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

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.

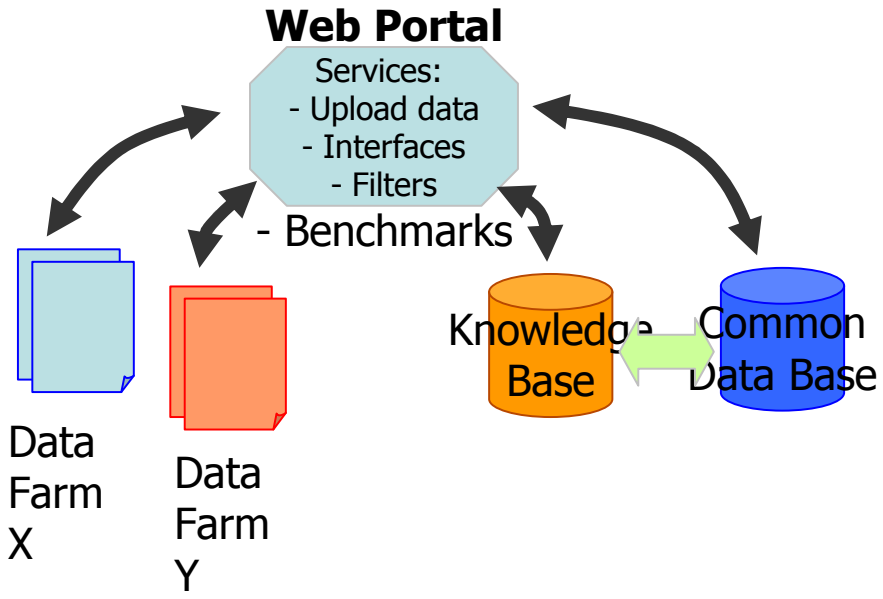
- **LACK of HATCHERY PERFORMANCE DATA**
- How to measure technical and economic performance improvements?
- How to understand the underlying causes?
- How to transform this knowledge in best available practices?

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.**

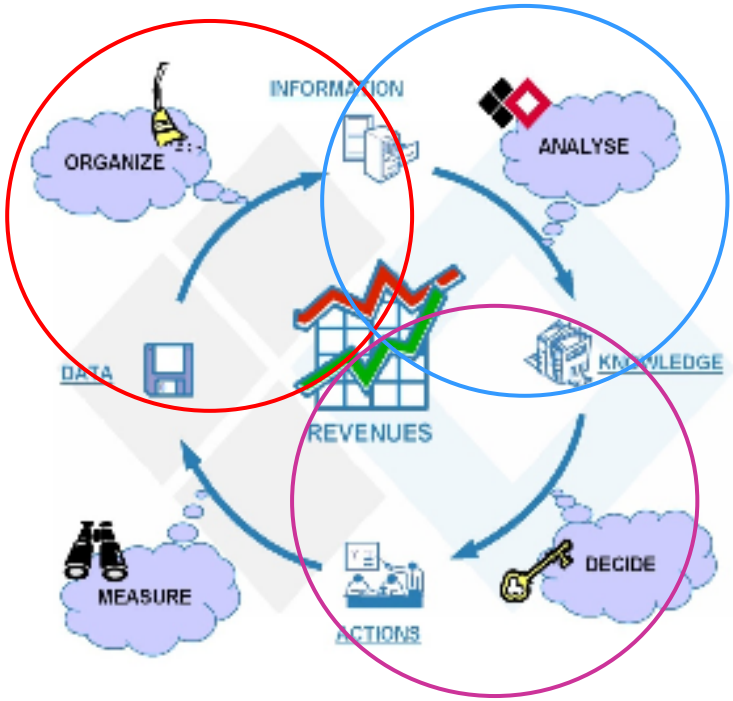
- **The main goal is to identify key factors affecting production performance and the underlying causes of malformations onset.**

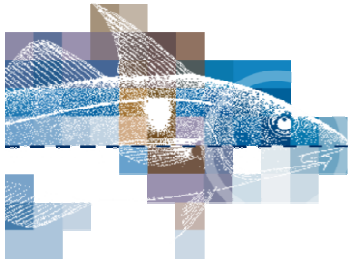
PEPITe is a company specialized in DATA MINING

