

## From data collection to improving performance

### What can we learn from production data ?

The FineFish development of a new tool for benchmarking and data mining

Philippe Mack - PEPITe

Francesca Margiotta & Courtney Hough - FEAP

*Athens*

*May 2009*





## 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 of problems?
  - ❑ How to transform this knowledge into best practice?

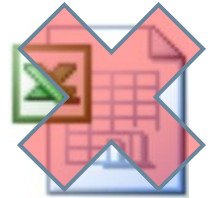


One of the objectives of FineFish is the systematic collection of hatchery data and the analysis of these with regard to the incidence of malformations in hatcheries and commercial fish farms.

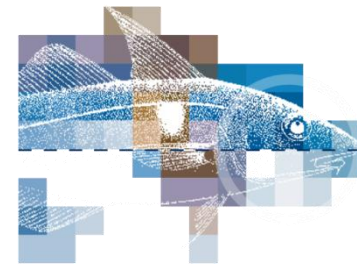
- **Help aquaculture operators to:**
  - ❑ **Benchmark** their activities/procedures
  - ❑ Be able to **share data and knowledge**
  - ❑ Implement **best practice** for their benefit

## A Professional Approach is necessary!

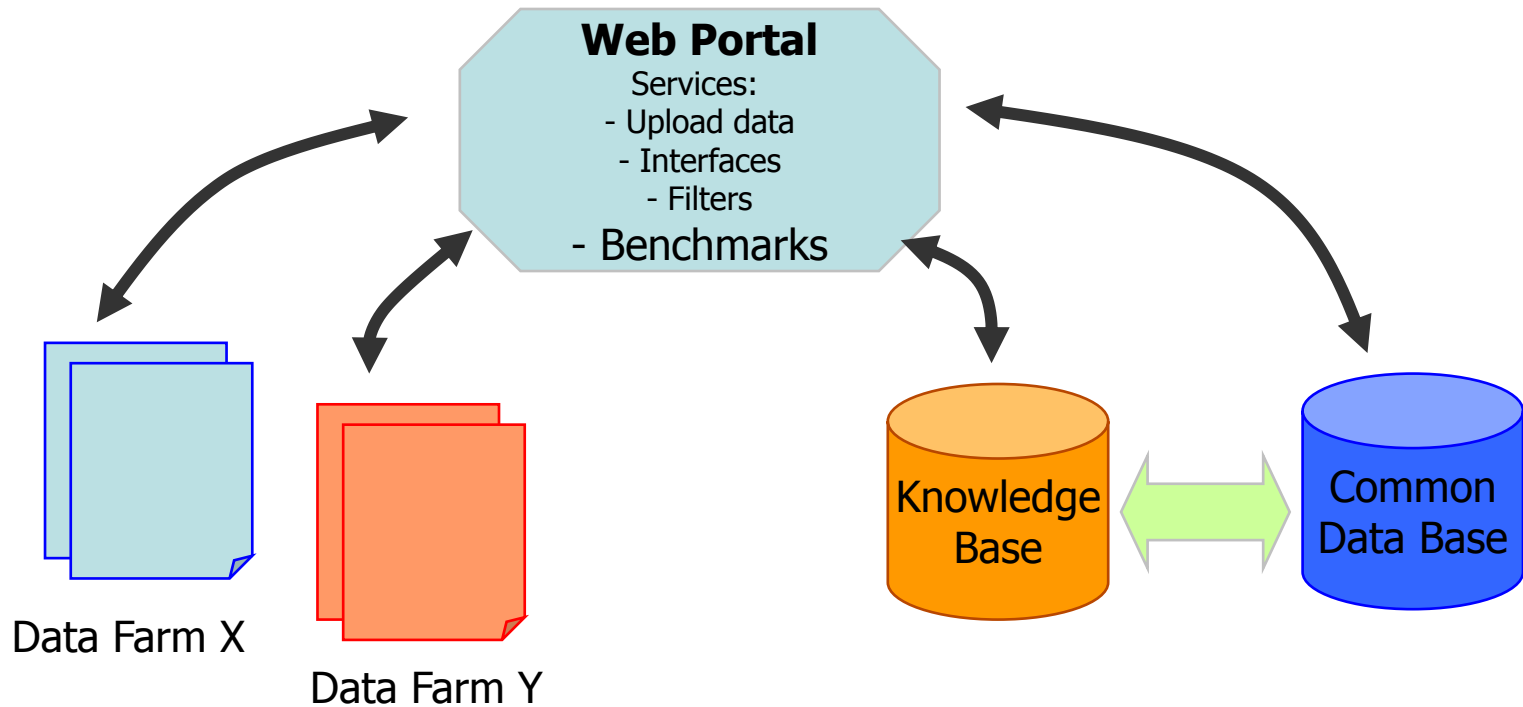
- The data collection for benchmarking started by using a relatively simple Excel worksheets for follow-up and reporting.
  - But Excel has its limits
- Software exists on the market that can recover and analyse production parameters (Novafish/ Superior Systems/ Fishtalk...)
  - But each SME has a different programme adapted to its own circumstances and realities.



➤ **Need for a common base  
(or platform)  
for all operators**

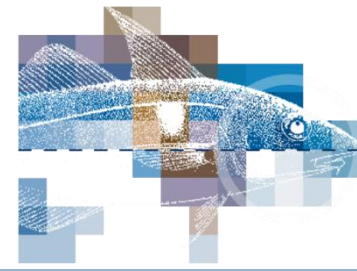


# A Professional Approach is necessary!



## Benefits of a common platform

- Having **all data available** stored in a single, standardised database will enable the **comparison** and **benchmarking** of data on production methodologies applied in the different hatcheries involved in the project.
- The ensuing 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 **causal factors** of malformations.

