

From data to performance:

What can we learn from production data ?

The FineFish development of a tool for data mining and benchmarking

Francesca Margiotta FEAP & Philippe Mack PEPITe

Bergen

March 2009



SIXTH FRAMEWORK
PROGRAMME





Some background information...

The FineFish project aims to generate new practical knowledge on how to reduce the incidence of malformations in the major species used in European Aquaculture and apply this to the professional sector.

- **LACK of HATCHERY PERFORMANCE DATA**
- ❑ How to measure technical and economic performance improvements?
- ❑ How to understand the underlying causes?
- ❑ How to transform this knowledge in best available practices?

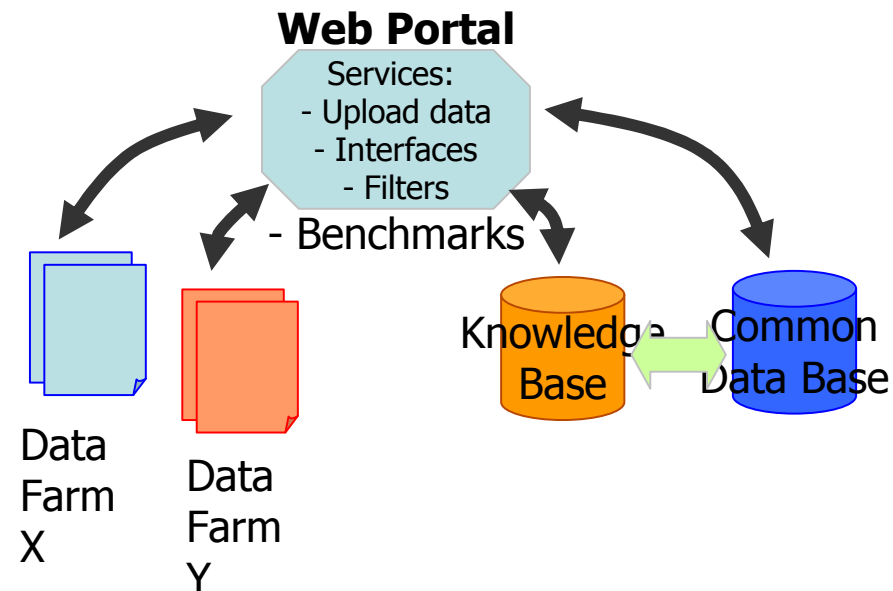
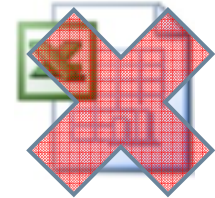


One of objective of FineFish is the systematic collection of hatchery data and the analysis of these data with regard to incidence of malformation in hatcheries and commercial fish farms.

- **Help aquaculture operators in:**
 - ❑ **Benchmarking their activities**
 - ❑ **Being able to share data and knowledge**
 - ❑ **Implement good practices**

A Professional Approach is necessary!

- The data collection for benchmarking started by using a relatively simple Excel worksheets for follow-up and reporting.
- Software exists on the market that can recover and analyse production parameters (Novafish/ Superior Systems/ Fishtalk...) but each SME has a different programme adapted to its reality.
 - **Need of a common platform for all fish producers**



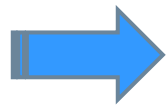


Benefits of a common platform

- **Having all data available stored in a single, standardized database will enable the comparison and benchmarking of data on production methodologies applied in the different hatcheries involved in the project.**
- **The following analysis of the data will enable the extraction of useful information and the improvement of current practices.**
- **The main goal is to identify key factors affecting production performance and the underlying causes of malformations onset.**

