="list-style-type: none"> <li>- the database is designed for scientific trawl surveys;</li> <li>- it requires installation of a webserver;</li> <li>- the exporter uses mostly proprietary formats (MS Office, Crystal reports, MS Access);</li> </ul>
PESCART	<ul style="list-style-type: none"> <li>- intuitive UI following the information flow;</li> <li>- it is a very complete system, since it covers: data acquisition, estimation and reporting.</li> </ul>	<ul style="list-style-type: none"> <li>- it targets artisanal fisheries only (the basic unit is fishing unit);</li> <li>- ”bulky” setup with a lot of requirements;</li> <li>- the application does not have any access restrictions (no permission system implemented);</li> <li>- the limitations of MS Access ”clash” with the design of the app (e.g. number of fields on a table does not allow the selection of a large number of species;)</li> </ul>
Geocrust	<ul style="list-style-type: none"> <li>- it combines positioning and landing data, from different sources (IGP, DGPA, logbooks);</li> <li>- extense use of GIS, not only for display but also for calculations (e.g.: buffers to define start and end of fishing trips);</li> <li>- it is generic enough, to be used with other fleets/geographic areas/species;</li> </ul>	<ul style="list-style-type: none"> <li>- it is based on proprietary technologies;</li> <li>- it relies on accurate positioning data (VMS);</li> <li>- it was not addressed a permission model for the UI (only superuser);</li> </ul>
<i>continued on next page</i>		

<i>continued from previous page</i>		
<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
Geocrust 2.0 (Geopescas)	- the procedure of identification of trips was enhanced with AI; - DBMS was migrated to a free, multi-platform software;	- it is still based on proprietary technologies; - the AI procedure needs adequate "training", in order to produce good results;
Geopescas Website	- it provides an interface for querying maps; - it exports georeferenced maps, images and XML, that can be integrated in third-party apps.	- it depends on technologies (Ggertools) that are heavy on the browser and require the installation of the Java plugin; - it was only released a demo version;
COST	- space dimension is included in the analysis; - uses well known technology; - promotes import and export of data; - open source and cross platform tool; - interfaces with Fishframe format;	- relies in good quality and well formatted data; - it requires expert knowledge;
Catch Assessment Survey and Statistical frame survey of the artisanal fisheries of Kontagora Reservoir	- it is a complete system, including data collection (survey) and statistical analysis of the results; - it comprehends both effort and catch;	Although the data is collected spatially, there is no indication that this component is taken in account in the sampling and analysis.
Artisanal Fisheries Data Collection And Management In Tanzania	- relatively inexpensive; - it does not need specialized staff to run and maintain it; - it contemplates the community participation;	There is no reference to GIS in the sampling procedure (although the sampling is on space), or in the estimation process.
IFMP	- it provides estimations of the catch over the years, which allows to detect trends, to detect need for further data and to evaluate management interventions.	- since data collection is through paper sheets, there is no data quality control on introduction; - there is no mention of a solution to store and manage the data over time (DB); - there is no use of GIS, although the PSU are spatial;

<i>continued from previous page</i>		
<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
Barahona Data Collection System	- the economic variables are well catered; - it gathers relatively accurate catch and effort information based on trip interviews; The effort variables, in particular, seem well thought out and adequate for monitoring CPUE and estimating fishing mortality; - the system was sustained, even after withdrawl of external funding;	- mortality and growth cannot be estimated with current data ; - there is the need for a new approach fro estimating discards; - the current sampling procedure does not follow a statistically rigorous method. - there is no database to store and manage data on a long term; - there is the need of training; - stock assessment is not contemplated in this system;
Samaná Fisheries Development Project (CEDEP)	- data collection is rigorous;	- the logbooks are not formally structured; - the data is not mantained in a database;
FISBOAT	- these tools provide robust and reliable scientific advice to fishery managers and policy-makers, including a management model driven by numerically defined harvesting rules; - it is provided an early-warning mechanism, giving policy-makers more time to respond proactively to evolving challenges.	- the methods are based exclusively on research vessel survey data; - it requires expert knoweledge to be used (statistics and computation);
<i>continued on next page</i>		

<i>continued from previous page</i>		
<b>Name</b>	<b>Advantages</b>	<b>Disadvantages</b>
Probability-based survey techniques in Mozambique	- the probability sampling produces unbiased estimates of catch, effort, and other key parameters; - the precision in such estimates can be quantified; - it is cost-effective; - the underlying theory is well developed and documented;	- no mentioning of use of GIS;
Programme for refining estimates of local fishery production in Western Samoa	- cost-effective method of gathering basic information that provides clues to changes in the fishery as a whole; - it requires a small input of resources for the information obtained;	- the data quality of the survey is threatened by the hiring of casual workers, without adequate supervision; the same for data entry in the database; - outdated database technology; - there is the need to collect more data (from the villages) to estimate the catch; - there is the need to design more complete entry datasheets;
Electronic Fish Catch Logbook Project (EFCL)	- innovative effort to automate and standardize logbook data. - it implements permissions through roles; - it supports secure connections (data encryption) - the use of a software to input the raw data, allows a first validation layer that does not exist in traditional systems using paper sheets; - customisable;	- it is based in proprietary technology; - it relies on internet access for data uploads and visualisation;
<i>continued on next page</i>		

<i>continued from previous page</i>		
Name	Advantages	Disadvantages

Table 3: Advantages and disadvantages of each system

In the next section (see 3) we present a set of fisheries management tools that pretend to fill some of these "gaps" on existent systems, reusing some of its (good) ideas and adding new technologies that are "giving cards" on other fields (like fuzzy mapping, on section 3.4, or artificial intelligence), having in mind the limitations and potentialities of the recipients (developing countries). Some of these ideas also arise during the development of Medfisis-CAS [11], an (almost) full and open-source system that reunites data from Logbook and Artisanal Fisheries in the same application/database; some of the concepts developed during this project such as abstractability of entities and configurability of components can be very useful in all stages of Fisheries Management Systems.

Figure 4 shows some of the challenges that can be addressed by a set of fisheries management tools. We propose a loose couple architecture, where we maximize the use of existing tools: to cut development time, to benefit from their value and also to involve a community of users that can take advantage of their expertise in specialized tools (like for instance R, or QGIS) to push the system further. The modularization of the system also means that we can distribute tasks, and till a certain extent develop them independently without compromising the rest of the system.

### 3 Project Objectives

The design of a software system for fisheries management, as described in section 2 (figure 4), should contemplate four stages: **input, validation, transformation and output**. The main target for each of this stages is the data about fisheries, in its raw or transformed state (catch, efforts), but it can also be other (complementary) types of data (meteorological, oceanographical, socio-economic, etc). In the following subsections we identify a series of tasks that are *on-demand*, and could result in very useful tools for fisheries scientists and managers. They are not limitative, and in our modular design they can be seen as a first approach rather than a complete description of the system.

#### 3.1 Data Input

The data input component contemplates the collection of information for the system, either by creating this information "from scratch" or by importing it from different supports/formats.