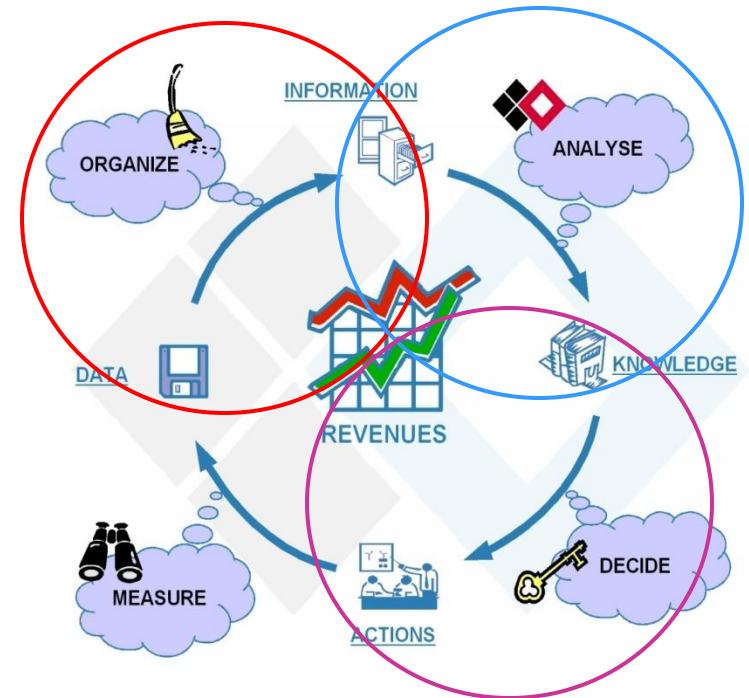


**PEPITe** is a company specialized in **DATA MINING**



### 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**  
(connecting 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:

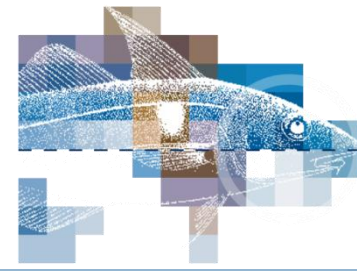
- ▣ Previously unknown
- ▣ Valid
- ▣ Comprehensible
- ▣ Useful

Wide range of tools available

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

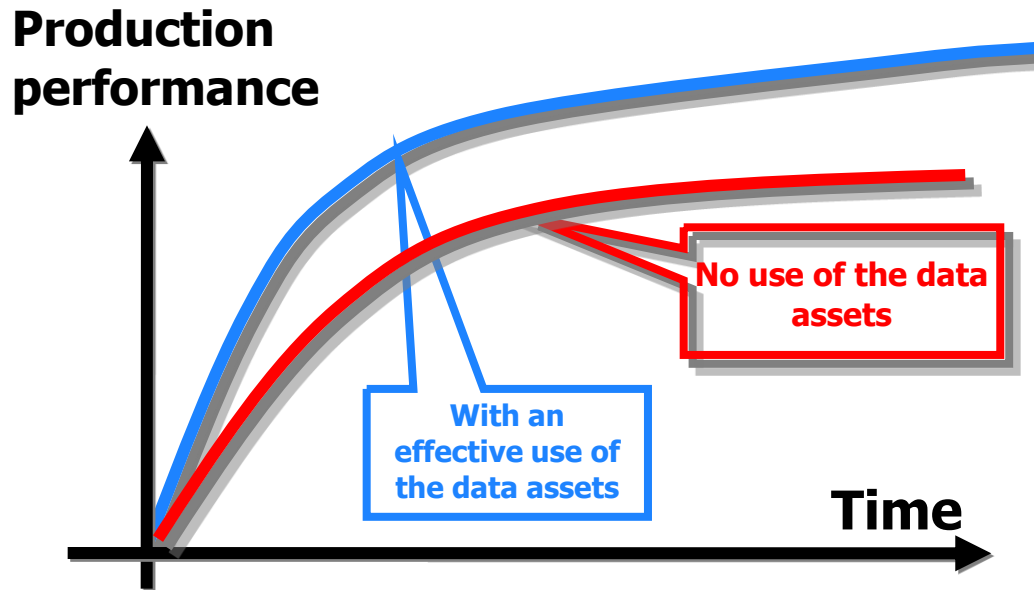
# Benefits of better data usage

- **Understand the past**
  - Explain **behaviour of key performance indicators [KPI]** (e.g. malformation rates, growth rates,...)
  - Transform **implicit knowledge** into **rules of procedure**
  - Identify past conditions that improved production performance (so as to reproduce these consistently)
  - Identify **process weaknesses** and **root causes of failure**
- **Address the present**
  - Take decisions based on **reliable KPI**
  - Track **process drifts** (early detection of abnormal fluctuations in malformation rates, production performance)
- **Foresee the future**
  - **Predict process states or KPI values** – ideally develop a “predictive model of malformation rate”
  - **Predict maintenance actions and 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 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:

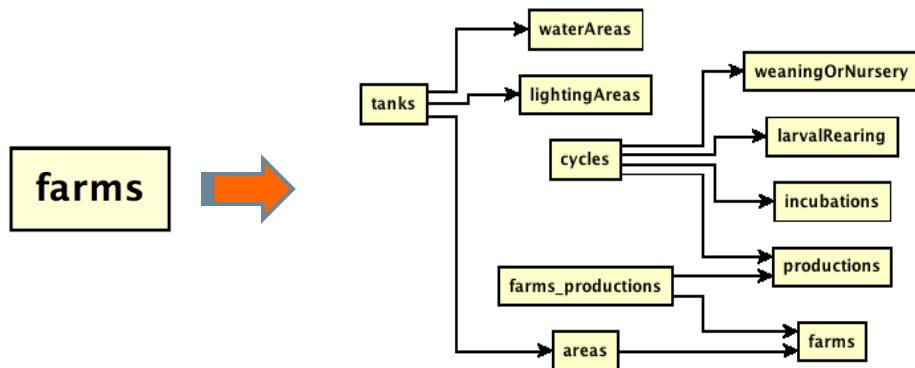
- a **measure of the performance** of specific goals that a business has defined to be of critical importance to their success → **malformation rate x batch**
- a **target** (or targets) → **set of a threshold value > than 10 %**
- an **action** resulting from that measurement → **corrective actions 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(s) between the data is also stored in form of tables.