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

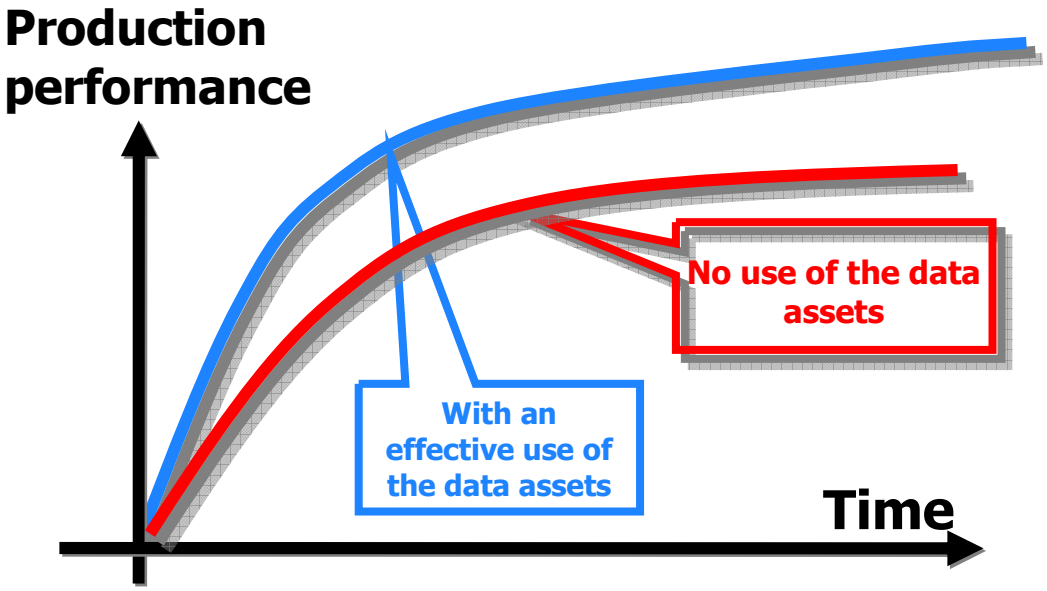
- Visualization, statistics, automatic learning from prediction models (forecasting models)

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:

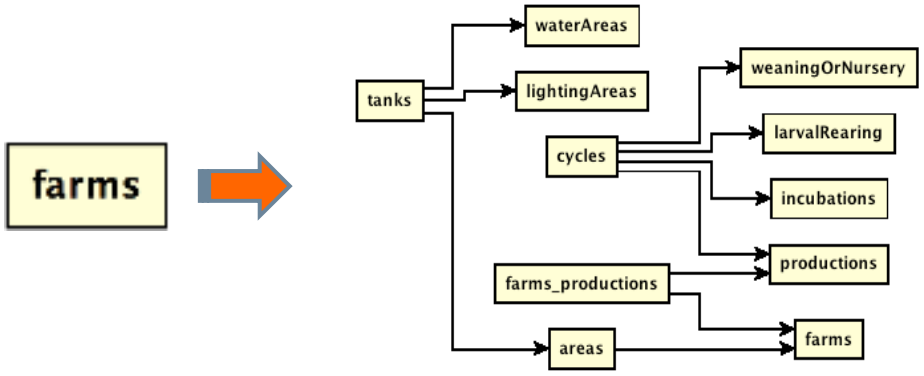
- a **measure of the performance** of specific goals that a business has defined to be of critical importance to their success → **malformation rate x batch**
- a **target** (or targets) → **set of a threshold value > than 10 %**
- an **action** resulting from that measurement → **corrective actions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