- **Tools for Sampling Design:** Before even introducing information in the system, it is often necessary to collect this information; the design of a unbiased sampling scheme, that minimizes the costs and maximizes the quality of the results it is an important and often overlooked task; overlooked, because many times the designer is not provided with the necessary tools to make it an easy and efficient process.



There are many advantages of using a spatial framework (GIS) on this particular problem, and suggestions can guide the sampler to informed and wise decisions.

- **Research on Open Standards and Distributed Models for Data Exchange:** The problem of exchanging data between different instances of the system is a pertinent and frequent one. JSON [12] and SQLite [13] are open and well-known standards that are "light" alternatives to XML. Also, the design and implementation of a server-client model, where we leave the "heavy processing load" to the server and create "light" data acquiring clients (for instance running on mobile phones and Ipads) can simplify the task of data integration, as we only update restricted parts of the system.
- **Data Recovering Tools:** There is a large quantity of historical data that is lost in terms of its value, just because it is not in an usable format. Data series take many years to assemble, and not only it costs a lot of money to generate new data, but it is not actually possible to recreate past information, so when this information is lost it is lost forever. It could be a very rewarding task to at least try to compile this information, that may not be usable for different reasons; hard disks that have bad sectors, obscure data formats whose specification was lost, data in non digital support and nearly illegible: these are all problems eligible for "data archaeology", that can make use of advanced software techniques such as: writing recognition, decryption algorithms and hard-disk recover tools.
- **Integrate data from different systems/sources:** The integration of different types of information provides a rich basis for the analysis and transformation of information; this includes having data from different sources such as oceanographical, meteorological, sociological, etc; apart from implementing a system that is able to import different formats (satellite images, excel spreadsheets, vector maps, etc) it is also relevant to design a system that can deal with the nature of this variability - for instance recreational fisheries data, logbooks and sales slips can be integrated with a unified and generic design and provide a richer basis for fisheries statistical analysis (see [11]).
- **Data Entry User Interfaces:** Database clients are a *milestone* in fisheries applications and common examples of this are digital logbooks and fleet register software's (see [11]). They reproduce the behavior of paper sheets for data collection, with all the advantages of data checking, and "feed" directly the database; if they are well designed and the data operators are trained, are also quicker and more friendly to use than the replaced papper sheets. There is still a room for improvement on this UIs, for instance enabling multiple choices or configurable validation rules <sup>1</sup>.

---

<sup>1</sup>As an example, some of this features were already implemented in CAS Medfisis [11]

## 3.2 Data Validation

Since the data entering the system may arrive from sources other than its data generator interfaces<sup>2</sup>, it is important to control the quality and validity of information we are inputting; this step is fundamental for all the following ones, since "bad quality" data will compromise the entire system ("garbage in, garbage out").

The Fishframe application [1], developed by the Danish Institute for Fisheries Research, presents an interesting and complete multi-step validation layer; on a first stage it is performed a basic check, looking at the structure of the data, ranges of values and duplicate values; on a second stage, a more complex check searches for the correctness of the dependency between fields; finally, the last check (which is not compulsory) performs a global analysis, looking at the general patterns in order to identify outliers. Apart from this, it also allows some customizable checks to be introduced by the user, that can be performed automatically. This flexibility and completeness of the validity step, inspired us to identify the following tasks in the data validation component of this system:

- **Detection of redundant data:** The detection (and consequent removal) of duplicate records is an important task, in order to have a "clean" database. It is not complicate to identify automatically records that are **exactly** identical, but the problems arise when we have for instance spelling mistakes, and records that are *almost* the same, but not exactly the same. As this situations are dubious, and may correspond or not to duplicate data, they require user supervision to confirm it; however, there is no reason why they can not be identified and presented to the operator has "possible duplicates"; for this task, it is very useful to use data mining techniques such as a fuzzy approach to language [14].
- **Detection of invalid data based on configurable rules:** Generic rules can be applied to check for the correctness of the data types, correctness of keys, etc, directly based on the database schema. However, some rules such as the format of strings, ranges of values, may change from system to system (for instance according to the geographic location) and the option to create them should be left to the user. To avoid "hard-coding" of rules, which would require a programmer's intervention and consequently difficult the process of exploratory analysis, we propose an interface for user configurable rules. This interface can be similar to a query builder such as the one represented in 7, which may be familiar to some users.
- **Detection of Outliers:** Outliers might arise due to careless data acquiring, instrument malfunction, wrong data processing routines, or any other reasons [15]; generic methods based on tabular data (for instance using a principal component analysis) can be used to perform this task, but there is also room for other techniques such as data mining, or GIS; for data with a spatial distribution (which is often the case in fisheries) GIS can provide a very valuable tool for capturing

---

<sup>2</sup>The data entry interfaces should have already their own validation routines removing, at least partially, the need for this step.

the general spatial pattern and consequently the values that fall outside this pattern (outliers)(see figure 8). After detecting and removing outliers values, we are left with the problem of imputation of "missing values" to replace the removed information (see section 3.3)

### 3.3 Data Transformation

Data transformation consists in using the information in the system and apply transformations to it, in order to generate new information. These transformations may be used to estimate values extending the scope of the dataset, or to generate indicators that characterize the dataset, like for instance the CPUE, or both. In this system we suggest facilitating the export of data and leave (some) more complex data analysis to specialized software such as R (figure 9) or QGIS(figure 10).

Bellow are examples of some specific (and common) problems with catch data, for which we propose to develop some solutions in the system.

- **Imputation of missing data:** Under-sampled or non sampled strata lead to an unknown bias, a problem that should be reduced with a good sampling design [16]. For the situations where we have to deal with missing values, it is possible to impute them, having in mind that also the method used for this operation can introduce some bias in the system. According to ICES [16] automatic methods should be avoid, in favor of expert knowledge, but they can be aided by spatial modelling techniques having in mind that to be use with success the spatial distribution should remain stable over time; this is a cue for developing GIS based models, that can provide the expert with a good basis for informed choices.
- **Allocate Catch in undefined areas:** On the Canadian Albacore Tuna Catch and Effort Relational Database [6] it is presented the problem of having sales or logbook catch data, whose geographic area is undefined. Instead of discarding these values from the calculations of catch by area, the authors found ways to distribute the catch and effort, following the overall geographic pattern of fleet; they allocate these values in proportion with the total catch and vessel distribution and for this purpose GIS can be very useful, for instance to produce a grid, or a contour map (see figure 11) with frequencies.
- **Estimate known unreported catch:** Another delicate problem arises from the incomplete vessel count. If there is an unreported effort than it will probably be reflected in an unreported catch. For the cases where it is known that an unreported effort exists, the authors of the the Canadian Albacore Tuna Catch and Effort Relational Database [6] suggest to extrapolate the total catch estimates based, on estimated number of unreported vessels. According to [17], unreported catch can be as significant as 50% of the total catch and therefore it is important to find suitable methods to approach this problem, like for instance the Monte Carlo Simulation (see figure 12).

### 3.4 Data Output

Data output, of raw or transformed information, is a component where it is possible (and desirable) to facilitate the link with other platforms by using "friendly" and well-known formats. This could include formats for tabular data (for excel, R, etc), but also formats for spatial data (shapefiles, GeoJSON, etc).