Test hatchery: La Ferme Marine de Douhet – France

## The “making” of the database

### → Analyse and model the hatchery production process

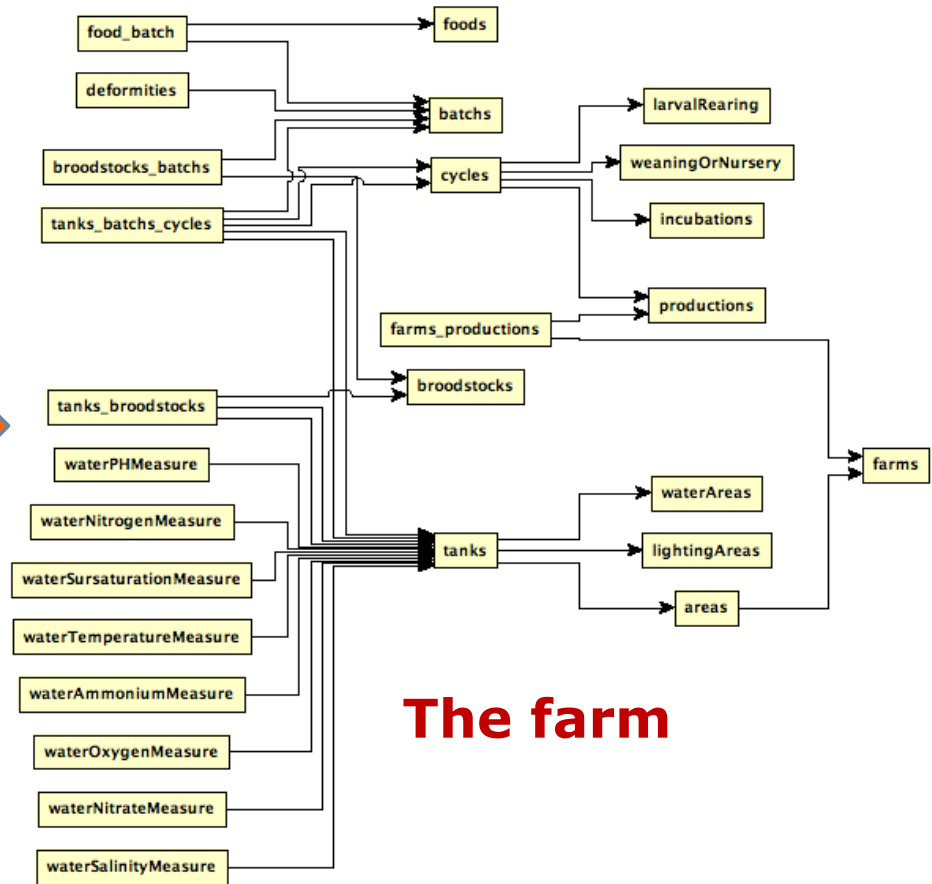
- Understand the farm structure and the production process
- Audit the data collection on site
- Discuss the database business logic with sector experts

### → Design a data model and implement it in a data “warehouse”

- A data “warehouse” is a repository of an organisation'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 adaptable to all farms, because it is a model of the real world and its relationships with reality....

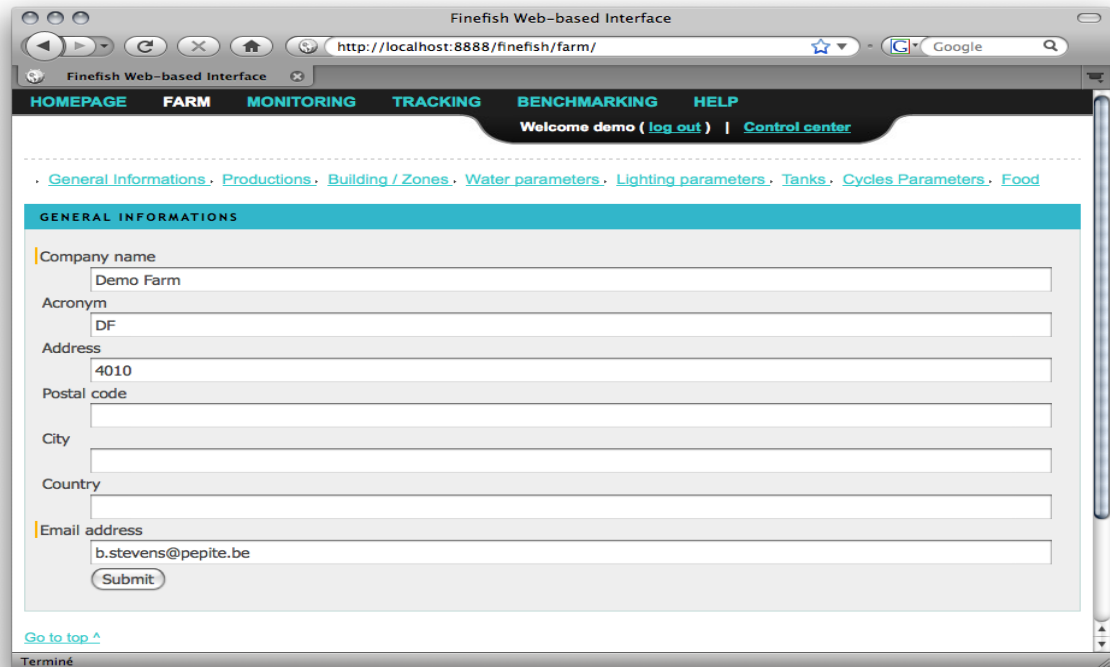
# Exploiting the database

## Configuration

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

## Data can be included on:

- Generic farm information
- Farm production (which sp.)
- Farm ‘area’ conditions  
e.g. Water parameters, lighting
- Tanks, ponds....
- Cycles of production
- Feeds....