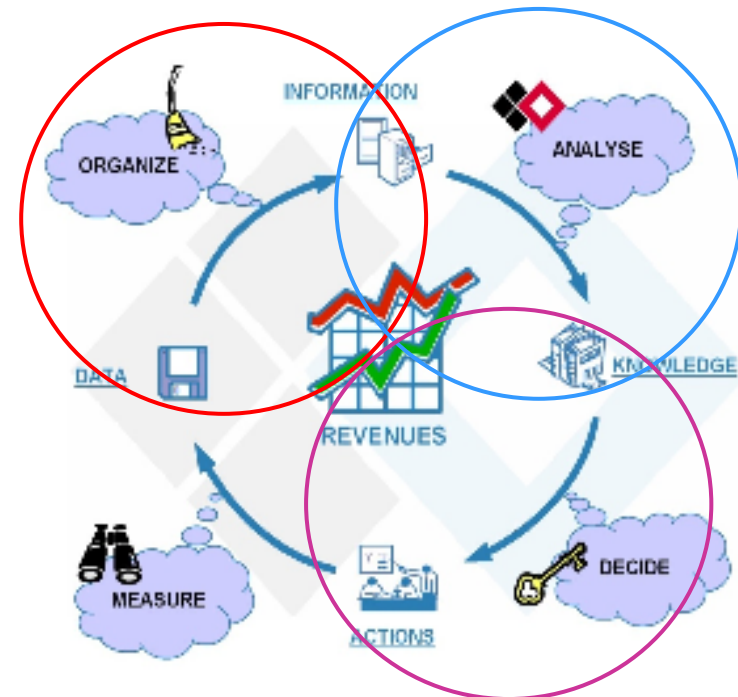


PEPITE is a company specialized in **DATA MINING**



Main objectives of PEPITE's work

- **develop a common database** with an accessible, **user friendly web interface** for the SME partners
- **apply data mining techniques to analyse production data**
- **apply predictive analysis**
(connects data to **effective action** by drawing reliable conclusions about current conditions and future events) to improve production performance.





Data mining

Is “the science of **extracting implicit, previously unknown, and useful information from large data sets** or databases”

Or “the **process of discovering meaningful new correlations, patterns and trends** by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.”

Methodology that aims to extract information from large databases, that is:

- ▣ Previously unknown
- ▣ Valid
- ▣ Comprehensible
- ▣ Useful

Wide range of tools

- ▣ Visualization, statistics, automatic learning from prediction models (forecasting models)



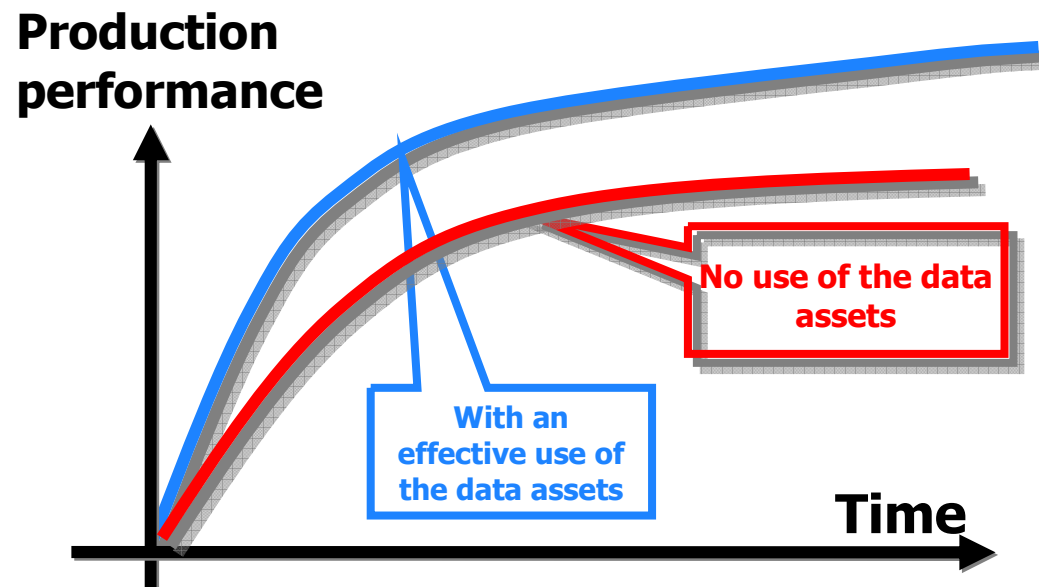
Benefits of better data usage

- **Understand the past**
 - Explain **key performance indicator** (malformation rate, growth rate,...) **behaviour**
 - Transform implicit knowledge into rules
 - Identify past conditions that improved production performance (in order to be able to reproduce it)
 - Identify process weaknesses and root causes of failure
- **Address the present**
 - Take a decision based on reliable KPI
 - Track process drifts (early detection of abnormal fluctuation in malformation rate, production performance)
- **Foresee the future**
 - Predict process states or KPI values – ideally “predictive model of malformation rate”
 - Predict maintenance actions – predict actions to improve performance



Benefits of better data usage

Understand the past → Address the present → Foresee the future





Key Performance Indicators



In a broad sense, a **key performance indicator (KPI)** is a tool for business improvement, focusing upon significant measurements within a company that indicate success or failure of that particular business.