GIS provide some very-rich and expressive visualization tools with a great potential for fisheries data, that we will discuss summarily in the next few paragraphs.

Web mapping is the process of designing, implementing, generating and delivering maps on the World Wide Web <sup>3</sup>. Compared to "traditional" maps, maps on the Internet have many other advantages such as "live updating", and link with many other sources of information; they provide the same interactivity as a GIS software, with the advantage of not requiring a specific software client, since they can run on an Internet browser. Web maps are normally incorporated in portals, together with other types of information, like in the case of ICES Ecosystem Data [18] (see figure 13), but they can also be embedded in a software application (see figure 14).

Continuous maps showing densities, can be created using grids or contours or other methods such as TINs. Grids (such as the one on figure 13) are very common to represent distributions of fishing effort or catches; On the Geocrust project (see [19]), the generation of grids with different sizes allowed the user to have a more generalized or more detailed perception of the fishing effort, which calls its attention to different properties (see figure 15).

GIS functions for cartography are generally based on boolean logic: that is, sharp frontiers between entities. However, when there is uncertainty and vagueness attached to a certain phenomena, this "false" precision may be awkward or simply inadequate (see [20]). The Zadeh's fuzzy set theory is an alternative to this boolean logic and has been proposed has a foundation to GIS design, that can be incorporated to cartographic display of phenomena with a certain imprecision associated to them. On figure 16 we see the comparison of a traditional map, with sharp boundaries and a fuzzy map.

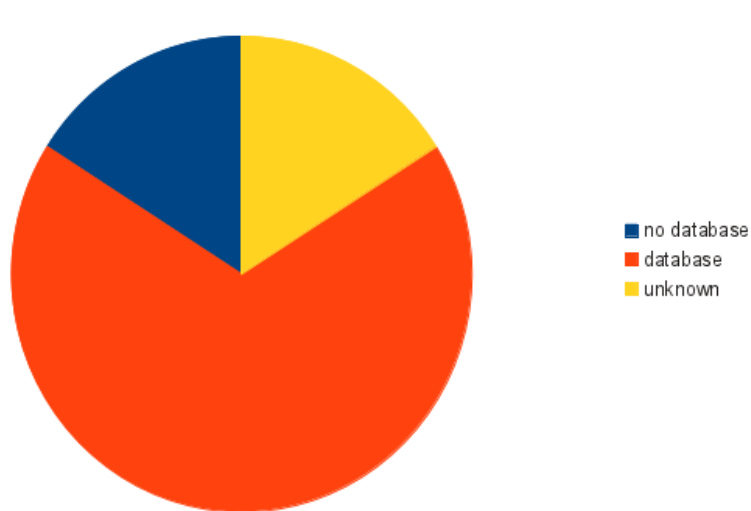
On fisheries surveys, where often spatial data relies on human description (for instance a fishermen describing where he was fishing), this could provide a very interesting approach.

## 4 Aknowledgegments

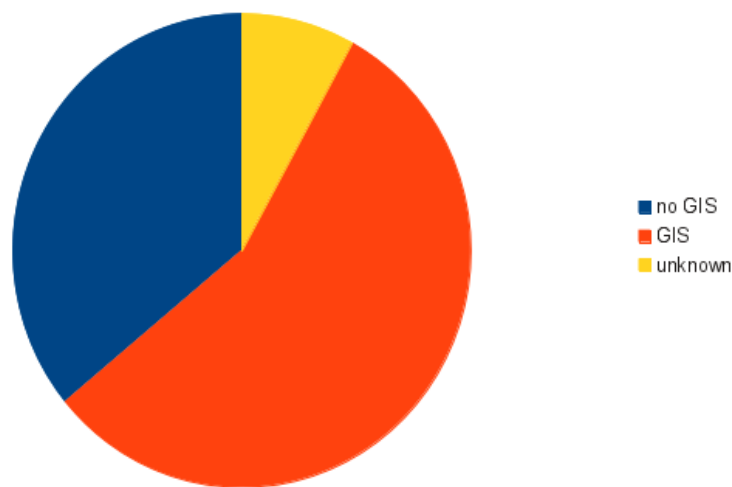
I would like to thank Henrik Degel, from the Fishframe project, for kindly providing me with a login that allowed me to try the Fishframe system; Carlos Pinto, from ICES, for amending and "filling the gaps" for the ICES systems reviewed on section 2.1; and Fabio Carocci, from FAO, for suggesting me to review a proprietary GIS fishery system.

---

<sup>3</sup>[http://en.wikipedia.org/wiki/Web\\_mapping](http://en.wikipedia.org/wiki/Web_mapping)

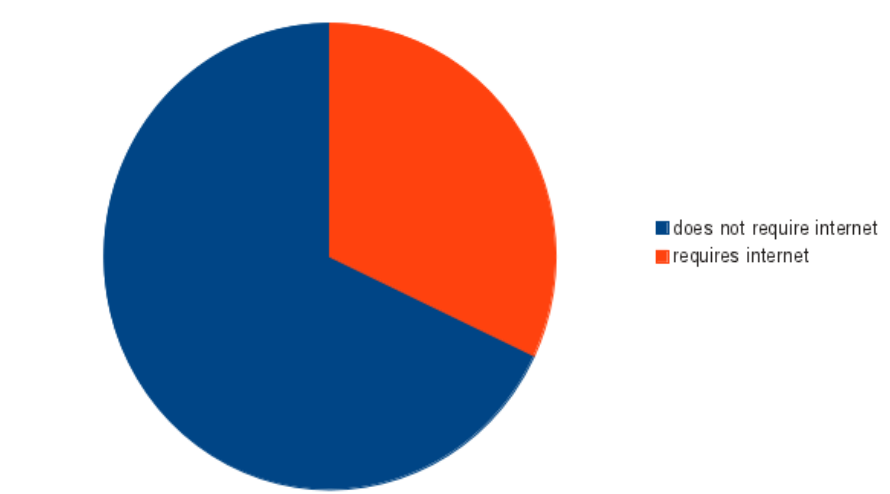


(a) Software use of database;

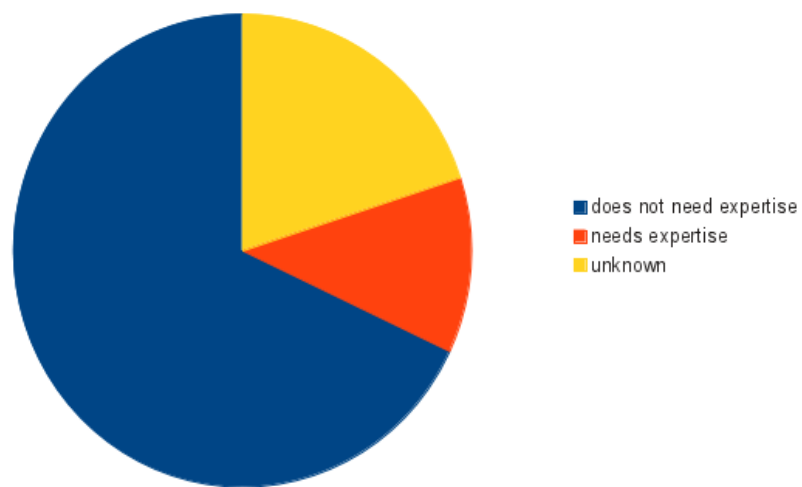


(b) Software use of GIS;