The screenshot shows a web browser window titled "Finefish Web-based Interface" with the URL "http://localhost:8888/finefish/farm/". The page has a navigation menu with links: HOMEPAGE, FARM, MONITORING, TRACKING, BENCHMARKING, and HELP. Below the menu, it says "Welcome demo (log out) | Control center". A breadcrumb trail includes: General Informations, Productions, Building / Zones, Water parameters, Lighting parameters, Tanks, Cycles Parameters, and Food. The main content area is titled "GENERAL INFORMATIONS" and contains a form with the following fields:

- Company name: Demo Farm
- Acronym: DF
- Address: 4010
- Postal code: (empty)
- City: (empty)
- Country: (empty)
- Email address: b.stevens@pepите.be

A "Submit" button is located below the email address field. At the bottom left of the page, there is a "Go to top ^" link and the word "Terminé" in the footer.

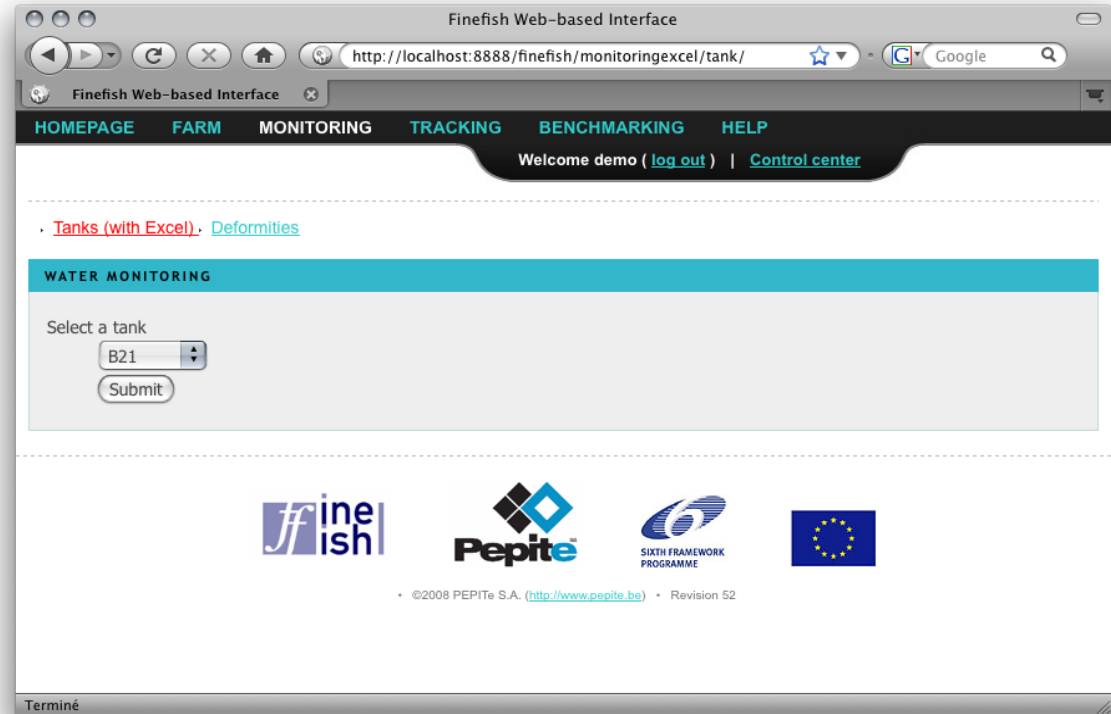
# Exploiting the database

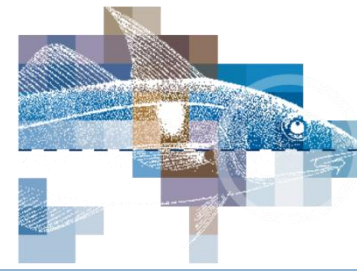
## Monitoring

Once the configuration of your farm is complete...

### You can import data on:

- Monitoring parameters of single units (tanks)... pH, Temp, O<sub>2</sub>, salinity...