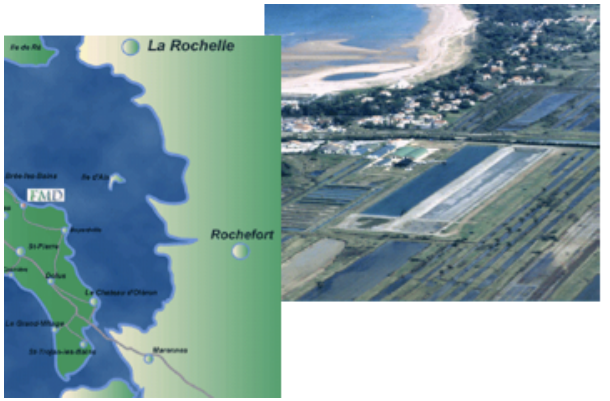
The "making" of the database

- **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

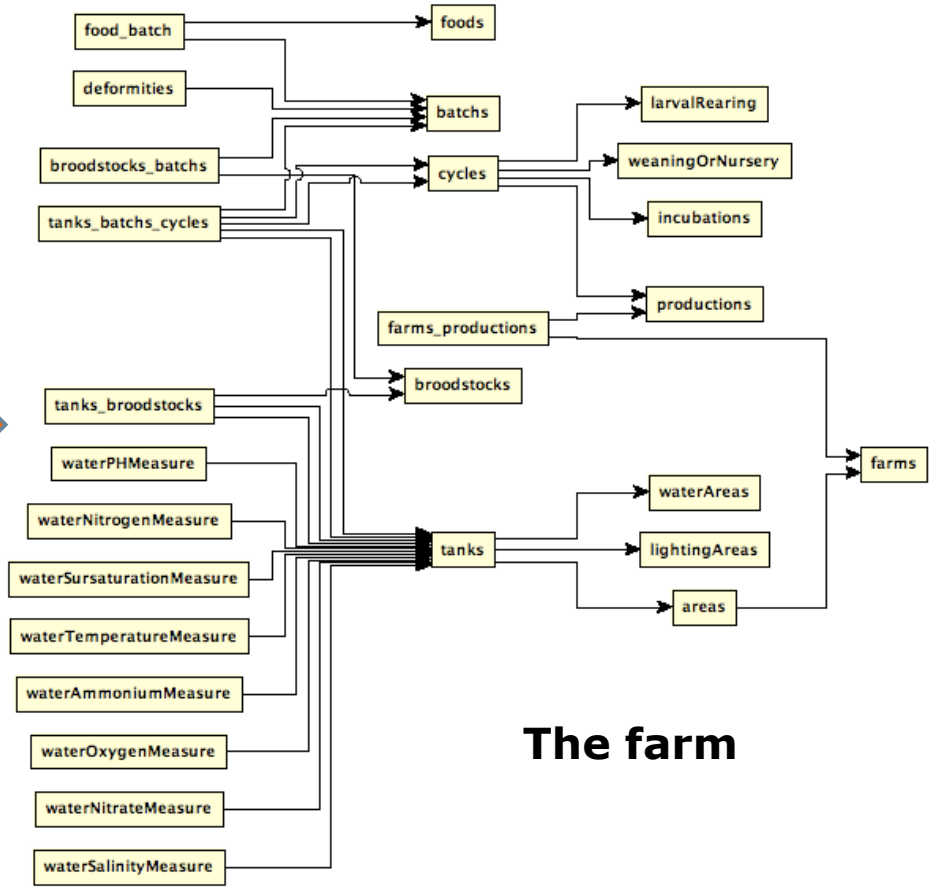
The making of the database

- **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.