Following a consultation with farm managers and technicians the main KPI identified in the scope of the FineFish project is  **MALFORMATION RATE x BATCH**

A KPI is a composite of the following:

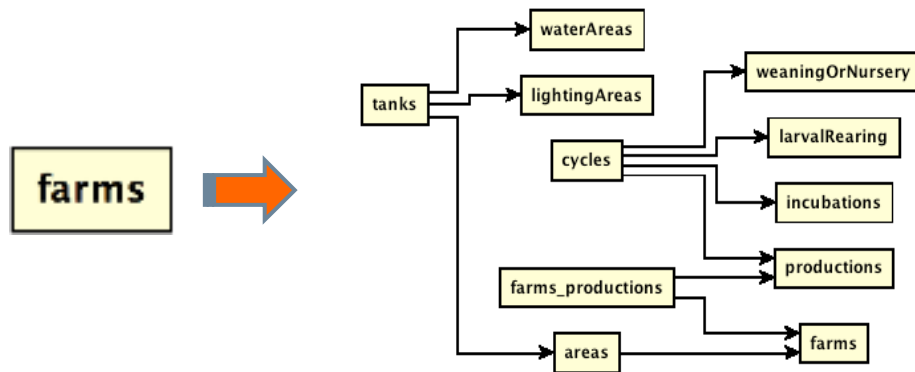
- a **measure of the performance** of specific goals that a business has defined to be of critical importance to their success → **malformation rate x batch**
- a **target** (or targets) → **set of a threshold value > than 10 %**
- an **action** resulting from that measurement → **corrective actions following the overtaking of the set threshold value**

The “making” of the database

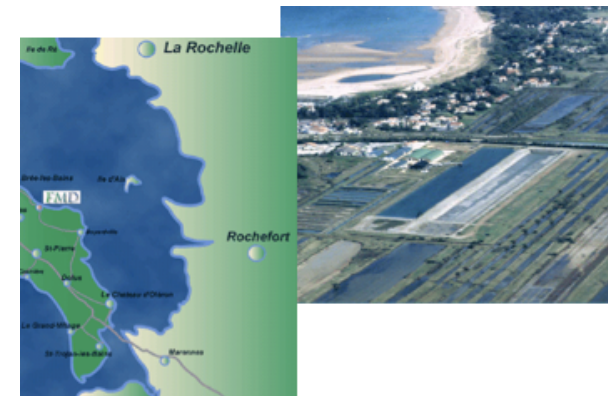
➤ TECHNICAL EXPERTISE WAS REQUIRED TO BUILD UP THE SYSTEM

After several days of observation and interviews in a test hatchery and detailed analysis of different hatchery structures and procedures, including interactions of the various areas (water area, tanks, lighting area...), LOGICAL RELATIONSHIPS between these components were used to **design a data model to be implemented in a RDBMS.**

A relational database is a database management system that is based on a relational model – data is stored in form of tables and the relationship among the data is also stored in form of tables.



The test hatchery: La Ferme Marine de Douhet – France





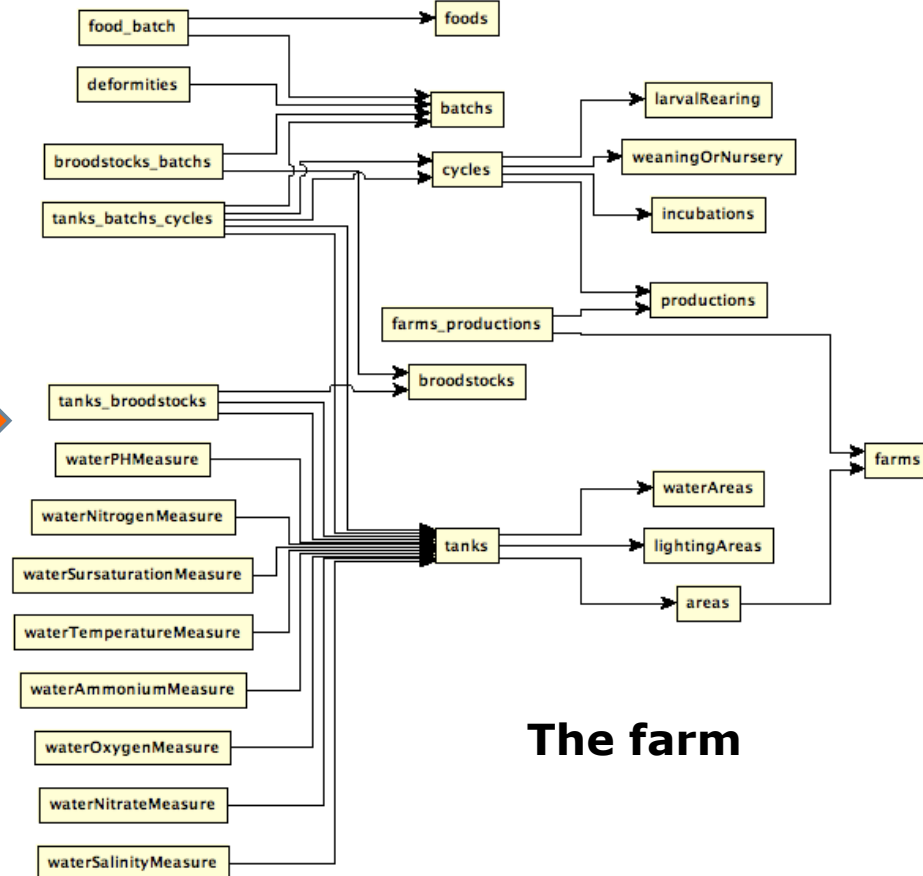
The making of the database

- ➔ **Analyse and model the hatchery production process**
 - Understand the farm structure and the production process
 - Audit the data collection on site
 - Discuss with sector experts the database business logic