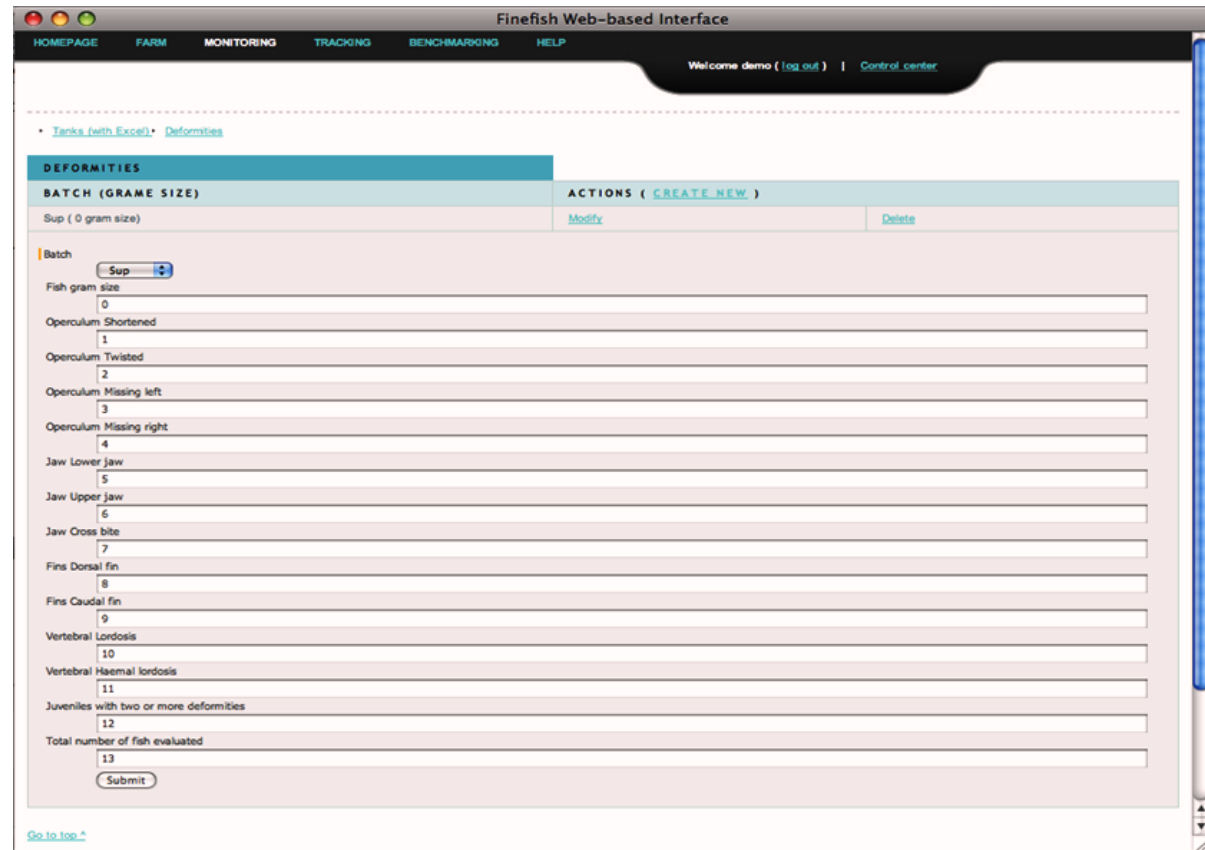


# Exploiting the database

## Monitoring

You can import data on:

- Malformations x batch
- Type of malformation



Finefish Web-based Interface

HOME PAGE FARM MONITORING TRACKING BENCHMARKING HELP

Welcome demo (log out) | Control center

• Tanks (with Excel) • Deformities

BATCH (GRAM SIZE)	ACTIONS (CREATE NEW)
Sup ( 0 gram size)	Modify Delete

Batch:

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:

Total number of fish evaluated:

[Go to top ^](#)

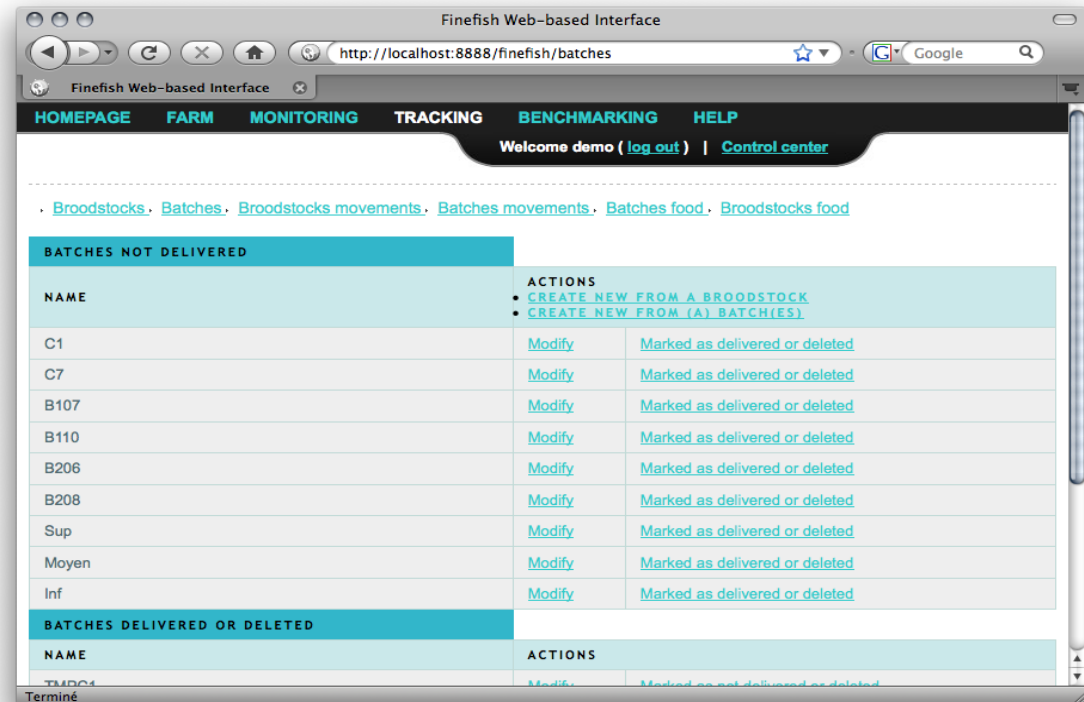
# Exploiting the database

## 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:

- Broodstock
- Batches
- Broodstock movements
- Batch movements
- Broodstock food/diet
- Batch food/diet