Figure 1: Aggregated analysis of the use of specific software components in the reviewed fisheries management systems;



(a) Need of internet connection, in order to run the software;



(b) Need for expert knowledge, in order to use the software

Figure 2: Aggregated analysis of specific characteristics the reviewed fisheries management systems;

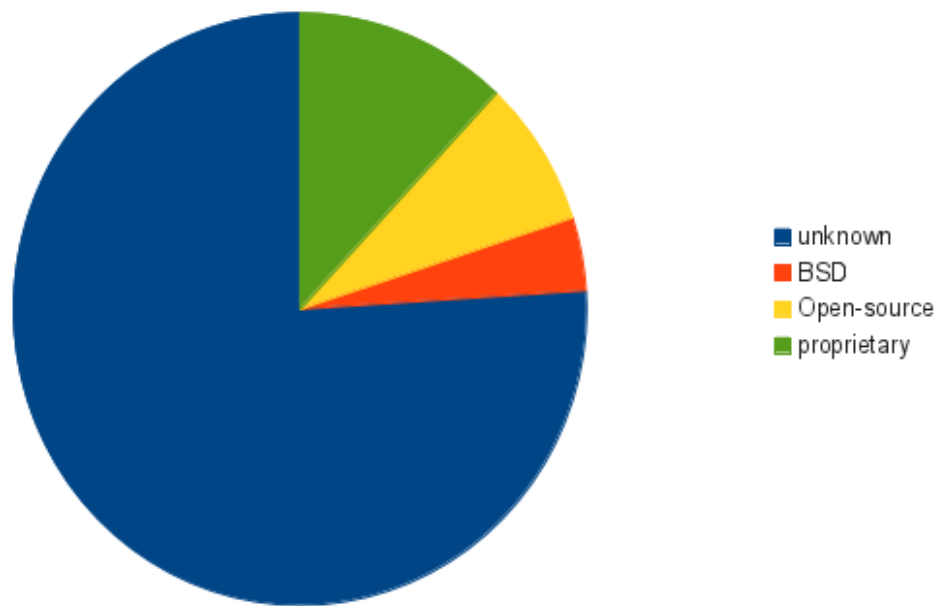


Figure 3: Software licenses;

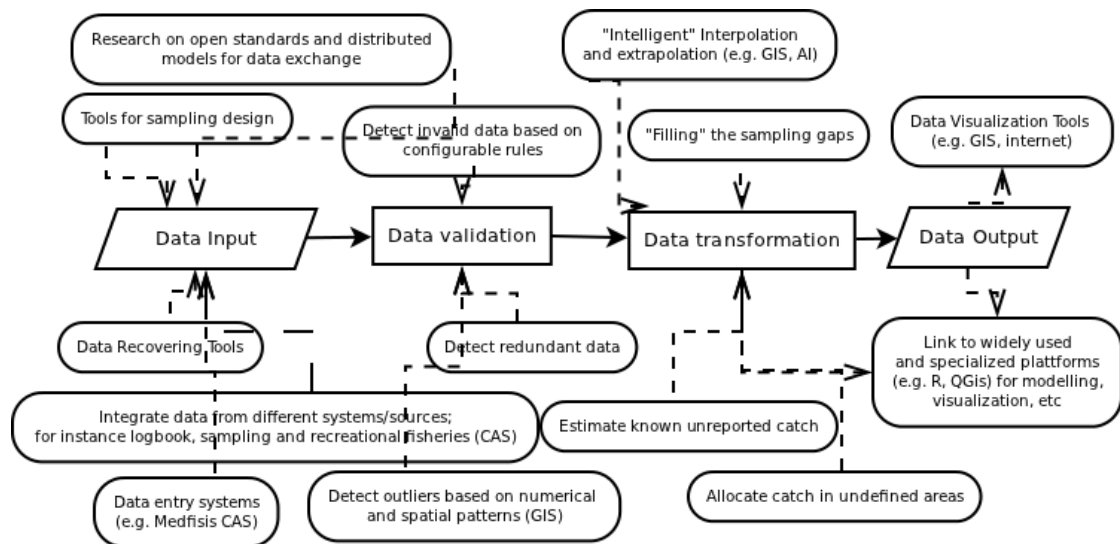


Figure 4: Some Interesting Challenges on Fisheries Management Systems;





Medfisis2-a

File Help

Frame Minor Strata Vessel Fishing Trip

Frame 1 -> no2 -> Name 6 Logbook

Choose Fishing Trip:

Name	Start Date	Start Time	End Date	End Time

Details

Landing Site:

Sampler:

Number of professional fishermen:

No part-time fishermen:

Start

Date:  Time:

Format: dd/MM/yyyy Format: HH:mm:ss

☒ Has time

End

Date:  Time:

Format: dd/MM/yyyy Format: HH:mm:ss

☒ Has time

Totals

	Estimated	Calculated	
No of operations	<input type="text" value="0"/>	<input type="text" value="0"/>	
Catch (weight)	<input type="text" value="0.00"/>	<input type="text" value="0.00"/>	Units: <input type="text" value="Kg"/>
Catch (no bones)	<input type="text" value="0.00"/>	<input type="text" value="0.00"/>	Weight/box: <input type="text" value="0.00"/> Units: <input type="text" value="Kg"/>

Comments:

No comments

Close Apply

Previous Next

Figure 6: Screenshot of a data entry form for fishing trip data, on CAS Medfisis [11]; on this system, forms were designed to accommodate at the same time logbook and artisanal fisheries data;

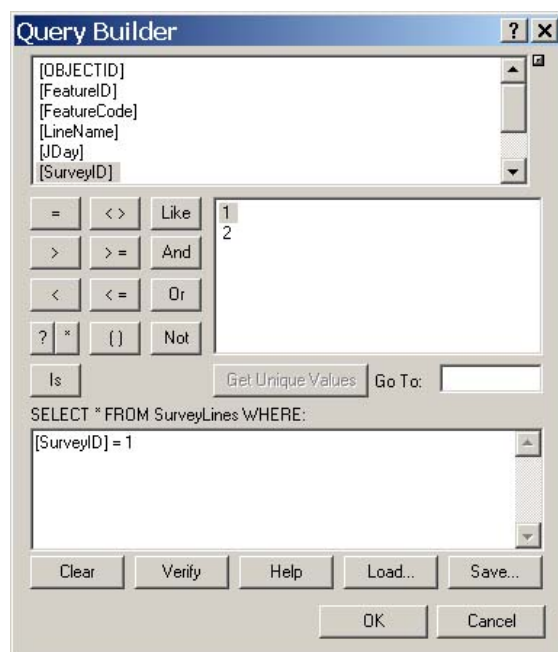


Figure 7: Query builder from ArcGIS; source: <http://woodshole.er.usgs.gov/pubs/of2006-1381/html/fig19.html>;