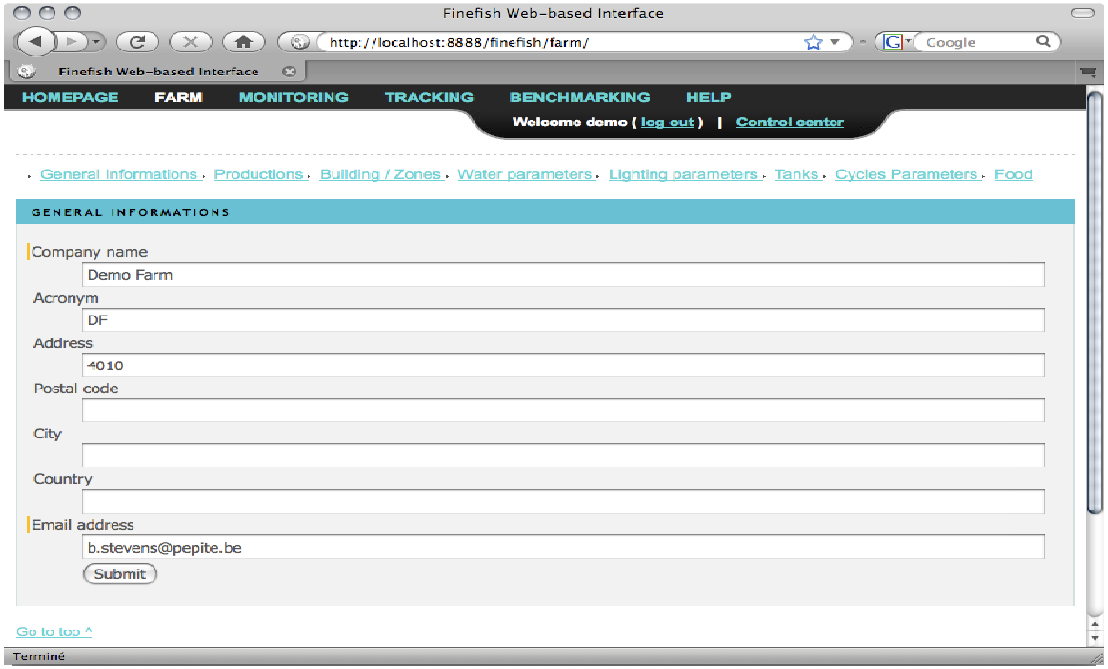
Exploiting the database

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



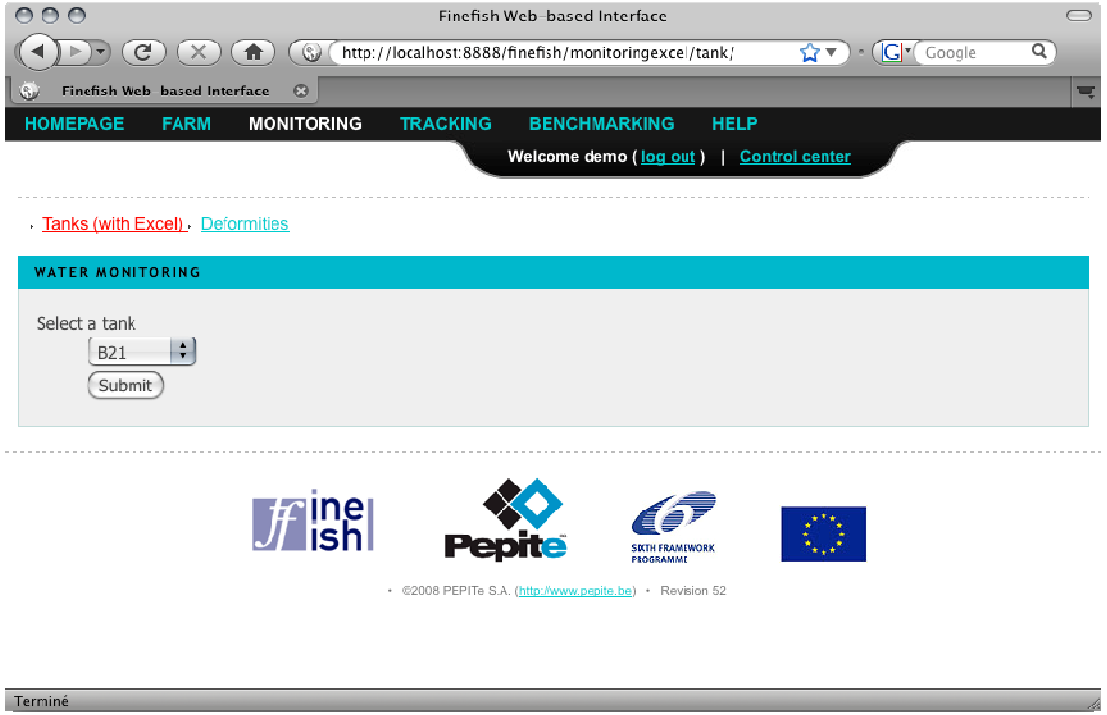
Exploiting the database

Monitoring

Once the configuration of your farm is complete…

You can import data on:

- Monitoring of single tanks (pH, T, $[O_2]$, salinity…)

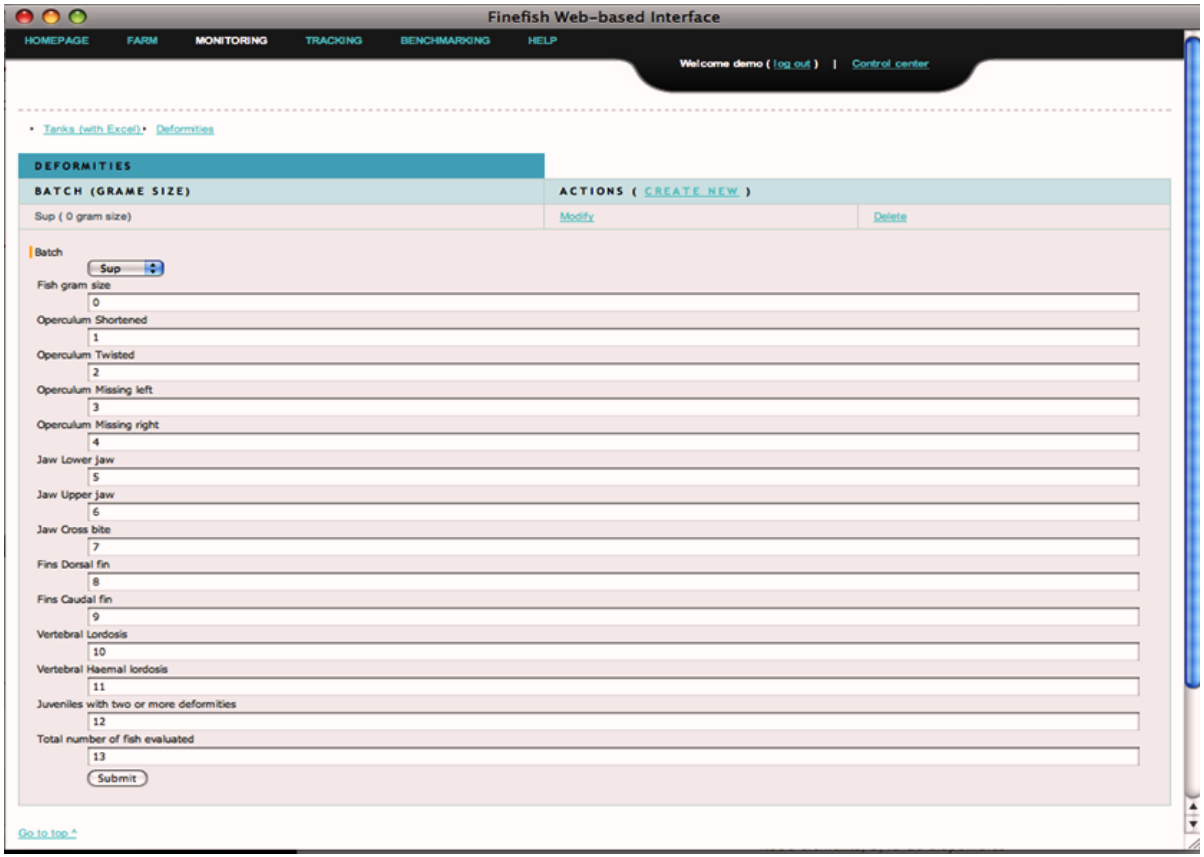


Exploiting the database

Monitoring

You can import data on:

- Malformations x batch



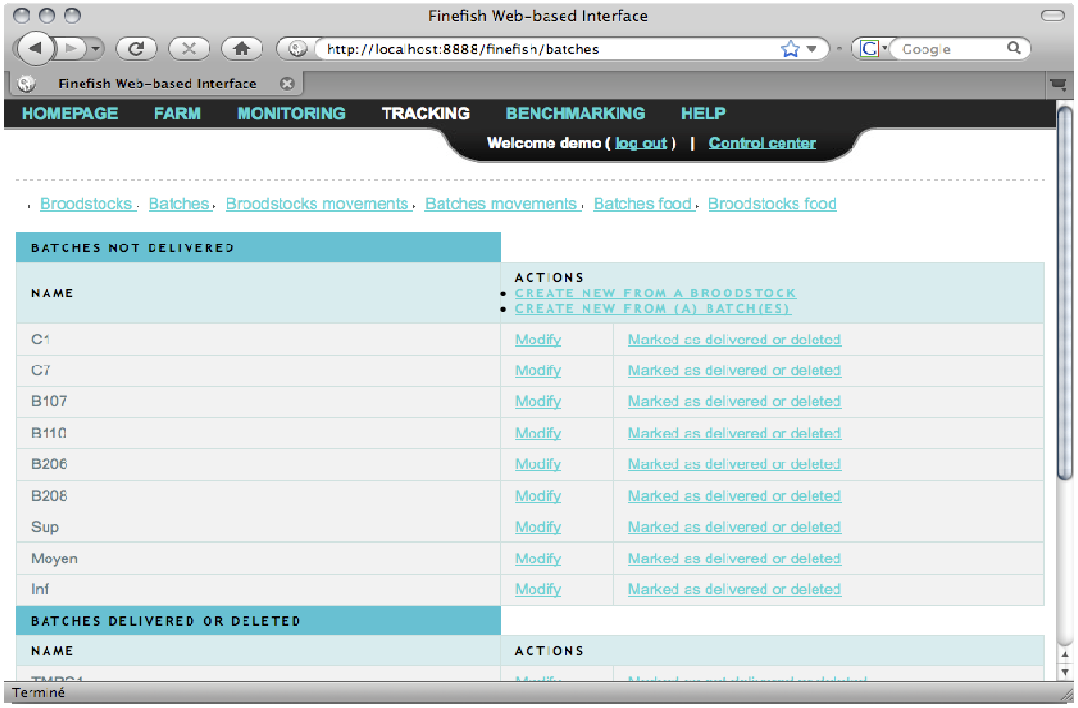
Exploiting the database

Tracking

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

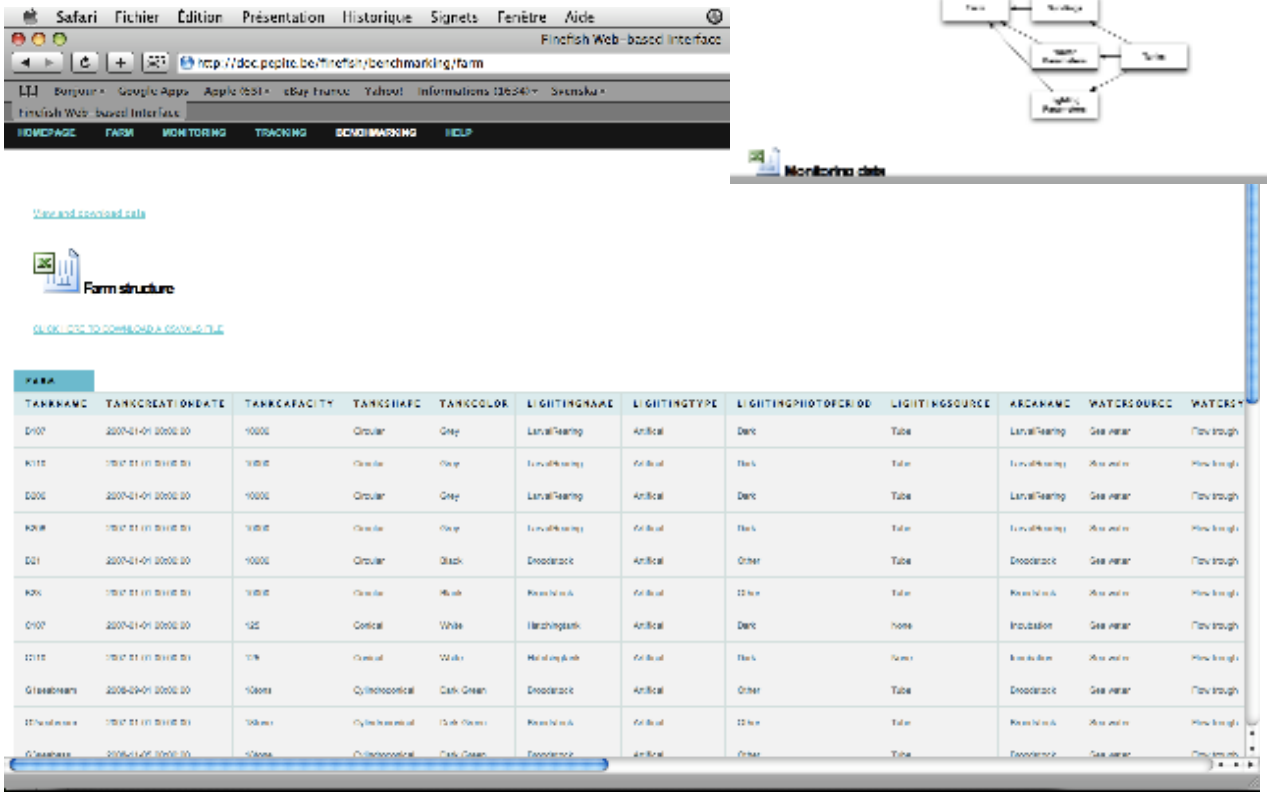
You can import data on:

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



Exploiting the database

Benchmarking



You can benchmark data on:

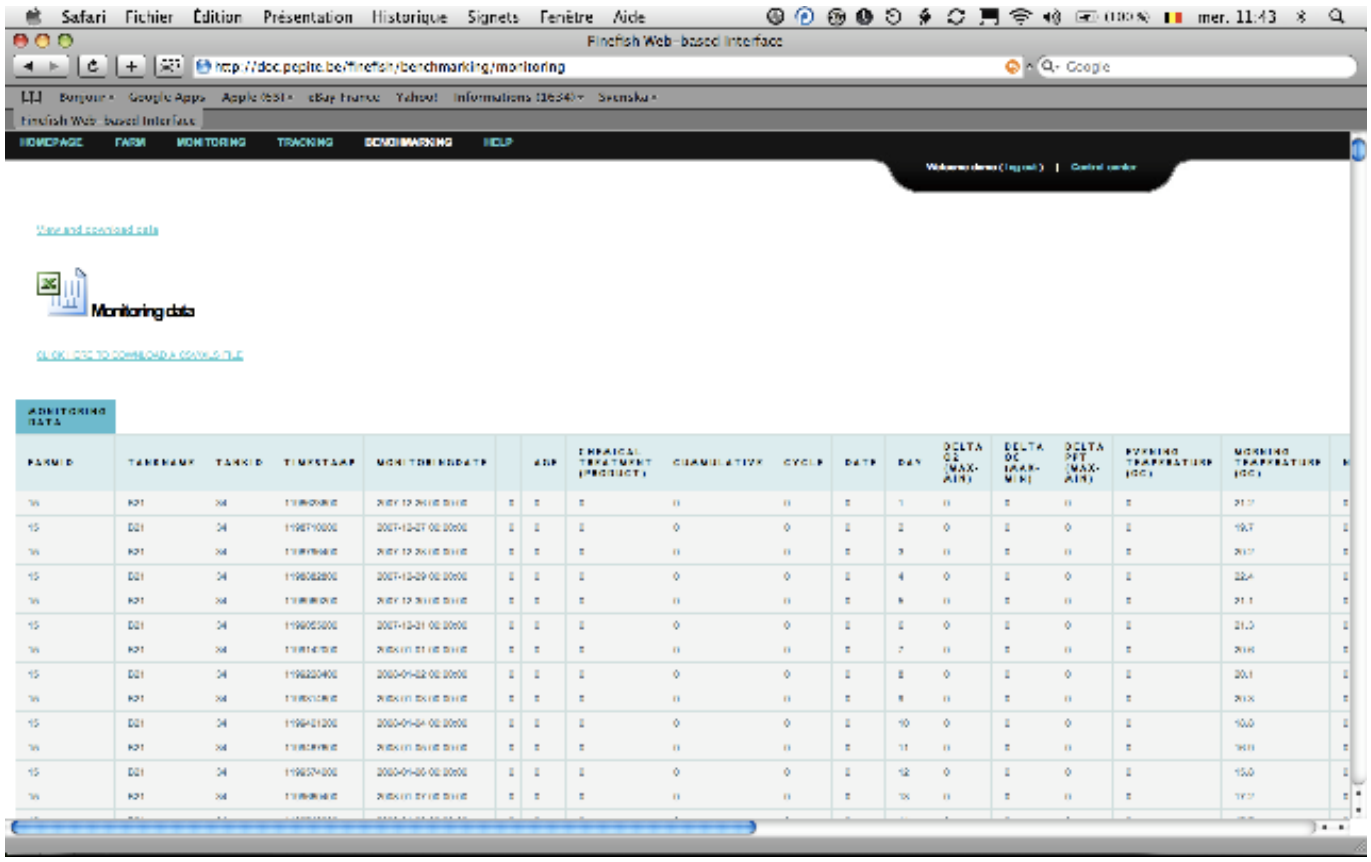
- **Farm structure**

Exploiting the database

Benchmarking

You can benchmark data on:

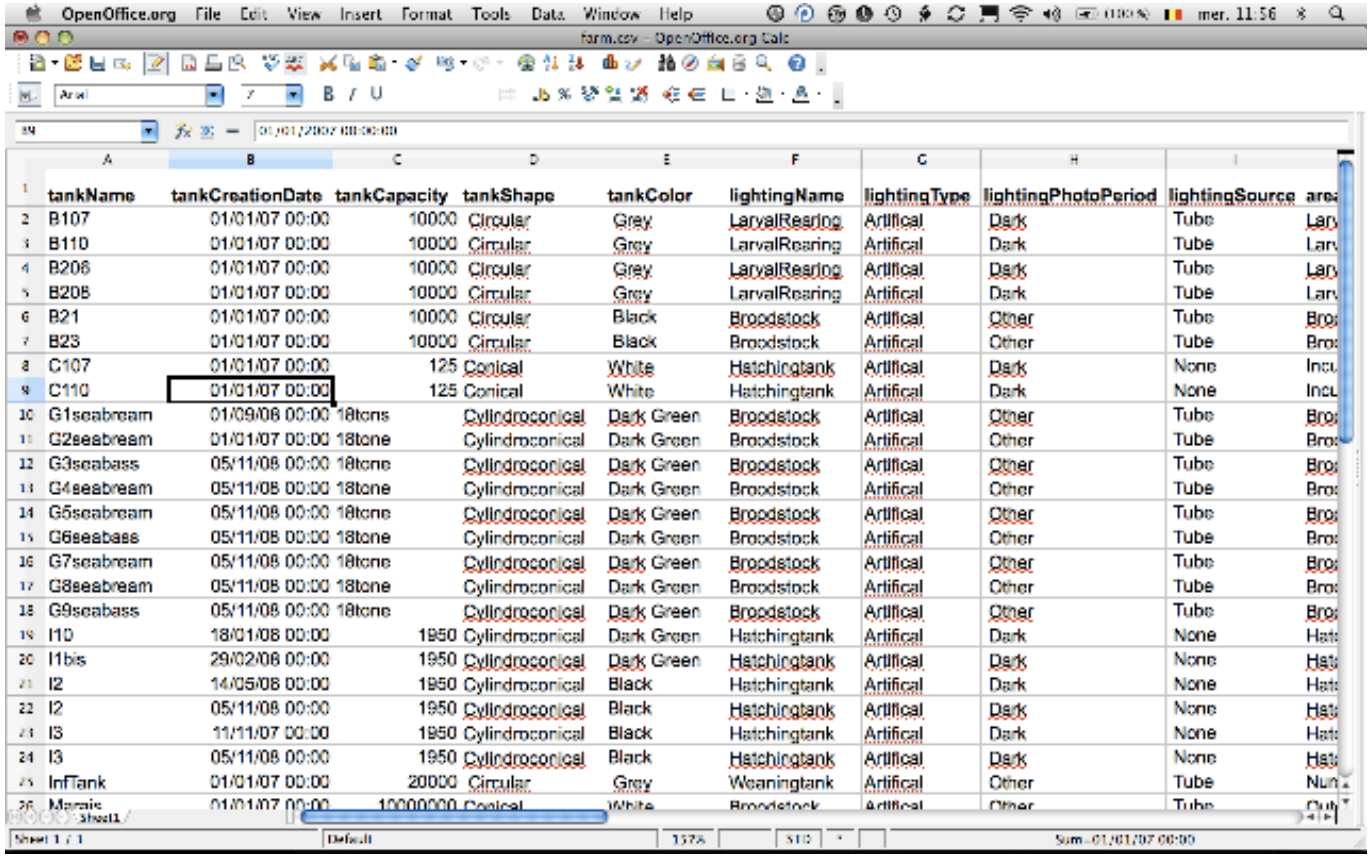
- **Monitoring**



…and actually all available data!

Exploiting the database

Output are tables containing data in an organized structure



	A	B	C	D	E	F	G	H	I	
1	tankName	tankCreationDate	tankCapacity	tankShape	tankColor	lightingName	lightingType	lightingPhotoPeriod	lightingSource	are
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	Bro
7	B23	01/01/07 00:00	10000	Circular	Black	Broodstock	Artifical	Other	Tube	Bro
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	01/09/06 00:00	18tons	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
11	G2seabream	01/01/07 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
12	G3seabass	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
13	G4seabream	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
14	G5seabream	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
15	G6seabass	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
16	G7seabream	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
17	G8seabream	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
18	G9seabass	05/11/06 00:00	18tone	Cylindroconical	Dark Green	Broodstock	Artifical	Other	Tube	Bro
19	I10	18/01/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artifical	Dark	None	Hat
20	I1bis	29/02/08 00:00	1950	Cylindroconical	Dark Green	Hatchingtank	Artifical	Dark	None	Hat
21	I2	14/05/08 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hat
22	I2	05/11/06 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hat
23	I3	11/11/07 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hat
24	I3	05/11/06 00:00	1950	Cylindroconical	Black	Hatchingtank	Artifical	Dark	None	Hat
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	Out

Exploiting the database

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

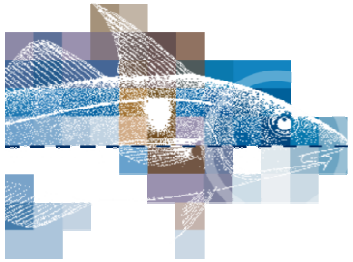
Queries are precise requests for information retrieval within a database and information system.