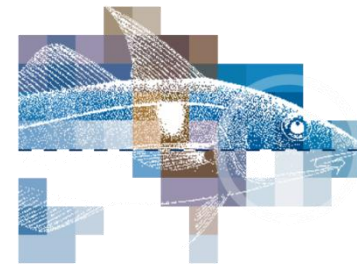


The screenshot shows a web browser window titled "Finefish Web-based Interface" with the URL "http://localhost:8888/finefish/batches". The interface has a navigation menu with "HOME PAGE", "FARM", "MONITORING", "TRACKING", "BENCHMARKING", and "HELP". Below the menu, there is a "Welcome demo (log out)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

The main content area displays a table titled "BATCHES NOT DELIVERED". The table has two columns: "NAME" and "ACTIONS". The "ACTIONS" column contains two links: "CREATE NEW FROM A BROODSTOCK" and "CREATE NEW FROM (A) BATCH(ES)".

NAME	ACTIONS
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
B208	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 titled "BATCHES DELIVERED OR DELETED" with a similar table structure. The first row shows "NAME" and "ACTIONS". The second row shows "TMO4" and "Modify | Marked as not delivered or deleted".

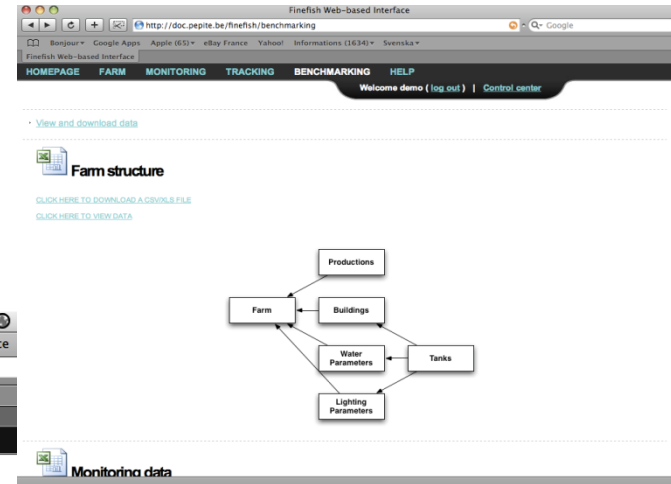
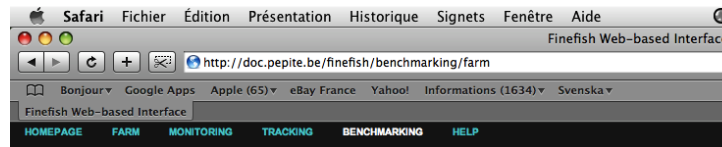


# Exploiting the database

## Benchmarking

You can benchmark data on:

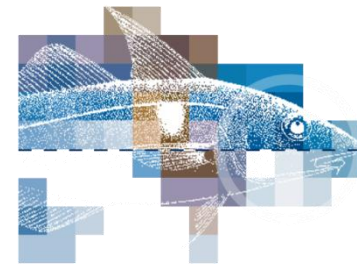
Farm structure



[View and download data](#)

**Farm structure**  
[CLICK HERE TO DOWNLOAD A CSV/XLS FILE](#)

FARM												
TANKNAME	TANKCREATIONDATE	TANKCAPACITY	TANKSHAPE	TANKCOLOR	LIGHTINGNAME	LIGHTINGTYPE	LIGHTINGPHOTOPERIOD	LIGHTINGSOURCE	AREANAME	WATERSOURCE	WATERSY	
B107	2007-01-01 00:00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	LarvalRearing	Sea water	Flow trough	
B110	2007-01-01 00:00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	LarvalRearing	Sea water	Flow trough	
B206	2007-01-01 00:00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	LarvalRearing	Sea water	Flow trough	
B208	2007-01-01 00:00:00	10000	Circular	Grey	LarvalRearing	Artificial	Dark	Tube	LarvalRearing	Sea water	Flow trough	
B21	2007-01-01 00:00:00	10000	Circular	Black	Broodstock	Artificial	Other	Tube	Broodstock	Sea water	Flow trough	
B23	2007-01-01 00:00:00	10000	Circular	Black	Broodstock	Artificial	Other	Tube	Broodstock	Sea water	Flow trough	
C107	2007-01-01 00:00:00	125	Conical	White	Hatchingtank	Artificial	Dark	None	Incubation	Sea water	Flow trough	
C110	2007-01-01 00:00:00	125	Conical	White	Hatchingtank	Artificial	Dark	None	Incubation	Sea water	Flow trough	
G1seabream	2008-09-01 00:00:00	18tons	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broodstock	Sea water	Flow trough	
G2seabream	2007-01-01 00:00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broodstock	Sea water	Flow trough	
G3seabass	2008-11-05 00:00:00	18tons	Cylindroconical	Dark Green	Broodstock	Artificial	Other	Tube	Broodstock	Sea water	Flow trough	



# Exploiting the database

## Benchmarking

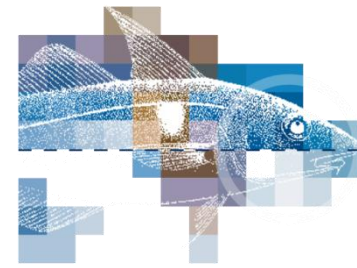
You can benchmark data on:

Monitoring data

The screenshot shows the 'Monitoring data' section of the Finefish Web-based Interface. It includes a navigation menu, a welcome message, and a table of monitoring data.