- ➔ **Design a data model and implement it in a data warehouse**
 - A data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis

Data model of a farm and the production process

The real world



The farm

The model is based on the test farm "FMD" but is adapted to all farms, 'cause it is a model of the real world and of relationships that are in the real world.



Exploiting the database

Configure

Access to the database is made through a **WEB BASED INTERFACE** where a “user” can design, define and update a hatchery’s configuration.

You can import data on:

- Generic farm information
- Farm production (which sp.)
- Farm areas (water, lighting)
- Tanks
- Cycles
- Food

The screenshot shows a web browser window titled "Finefish Web-based Interface" with the URL "http://localhost:8888/finefish/farm/". The browser's address bar also shows "Google". The page has a navigation menu with links: [HOMEPAGE](#), [FARM](#), [MONITORING](#), [TRACKING](#), [BENCHMARKING](#), and [HELP](#). Below the menu, there is a welcome message: "Welcome demo (log out) | Control center". A breadcrumb trail is visible: [General Informations](#) > [Productions](#) > [Building / Zones](#) > [Water parameters](#) > [Lighting parameters](#) > [Tanks](#) > [Cycles Parameters](#) > [Food](#). The main content area is titled "GENERAL INFORMATIONS" and contains a form with the following fields: Company name (filled with "Demo Farm"), Acronym (filled with "DF"), Address (filled with "4010"), Postal code, City, Country, and Email address (filled with "b.stevens@pepite.be"). A "Submit" button is located at the bottom of the form. At the bottom of the browser window, there is a "Go to top" link and the text "Terminé".



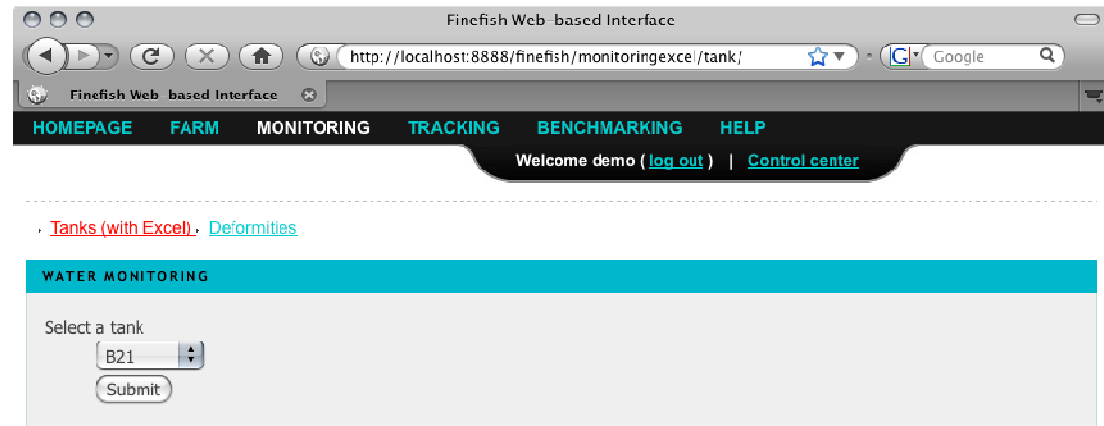
Exploiting the database

Monitoring

Once the configuration of your farm is complete...

You can import data on:

- Monitoring of single tanks (pH, T, [O₂], salinity...)



©2008 PEPITE S.A. (<http://www.pepите.be>) - Revision 52



Exploiting the database

Monitoring

You can import data on:

- Malformations x batch

The screenshot shows the 'Finefish Web-based Interface' with a navigation menu (HOMEPAGE, FARM, MONITORING, TRACKING, BENCHMARKING, HELP) and a user status bar (Welcome demo (log out) | Control center). The main content area is titled 'DEFORMITIES' and includes a table with columns 'BATCH (GRAM SIZE)' and 'ACTIONS (CREATE_NEW)'. The table contains one row for 'Sup (0 gram size)' with 'Modify' and 'Delete' actions. Below the table is a form for entering data, with a 'Batch' dropdown set to 'Sup'. The form includes 13 input fields for various deformities: Fish gram size, Operculum Shortened, Operculum Twisted, Operculum Missing left, Operculum Missing right, Jaw Lower Jaw, Jaw Upper Jaw, Jaw Cross bite, Fins Dorsal fin, Fins Caudal fin, Vertebral Lordosis, Vertebral Haemal lordosis, Juveniles with two or more deformities, and Total number of fish evaluated. A 'Submit' button is at the bottom of the form. A 'Go to top' link is visible at the bottom left of the page.