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

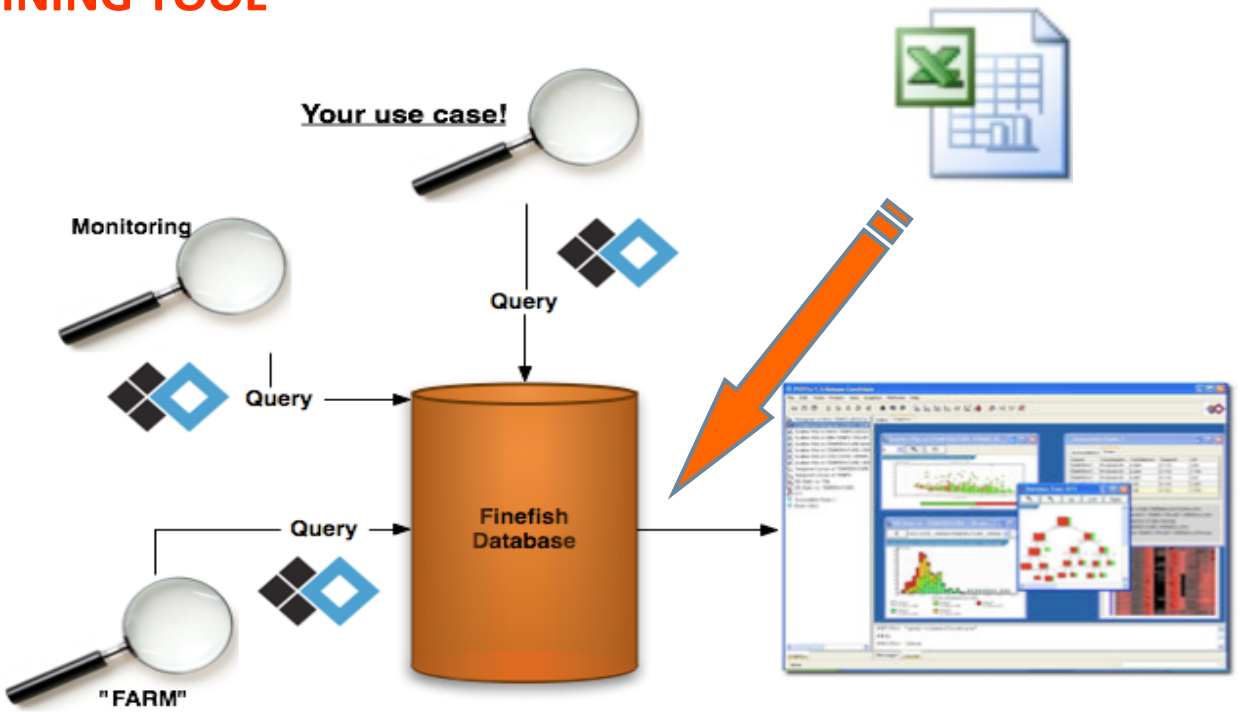
QUERYS

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

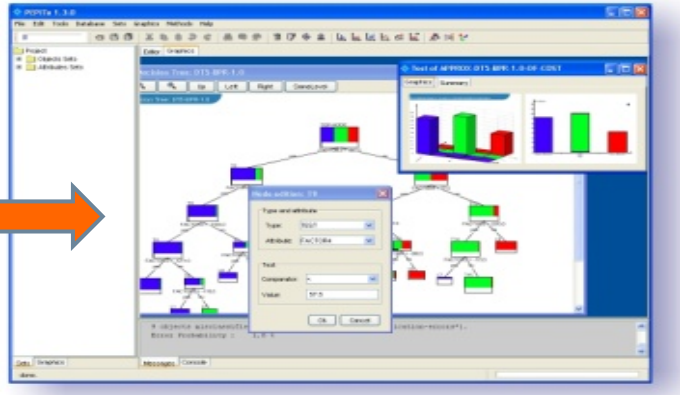


The result of a query is a table, a CSV file that can be easily converted in an Excel sheet and uploaded in the analitical software PEPITo for further analysis.

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

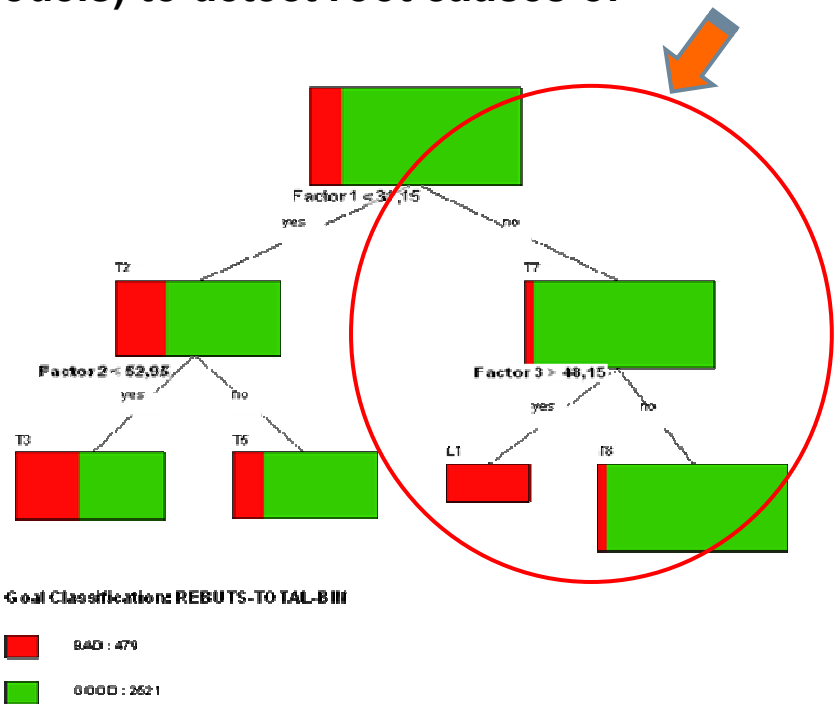
PREDICTIVE AND ROOT CAUSE ANALYSIS

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