FARMID	TANKNAME	TANKID	TIMESTAMP	MONITORINGDATE	AGE	CHEMICAL TREATMENT (PRODUCT)	CUMMULATIVE	CYCLE	DATE	DAY	DELTA O2 (MAX-MIN)	DELTA OC (MAX-MIN)	DELTA PPT (MAX-MIN)	EVENING TEMPERATURE (OC)	MORNING TEMPERATURE (OC)
15	B21	34	1198623600	2007-12-26 00:00:00	0	0	0	0	0	1	0	0	0	0	21.2
15	B21	34	1198710000	2007-12-27 00:00:00	0	0	0	0	0	2	0	0	0	0	19.7
15	B21	34	1198796400	2007-12-28 00:00:00	0	0	0	0	0	3	0	0	0	0	20.2
15	B21	34	1198882800	2007-12-29 00:00:00	0	0	0	0	0	4	0	0	0	0	22.4
15	B21	34	1198969200	2007-12-30 00:00:00	0	0	0	0	0	5	0	0	0	0	21.1
15	B21	34	1199055600	2007-12-31 00:00:00	0	0	0	0	0	6	0	0	0	0	21.3
15	B21	34	1199142000	2008-01-01 00:00:00	0	0	0	0	0	7	0	0	0	0	20.6
15	B21	34	1199228400	2008-01-02 00:00:00	0	0	0	0	0	8	0	0	0	0	20.1
15	B21	34	1199314800	2008-01-03 00:00:00	0	0	0	0	0	9	0	0	0	0	20.3
15	B21	34	1199401200	2008-01-04 00:00:00	0	0	0	0	0	10	0	0	0	0	18.8
15	B21	34	1199487600	2008-01-05 00:00:00	0	0	0	0	0	11	0	0	0	0	16.9
15	B21	34	1199574000	2008-01-06 00:00:00	0	0	0	0	0	12	0	0	0	0	15.8
15	B21	34	1199660400	2008-01-07 00:00:00	0	0	0	0	0	13	0	0	0	0	17.2

...and actually all available data!



# Exploiting the database

Output are tables containing data in an organized structure

The screenshot shows the OpenOffice.org Calc application window with the file 'farm.csv' open. The table contains the following data:

	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	Artifical	Dark	Tube	Lary
3	B110	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artifical	Dark	Tube	Lary
4	B206	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artifical	Dark	Tube	Lary
5	B208	01/01/07 00:00	10000	Circular	Grey	LarvalRearing	Artifical	Dark	Tube	Lary
6	B21	01/01/07 00:00	10000	Circular	Black	Broodstock	Artifical	Other	Tube	Broo
7	B23	01/01/07 00:00	10000	Circular	Black	Broodstock	Artifical	Other	Tube	Broo
8	C107	01/01/07 00:00	125	Conical	White	Hatchingtank	Artifical	Dark	None	Incu
9	C110	01/01/07 00:00	125	Conical	White	Hatchingtank	Artifical	Dark	None	Incu
10	G1seabream	01/09/08 00:00	18tons	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
11	G2seabream	01/01/07 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
12	G3seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
13	G4seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
14	G5seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
15	G6seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
16	G7seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
17	G8seabream	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
18	G9seabass	05/11/08 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Broo
19	I10	18/01/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artifical	Dark	None	Hate
20	I1bis	29/02/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artifical	Dark	None	Hate
21	I2	14/05/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hate
22	I2	05/11/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hate
23	I3	11/11/07 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hate
24	I3	05/11/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hate
25	InfTank	01/01/07 00:00	20000	Circular	Grey	Weaningtank	Artifical	Other	Tube	Nur
26	Marais	01/01/07 00:00	10000000	Conical	White	Broodstock	Artifical	Other	Tube	Quit