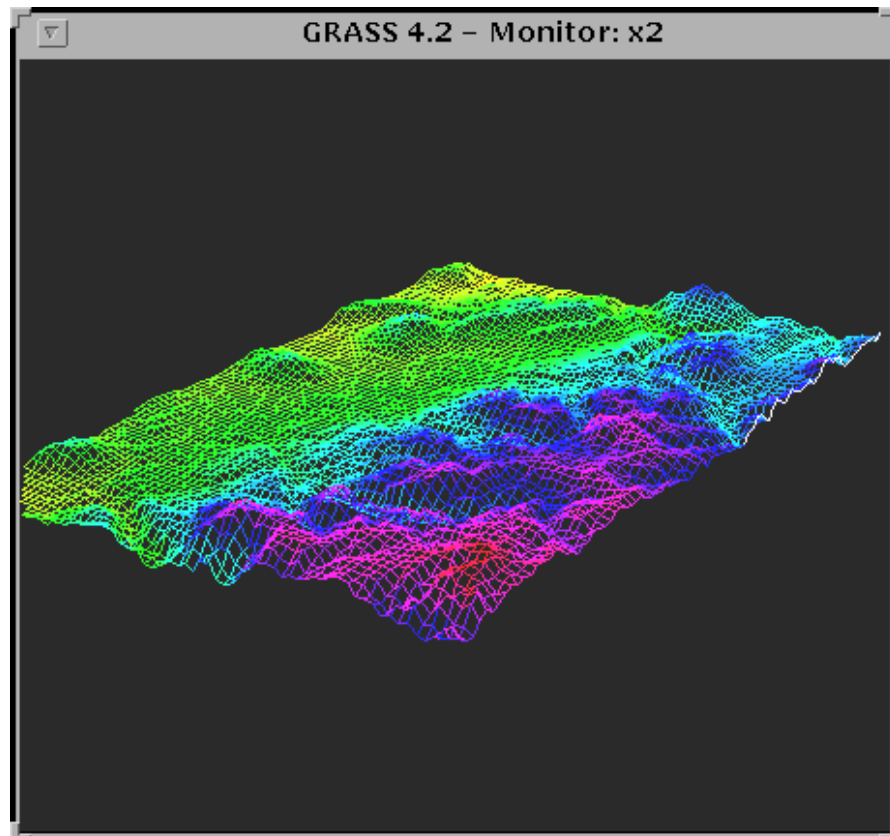


Figure 8: A Digital Elevation Model (DEM) like the one above, can be used to generate a surface of values in order to detect outliers; source: <http://www.grass-kr.org/research/demo1/index.html>;

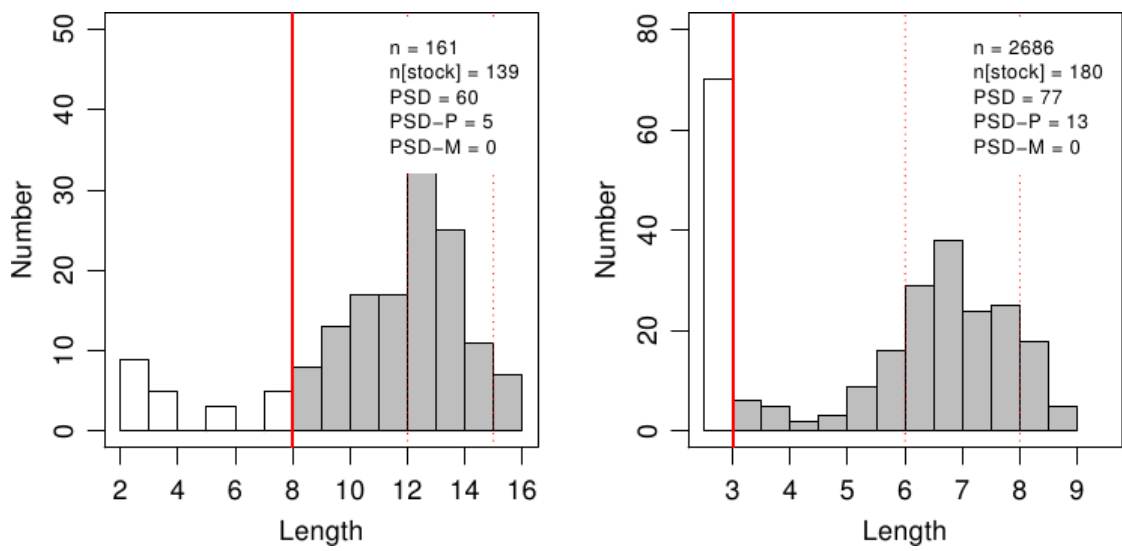


Figure 9: This analysis of the size structure was created using the R package; source: <http://www.ncfaculty.net/dogle/fishR/index.html>;

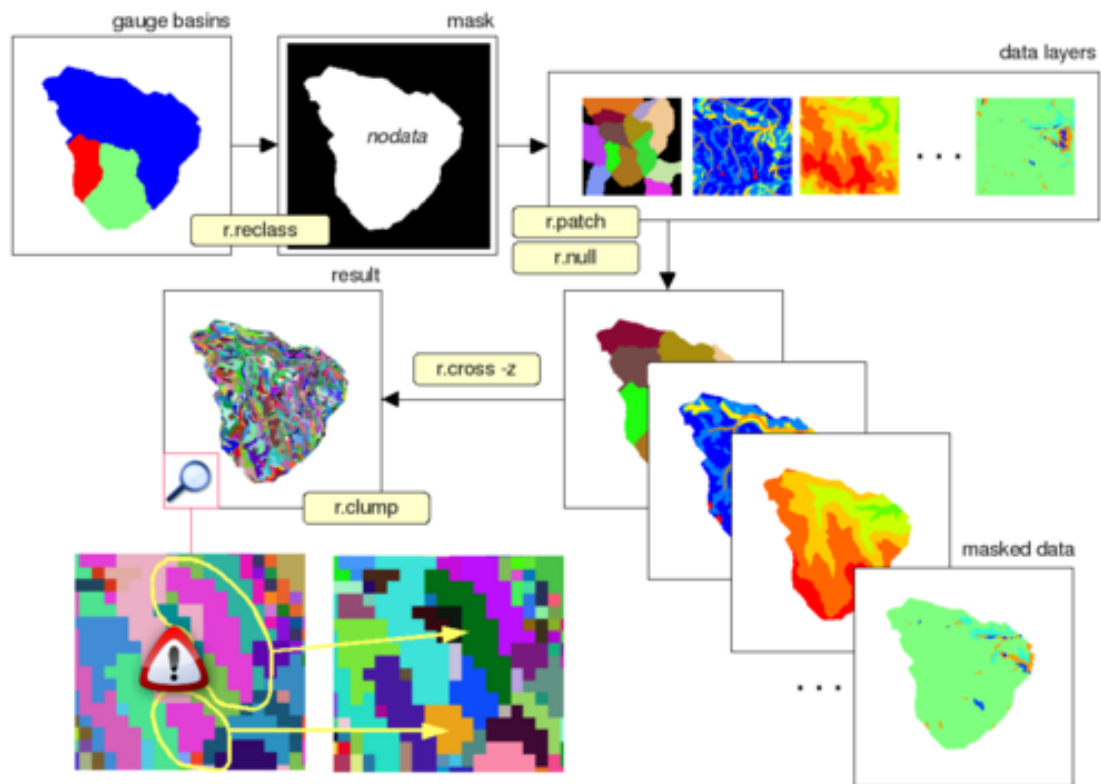


Figure 10: QGIS (in this example using the grass functionality) can be used to overlay different types of information; source: [http://geoinformatics.fsv.cvut.cz/gwiki/Deriving\\_Hydrological\\_Response\\_Units\\_\%28HRUs\%29\\_using\\_a\\_Web\\_Processing\\_Service\\_implementation\\_based\\_on\\_GRASS\\_GIS](http://geoinformatics.fsv.cvut.cz/gwiki/Deriving_Hydrological_Response_Units_\%28HRUs\%29_using_a_Web_Processing_Service_implementation_based_on_GRASS_GIS);