Exploiting the database

Tracking

You can enter data about broodstocks and batches and set their position (in tanks) and their cycle parameters (nursery, larval rearing...)

You can import data on:

- Broodstocks
- Batches
- Broodstocks' movements
- Batches' movements
- Broodstocks' food
- Batches' food

The screenshot shows a web browser window titled "Finefish Web-based Interface" with the URL "http://localhost:8888/finefish/batches". The interface has a navigation menu with "HOME PAGE", "FARM", "MONITORING", "TRACKING", "BENCHMARKING", and "HELP". Below the menu, there is a "Welcome demo (log out) | Control center" message. A breadcrumb trail shows: "Broodstocks > Batches > Broodstocks movements > Batches movements > Batches food > Broodstocks food".

BATCHES NOT DELIVERED	
NAME	ACTIONS
	<ul style="list-style-type: none">CREATE NEW FROM A BROODSTOCKCREATE NEW FROM (A) BATCH(ES)
C1	Modify Marked as delivered or deleted
C7	Modify Marked as delivered or deleted
B107	Modify Marked as delivered or deleted
B110	Modify Marked as delivered or deleted
B206	Modify Marked as delivered or deleted
B206	Modify Marked as delivered or deleted
Sup	Modify Marked as delivered or deleted
Moyen	Modify Marked as delivered or deleted
Inf	Modify Marked as delivered or deleted

Below the table, there is a section for "BATCHES DELIVERED OR DELETED" with a similar header and actions.

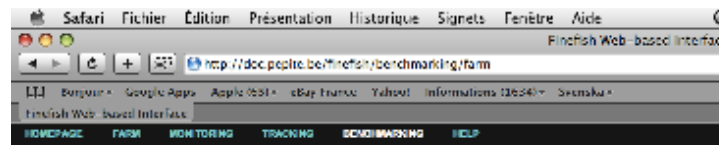


Exploiting the database

Benchmarking

You can benchmark data on:

- Farm structure



[View and download data](#)

Farm structure

[CLICK HERE TO DOWNLOAD DATA FILE](#)

TANKNAME	TANKCREATIONDATE	TANKCAPACITY	TANKSHAPE	TANKCOLOR	LIGHTINGNAME	LIGHTINGTYPE	LIGHTINGPHOTOCELL	LIGHTINGSOURCE	ARCNAME	WATERSOURCE	WATERFLOW
D107	2007-01-01 00:00:00	10000	Circular	Grey	Larval/Rearing	Artificial	Dark	Tube	Larval/Rearing	Sea water	Flow trough
K112	1987-01-01 01:00:00	10000	Circular	Grey	Larval/Rearing	Artificial	Dark	Tube	Larval/Rearing	Sea water	Flow trough
D200	2007-01-01 00:00:00	10000	Circular	Grey	Larval/Rearing	Artificial	Dark	Tube	Larval/Rearing	Sea water	Flow trough
K208	1987-01-01 01:00:00	10000	Circular	Grey	Larval/Rearing	Artificial	Dark	Tube	Larval/Rearing	Sea water	Flow trough
D21	2007-01-01 00:00:00	10000	Circular	Black	Deodorize	Artificial	Other	Tube	Deodorize	Sea water	Flow trough
K26	1987-01-01 01:00:00	10000	Circular	Black	Rear/Smolt	Artificial	Dark	Tube	Rear/Smolt	Sea water	Flow trough
D107	2007-01-01 00:00:00	52	Circular	White	Harvesting	Artificial	Dark	None	Incubation	Sea water	Flow trough
D112	1987-01-01 01:00:00	100	Circular	White	Harvesting	Artificial	Dark	None	Incubation	Sea water	Flow trough
Glasshouse	2008-09-01 00:00:00	50000	Cylindrical	Dark Green	Deodorize	Artificial	Other	Tube	Deodorize	Sea water	Flow trough
D110	1987-01-01 01:00:00	10000	Cylindrical	Dark Green	Rear/Smolt	Artificial	Dark	Tube	Rear/Smolt	Sea water	Flow trough
Glasshouse	2008-09-01 00:00:00	50000	Cylindrical	Dark Green	Deodorize	Artificial	Other	Tube	Deodorize	Sea water	Flow trough



Exploiting the database

Benchmarking

You can benchmark data on:

- Monitoring

ADDITIONAL DATA															
FARM ID	TARENAME	TANK ID	TIMESTAMP	MONITORING DATE	ADR	EMPAICAL TREATMENT (PRODUCT)	CUMULATIVE	CYCLE	DATE	DAY	DELTA DE (MAX-AIR)	DELTA DE (MAX-MIN)	DELTA PPT (MAX-AIR)	WORKING TRAPPERING (CC)	WORKING TRAPPERING (CC)
16	K21	34	13862080	2007-12-26 00:00:00	0	0	0	0	0	0	1	0	0	0	21.7
15	D01	04	13861000	2007-12-27 00:00:00	0	0	0	0	0	0	2	0	0	0	19.7
16	K21	34	13861940	2007-12-26 00:00:00	0	0	0	0	0	0	3	0	0	0	20.7
15	D01	04	138608200	2007-12-29 00:00:00	0	0	0	0	0	0	4	0	0	0	22.4
16	K21	34	13860800	2007-12-29 00:00:00	0	0	0	0	0	0	5	0	0	0	21.1
15	D01	04	138605200	2007-12-31 00:00:00	0	0	0	0	0	0	6	0	0	0	21.0
16	K21	34	13861400	2008-01-02 00:00:00	0	0	0	0	0	0	7	0	0	0	20.6
15	D01	04	138620400	2008-01-02 00:00:00	0	0	0	0	0	0	8	0	0	0	20.1
16	K21	34	13861080	2008-01-02 00:00:00	0	0	0	0	0	0	9	0	0	0	20.6
15	D01	04	138611200	2008-01-04 00:00:00	0	0	0	0	0	0	10	0	0	0	19.0
16	K21	34	13860800	2008-01-05 00:00:00	0	0	0	0	0	0	11	0	0	0	19.0
15	D01	04	138627400	2008-01-05 00:00:00	0	0	0	0	0	0	12	0	0	0	15.0
16	K21	34	13860840	2008-01-05 00:00:00	0	0	0	0	0	0	13	0	0	0	19.7

...and actually all available data!



Exploiting the database

Output are tables containing data in an organized structure

The screenshot shows an OpenOffice Calc spreadsheet with a table of tank data. The table has 11 columns: tankName, tankCreationDate, tankCapacity, tankShape, tankColor, lightingName, lightingType, lightingPhotoPeriod, lightingSource, and area. The data is organized into rows, with the first row being the header. The table contains 26 rows of data, including various tank types like B107, B110, B206, B208, B21, B23, C107, C110, G1seabream, G2seabream, G3seabass, G4seabream, G5seabream, G6seabass, G7seabream, G8seabream, G9seabass, I10, I1bis, I2, I3, InfTank, and Mosaic.

	A	B	C	D	E	F	G	H	I	J
1	tankName	tankCreationDate	tankCapacity	tankShape	tankColor	lightingName	lightingType	lightingPhotoPeriod	lightingSource	area
2	B107	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	Larv
3	B110	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	Larv
4	B206	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	Larv
5	B208	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	Larv
6	B21	01/01/07 00:00	10000	Circular	Black	Broodstock	Artificial	Other	Tube	Broo
7	B23	01/01/07 00:00	10000	Circular	Black	Broodstock	Artificial	Other	Tube	Broo
8	C107	01/01/07 00:00	125	Conical	White	Hatchingtank	Artificial	Dark	None	Incu
9	C110	01/01/07 00:00	125	Conical	White	Hatchingtank	Artificial	Dark	None	Incu
10	G1seabream	01/09/08 00:00	18tons	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
11	G2seabream	01/01/07 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
12	G3seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
13	G4seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
14	G5seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
15	G6seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
16	G7seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
17	G8seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
18	G9seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broo
19	I10	18/01/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artificial	Dark	None	Hati
20	I1bis	29/02/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artificial	Dark	None	Hati
21	I2	14/05/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artificial	Dark	None	Hati
22	I2	05/11/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artificial	Dark	None	Hati
23	I3	11/11/07 00:00	1950	Cylindroconical	Black	Hatchingtank	Artificial	Dark	None	Hati
24	I3	05/11/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artificial	Dark	None	Hati
25	InfTank	01/01/07 00:00	20000	Circular	Grey	Weaningtank	Artificial	Other	Tube	Nun
26	Mosaic	01/01/07 00:00	1000000	Conical	White	Broodstock	Artificial	Other	Tube	Out



Exploiting the database

Depending on what information farmers want to extract from the database queries can be applied.

Queries are precise requests for information retrieval within a database and information system.