# 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

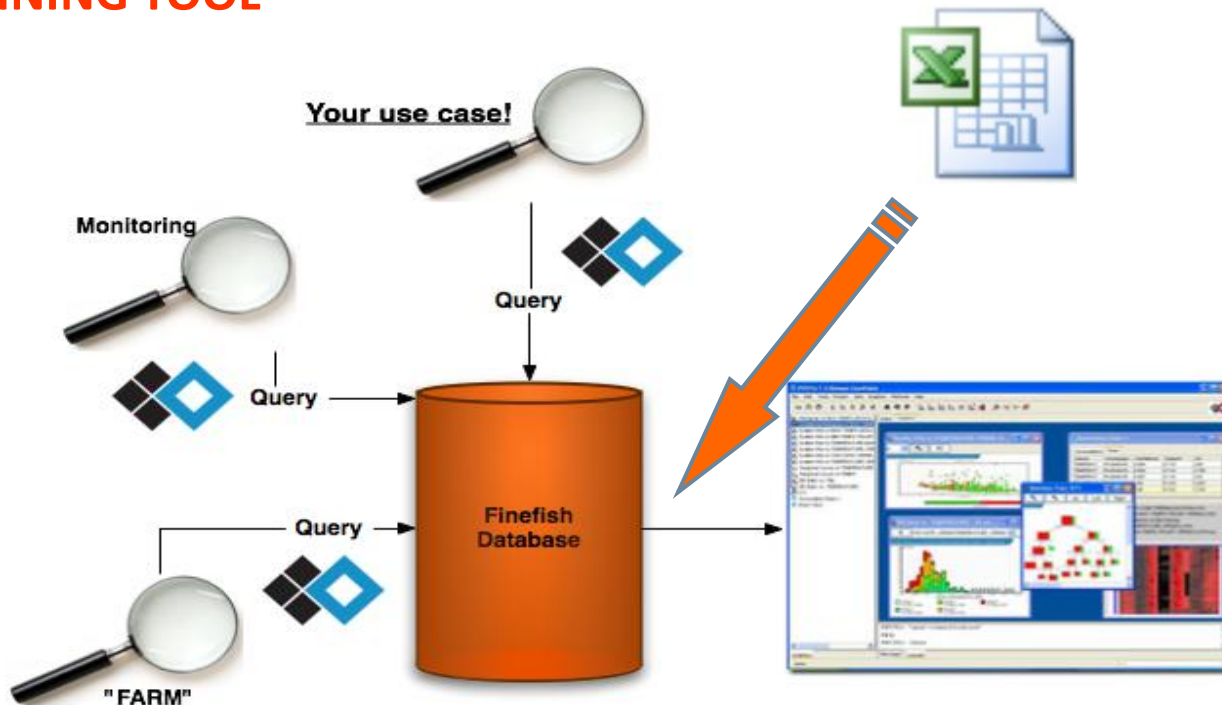
## QUERIES

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

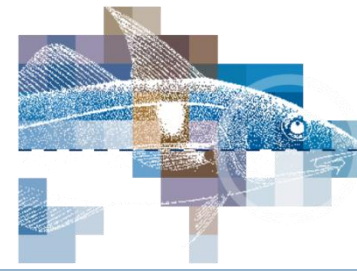
identifier	tank_departure	tank_destination	cycle_from	cycle_to
batch one	C1	B110	Incubation	Larval rearing
batch two	C7	B107	Incubation	Larval rearing

# Data analysis - PEPITo

A DATA MINING TOOL



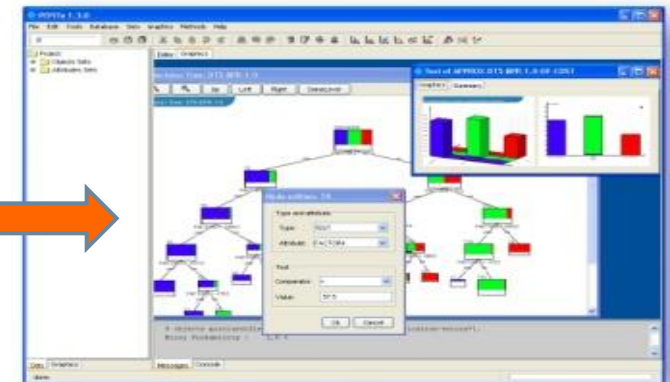
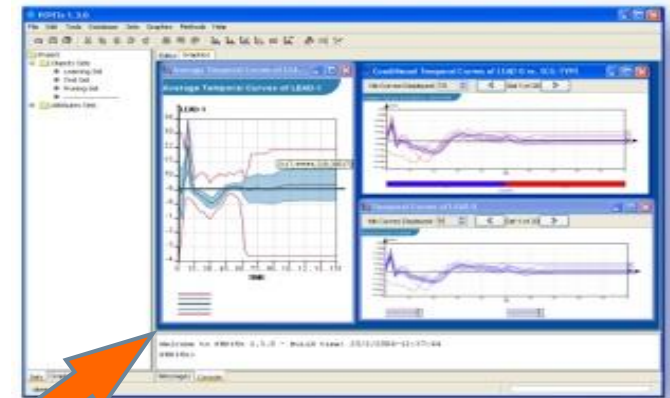
The result of a query is a table, a CSV file that can be easily converted in an Excel sheet and uploaded in the analytical software 'PEPITo' for further analysis.



# 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,...



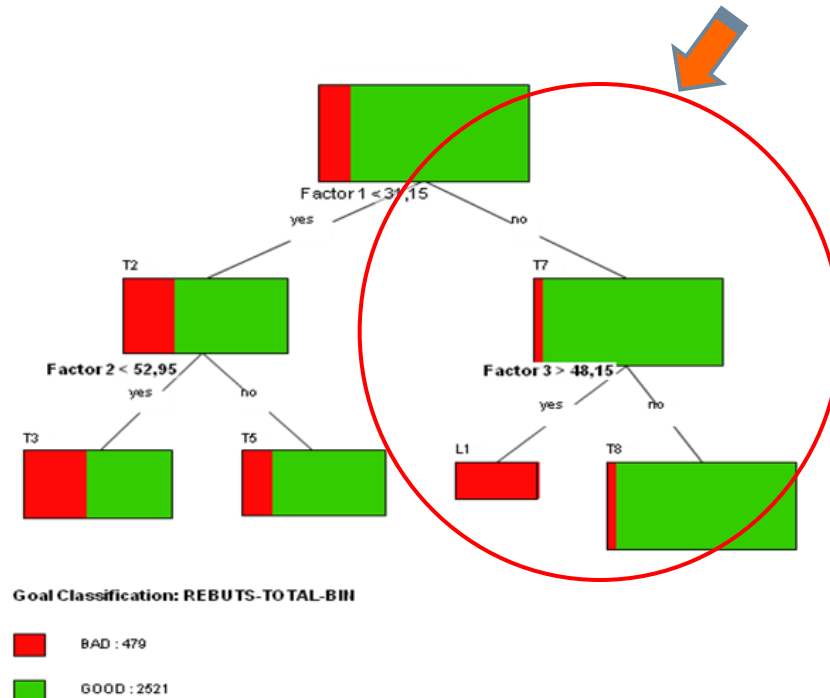
# Data analysis - PEPITo

## PREDICTIVE AND ROOT CAUSE ANALYSIS

Once enough data is stored in the system, we will be able to apply predictive analysis tools (like decisions tree 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, the conclusion would probably be **specific to the farm**)
- detect in the whole set of data within 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 improvement 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 farm 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**



## What next?

- **Project finishes soon – does this part end with the project?**
- **NO – looking to finish testing (within timeframe) and continue to put in data (2 year commitment)**
- **Looking to expand data sources (outside of the Consortium)**
- **Build up competence for data input/monitoring**
- **Consideration of new project(s) including this aspect**
  
- **A fine example of a new cross-cutting & innovative approach**