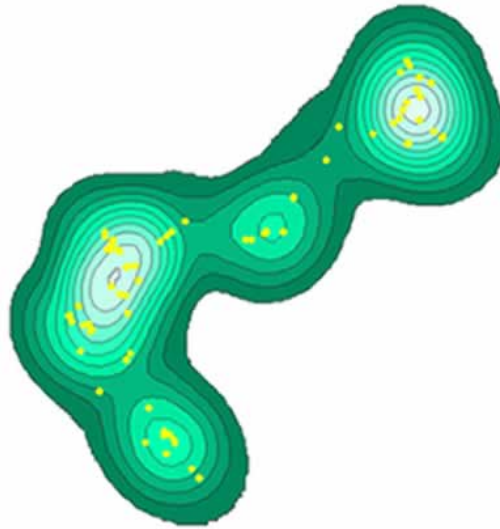


Figure 11: The Kernel Home Range Method is a probability measurement that outputs contours, based on ponctual observations; source: <http://www.bio.davidson.edu/people/midorcas/GISclass/GISprojects/grayson/GISMethods.htm>;

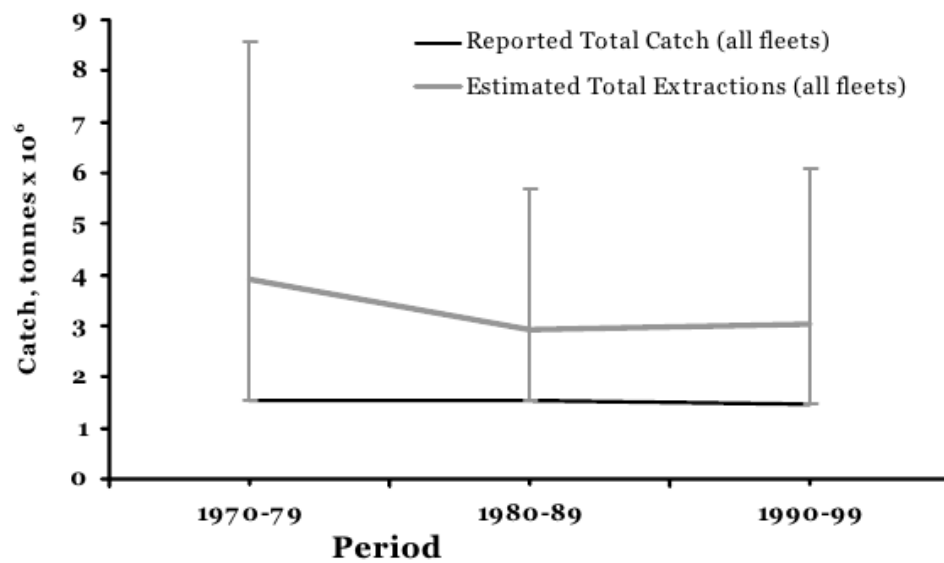


Figure 12: In this example [17], Monte Carlo Simulation was used to elaborate estimates of total extractions of all species from Moroccan waters; in the image we see the comparison with the total reported catch;

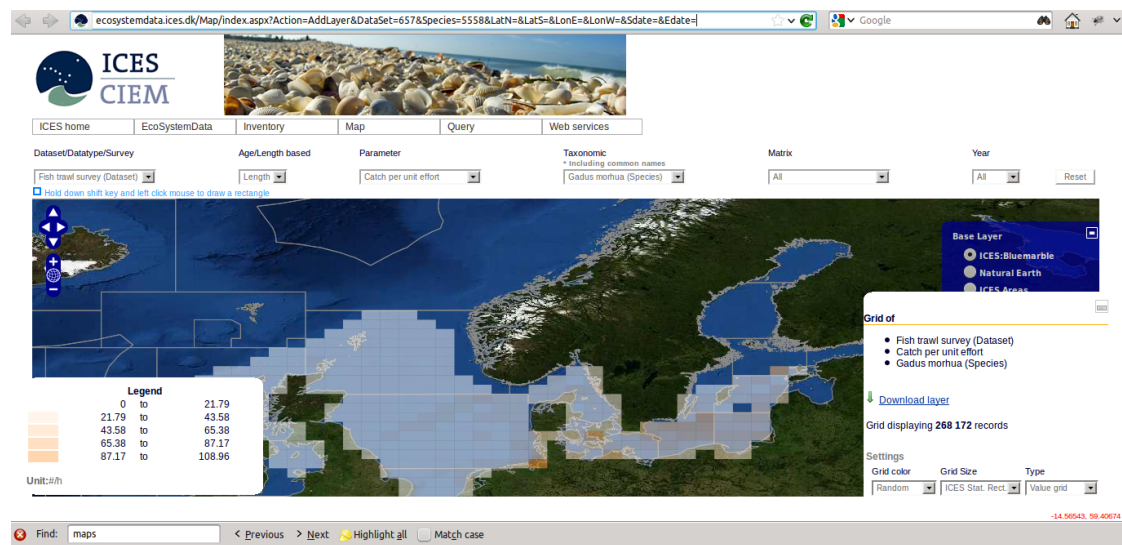


Figure 13: Web map of a ICES fish trawl Survey dataset, overlaying a grid of CPUE on top of a layer of Marble (Nasa Worldwind) [18];