Ex. I want to see all fish species produced in the farms in the database!

name	specie
Ferme marine du douhet	Sea bream
Viviers de france	Trout
A very big farm	Salmon
A very big farm	Sea bream
A very big farm	Trout

Purely as a technical indication, PEPITE used this query to create the result:
SELECT farms.name, productions.specie
FROM farms
LEFT JOIN farms_productions ON farms.id = farms_productions.id_farm
LEFT JOIN productions ON farms_productions.id_production = productions.id;



Exploiting the database

QUERYS

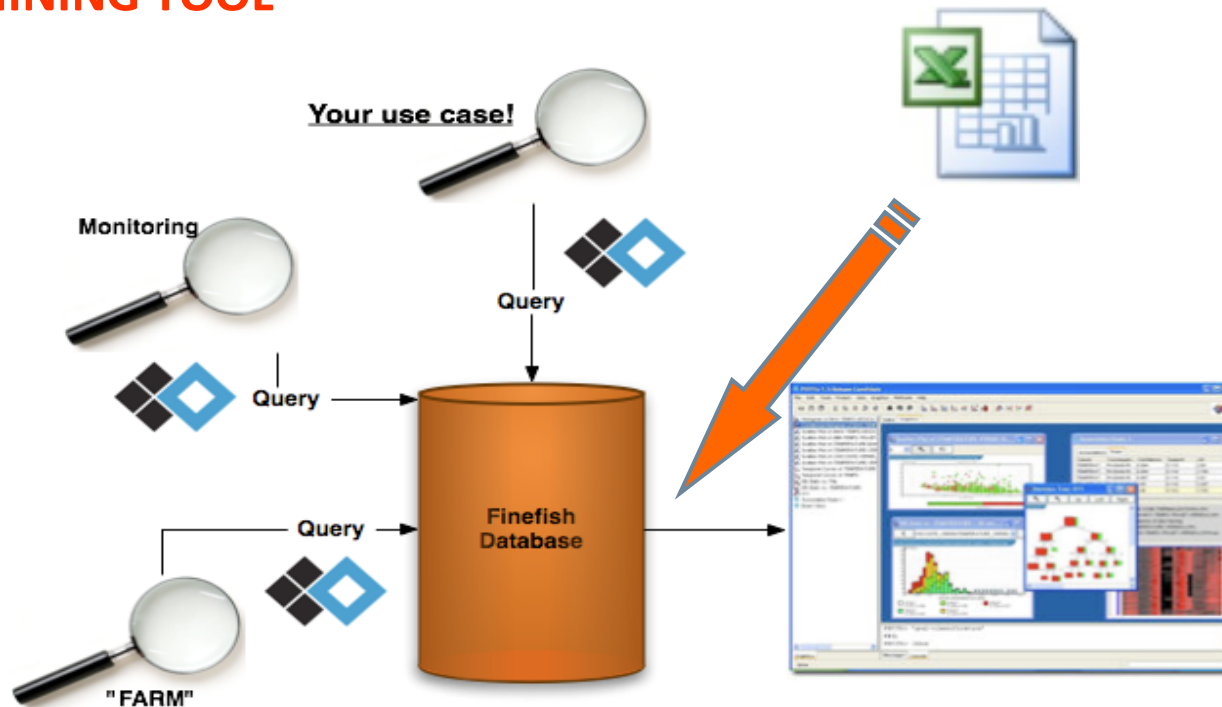
In the same way, through a query we can retrieve batch paths in the farm and obtain a table like this one:

identifier	tank_departure	tank_destination	cycle_from	cycle_to
batch one	C1	B110	Incubation	Larval rearing
batch two	C7	B107	Incubation	Larval rearing



Data analysis - PEPITo

A DATA MINING TOOL



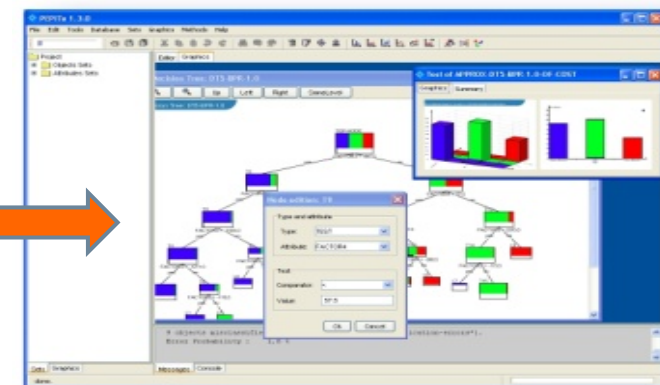
The result of a query is a table, a CSV file that can be easily converted in an Excel sheet and uploaded in the analitical software PEPITo for further analysis.



Data analysis

to be performed with the SOFTWARE PEPITO

- Data validation and filtering
- Data transformation: FFT, sampling,...
- Data visualisation: distribution plots, scatter plots, temporal curves,...
- Statistical analysis: analysis of variance, correlations analysis,...
- Predictive analysis: neural networks, decision trees, association rules,...