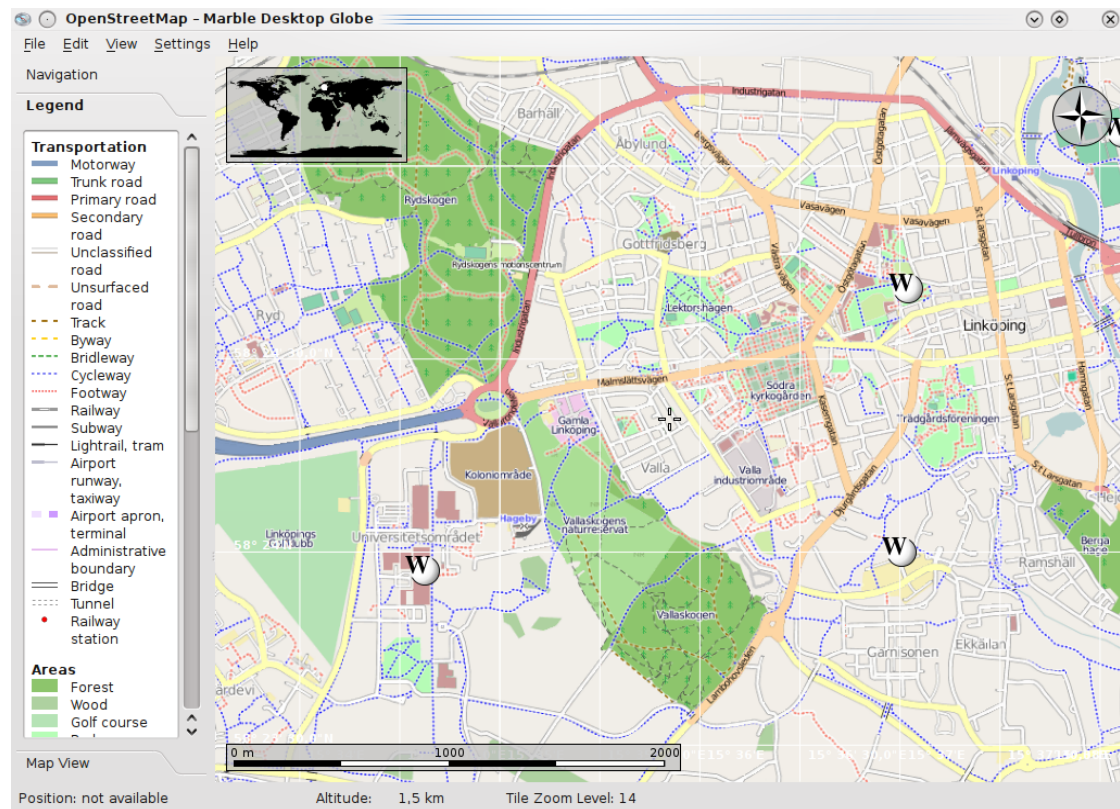
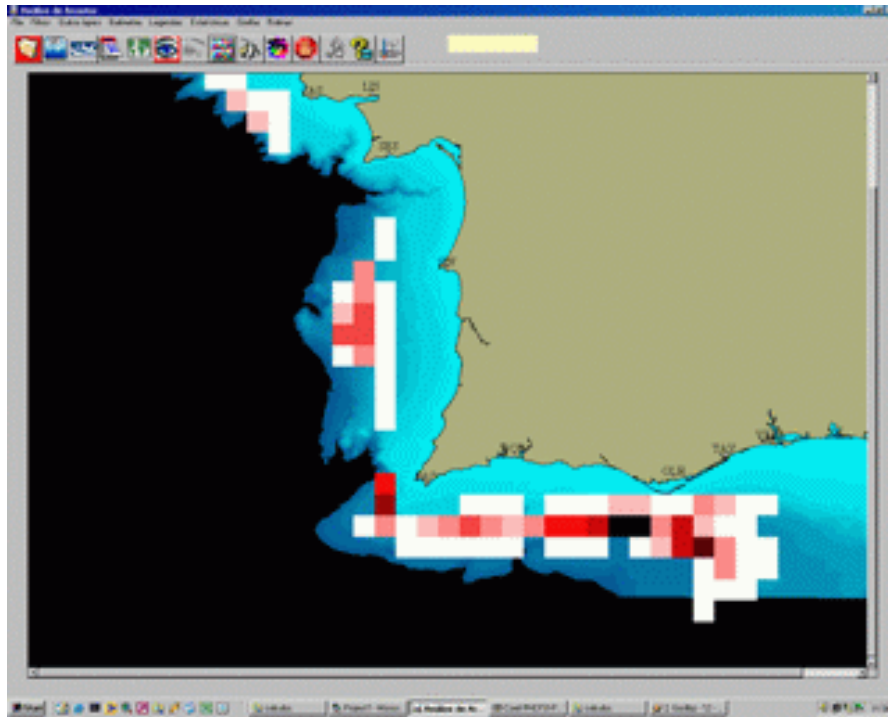
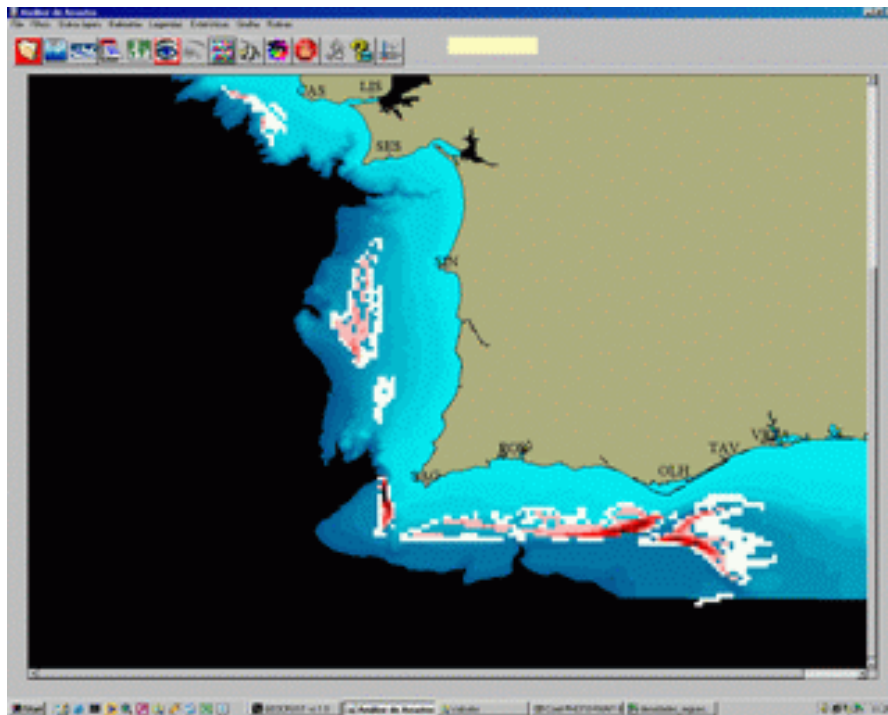


Figure 14: Marble <sup>4</sup> is a free and open source application that displays webmaps, along with other types of maps in a C++ application; source: <http://edu.kde.org/marble/screenshots/generic/marble-linkoeeping.png>;





(a) Density map of fishing effort, using a large grid;



(b) Density map of fishing effort, using a small grid;

Figure 15: Fishing effort maps generated by GeoCrust 2.0. To produce these maps the user can choose one or more vessels for a given period of time (quarter, semester, and year); VMS data from valid fishing trips <sup>41</sup>trawl hauls. Density grid size set by the user (minimum 0.06 x 0.06 nm);