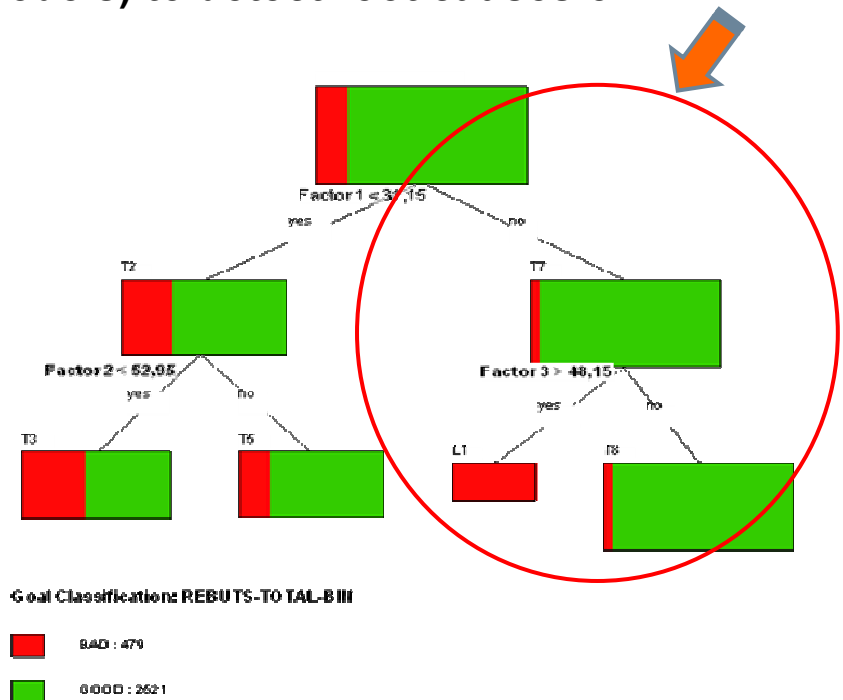
Data analysis - PEPITo

PREDICTIVE AND ROOT CAUSE ANALYSIS

Once enough data will be stored in the system, we will be able to apply predictive analysis tools (like decisions trees models) to detect root causes of malformation rate.

Possible analysis:

- detect in a farm the parameters explaining the malformation rate drift between two production cycles (in this case conclusion would probably be specific to the farm)
- detect in the whole set of farm recorded in the database why malformation rate is higher in some farms' production (in this case we can expect that the conclusion would be broader, and that the improvements actions could be applied to every farm)





PEPITe & FEAP working with Fish Farms

VALIDATING DATA MODEL & COLLECTION PROCESS

5 hatcheries have configured their structural inputs and are registering production cycle data (in France, Italy, Spain, Norway and Israel).

PEPITe and FEAP are working in synergy with these farms:

- Giving individual support and training to upload data and perform analysis.
- Using the experience and knowledge of their technical and scientific staff so as to improve the system

Through farms' feedback it will be possible to:

- Improve the web interface and make it more user-friendly and responsive to farmer's necessities
- Identify possible bugs in the database and correct these in order to enable and facilitate a correct and easy input of data
- Modify the database and the data organization in a way to enable the creation of specific queries to be analysed through the data mining tool.



FEEDBACK from HATCHERIES

Some feedback on queries:

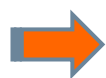
- Panittica pugliese SpA - Italy (Sea bass and Sea bream):

“we want to be able to keep track of parameters such as T, pH, salinity, food quality and quantity (fed to fish larvae) per tank in time.”

- Bolaks AS - Norway (salmon):

“we want to be able to keep track of different light regimes in tanks in time”

“we want to keep track in time of the different treatments reserved to fish in tanks. ”



These are requests that can easily be achieved through the database and the data mining software!



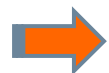
APPLICATIONS

An example:

- THE DATABASE RECORDS ALL MOVEMENTS OF FISH FROM ONE TANK TO ANOTHER and THE DIFFERENT CONDITIONS IN EACH TANK VARYING IN TIME.

FOR EVERY SINGLE BATCH OF FISH PRODUCED BY THE HATCHERY IT IS POSSIBLE TO EXTRACT INFORMATION FROM HISTORICAL DATA REGARDING CHANGE in TIME of:

- **POSITION (tank id)**
- **TREATMENTS (chemical treatment, antibiotics...)**
- **MONITORING PARAMETERS (T, pH, light...)**
- **FOOD (rotifers, algae, different feeds..)**



This is very important since these are the variables potentially influencing malformation incidence!



POTENTIAL of THE SYSTEM

- **discover unexpected correlations between parameters**
- **benchmark different farms and point out good practices (BMP)**
- **verify ad hoc knowledge with historical data in order to early detect abnormal situations**
- **expand the system to other KPIs (identify new KPI)**
- **include genetic information in the system**
- **automate the creation of reports**
- **automate the import of data in the system collected through other software available on the market**