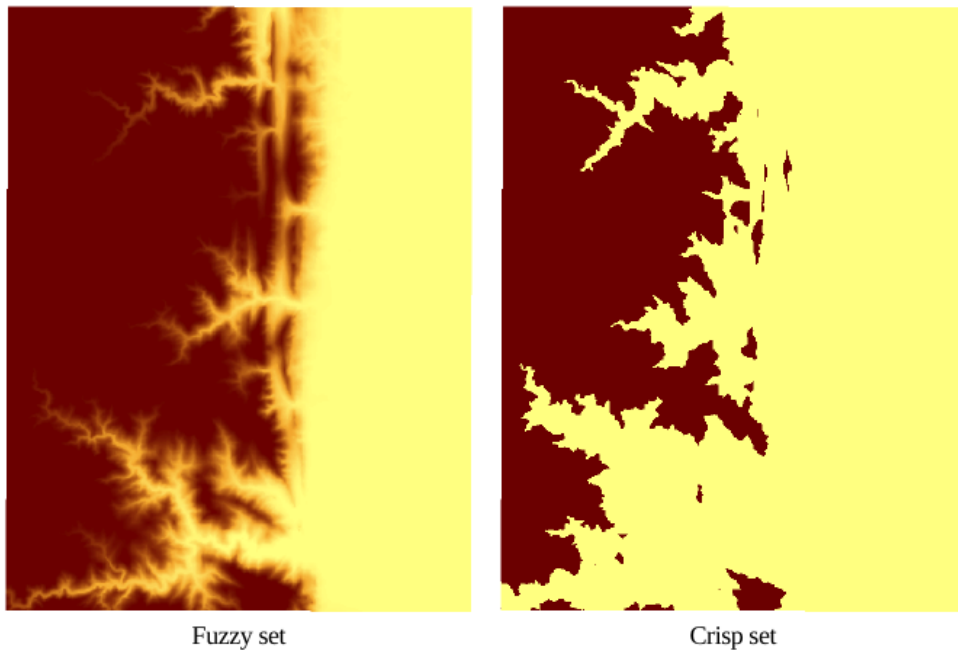


Figure 16: This is an illustrative example of an analysis with a fuzzy logic approach (on the left) and a crisp approach (on the right) [21];

## References

- [1] F. Degel, T. Jansen *FishFrame Fisheries and stock assessment data framework*. ICES CM 2006/M:02 , 2006.
- [2] F. Degel, T. Jansen *FishFrame Licensing*. Danish Institute for Fisheries Research Charlottenlund castle, Charlottenlund, Denmark. 5th March, 2006.
- [3] H. Kjems-Nielsen, L. Inger Larsen, M. Zarecki1, T. Jansen, B. Cowan, P. Sandbeck, M. Dueholm, O. Skov. *InterCatch - a tool for fish stock assessment, status and methods*. ICES CM 2006/M:29, 2006.
- [4] ICES. *InterCatch - User Manual..* Document version 1.9.
- [5] ICES *EcoSystemData3.0\_Release*. [http://ecosystemdata.ices.dk/EcoSystemData3.0\\_Release.pdf](http://ecosystemdata.ices.dk/EcoSystemData3.0_Release.pdf)
- [6] M. Stocker, H. Stiff, W. Shaw, A.W. Argue *The Canadian Albacore Tuna Catch and Effort Relational Database*. Canadian Technical Report of Fisheries and Aquatic Sciences 2701, 2007.
- [7] Mappamond *Fishery\_Analyst\_V2\_manual*. [http://www.mappamondogis.it/pdf/Fishery\\_Analyst\\_V2\\_manual.pdf](http://www.mappamondogis.it/pdf/Fishery_Analyst_V2_manual.pdf)
- [8] L.I. Muhoozi *Implementation of a Fisheries Management Plan (IFMP) Project for Lake Victoria: A Report Of The Fisheries Catch Assessment Survey In The Ugandan Waters Of Lake Victoria For The February 2008 Survey*. National Fisheries Resources Research Institute (NaFIRRI) and National Agricultural Research Organization (NARO), February, 2008.
- [9] P. A. Medley *Review Of The Data Collection And Management Systems Of The Marine Fisheries In The Dominican Republican: Final Report*.
- [10] N. Helm *A Report on The Market Survey of Reef and Lagoon Fish Catches In Western Samoa*. South Pacific Comission, SPC/lnshoreFish.Res./BP 30, 7 March 1988.
- [11] FAO. *Medfisis*, <http://www.faomedfisis.org/>.
- [12] JSON. *JSON*, <http://json.org/>.
- [13] SQLite. *SQLite*, <http://www.sqlite.org/>.
- [14] H.H. Shahri, A.A.Z. Barforush. *Data mining for removing fuzzy duplicates using fuzzy inference*. Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting, 27-30 June 2004. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1336319](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1336319)

- [15] C. López. *Quality of Geographic Data Detection of Outliers and Imputation of Missing Values*. PhD Dissertation on the Royal Institute of Technology, Department of Geodesy and Photogrammetry, Stockholm, Sweden 1997.
- [16] ICES. *Workshop on methods for merging metiers for fishery based sampling (WK-MERGE)* . ICES WKMERGE REPORT 2010, Copenhagen, Denmark, 19-22 January 2010.
- [17] R. Forrest, T. Pitcher, R. Watson, H. Valtýsson, S. Guénette *Estimating Illegal and Unreported Catches from Marine Ecosystems: Two Case Studies*. Sea Around Us: North Atlantic, Page 81, 19 DEC 2002.
- [18] ICES. *ICES Data Centre*, <http://www.ices.dk/datacentre/Submissions/index.aspx>.
- [19] J. Simoes, C. Pinto, M. Afonso-Dias. *Methodology for Monitoring and Management of the Crustacean Trawl Fleet. Example of GeoCrust 1.0 GIS* in Finisterra special edition Cartography and Geographic Information Systems, 2005.
- [20] D. Z. Sui *A Fuzzy Gis Modeling Approach Land Evaluation For Urban*. Environ. and Urban Systems, Vol. 16, pp. IOI-115, USA, 1992.
- [21] W. Kainz. *The mathematics of GIS*. <http://goo.gl/fb/CWk1R>
- [22] FAO/DANIDA. *Guidelines for the Routine Collection of Capture Fishery Data*. FAO FISHERIES TECHNICAL PAPER 382. Food and Agriculture Organization of the United Nations. Rome, 1999.
- [23] *fisheries\_data\_mng\_sys.sqlite*.
- [24] L. R. Garces, G. T. Silvestre, I. Stobutzki, F. C. Gayanilo Jr, F. Valdez, M. Saupi, T. Boonvanich, M. Roongratri, P. Thouc, Purwanto, I. Haroon, K. N. Kurup, M. Srinath, H. A. B. Rodrigo, M. D. Santos, F. S. B. Torres Jr., M. K. Tan, D. Pauly *A regional database management system—the fisheries resource information system and tools (FiRST): Its design, utility and future directions*. Fisheries Research, 78 (1-2). pp. 119-129.
- [25] IIP. *PESCAART: Documentacao tecnica*. 2006.
- [26] European Comission. *Common tool for raising and estimating properties of statistical estimates derived from the Data Collection Regulation (COST)*. Studies and Pilot projects for carrying out the common fisheries policy. Call for proposal FISH/2006/15 – lot 2.
- [27] COST Team and various contributors. *Package ‘COSTcore’ documentation*. February, 15, 2008.

- [28] B. U. Ibrahim, J. Auta, J. K. Balogun, *A Survey Of The Artisanal Fisheries Of Kontagora Reservoir, Niger State, Nigeria*. Bayero Journal of Pure and Applied Sciences, 2(1): 47 - 51. 2009.