Moving from data collection to improving performance

Philippe Mack, Francesca Margiotta & Courtney Hough

Introduction

The core objective of the FineFish project was 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. While preparing the project's workplan at the start, some very simple questions were posed such as 'What is the current level of malformations in hatcheries?', 'How much do malformations cost the sector?' and 'How can you measure improvements?'

During the project start-up period it was quickly realised that the monitoring procedures within the different hatchery activities covered by the project (salmon, trout, cod, seabass, seabream) were all different and not standardized. It was also observed that none of the hatcheries used proprietary software for following production profiles, most using spreadsheet software for data entry and analysis.

Generally, there is a lack of publicly available data and information on hatchery performance in all of the sectors covered, making comparative work extremely subjective. This is for evident professional reasons concerning corporate performance.

In addition, appreciation of malformations has been very subjective, with diagnostics being limited to the nature of the malformation in question – being measured as severe rather than on a scale of influence (e.g. 1-5).

Finally, from an analytical point of view, the complexity of the different influences on malformation incidence – ranging from diets and nutritional components, through the complexity of in-tank abiotic factors, to temperature profiles at different stages, while not forgetting genetic aspects – means that there is no simple single and evident answer to monitoring.

To respond to this position, within Finefish, 3 main questions were posed

- ✧ How to measure technical and economic performance improvements?
- ✧ How to understand the underlying causes of malformation problems?
- ✧ How to transform new knowledge into best practice?

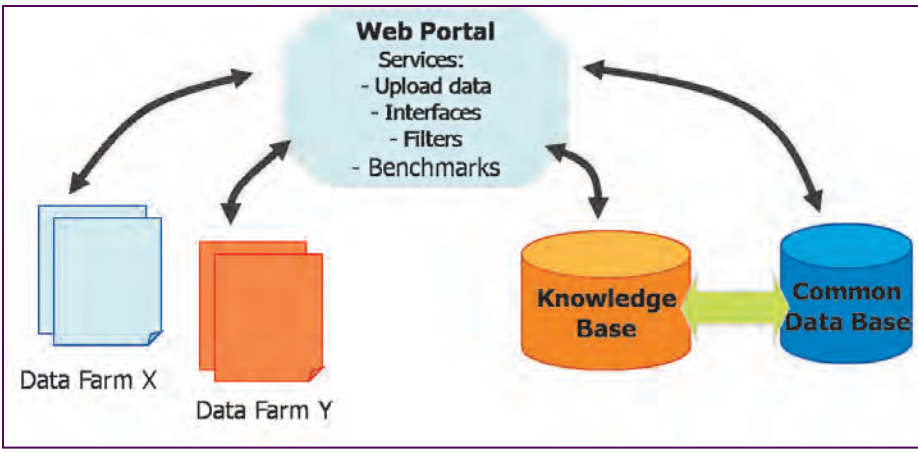
It was thus that FineFish examined how to organise the systematic collection of hatchery data and its analysis with regard to the incidence of malformation in hatcheries and commercial fish farms, the ultimate target being to help aquaculture operators in:

- ✧Benchmarking their activities so as to measure variations
- ✧Being able to share data and knowledge without compromising confidential information;
- ✧Implementing good practices and new knowledge

Approach

The data collection for benchmarking started by using relatively simple worksheets for follow-up and reporting. However, it was rapidly realised that this approach did not meet the expectations of the SMEs and that a rather professional approach was necessary. Software exists on the market that can recover and analyse production parameters but each SME has a different programme that is adapted to its own operations. Consequently, it was decided to develop a **common platform** that could be applied for all hatchery operators.

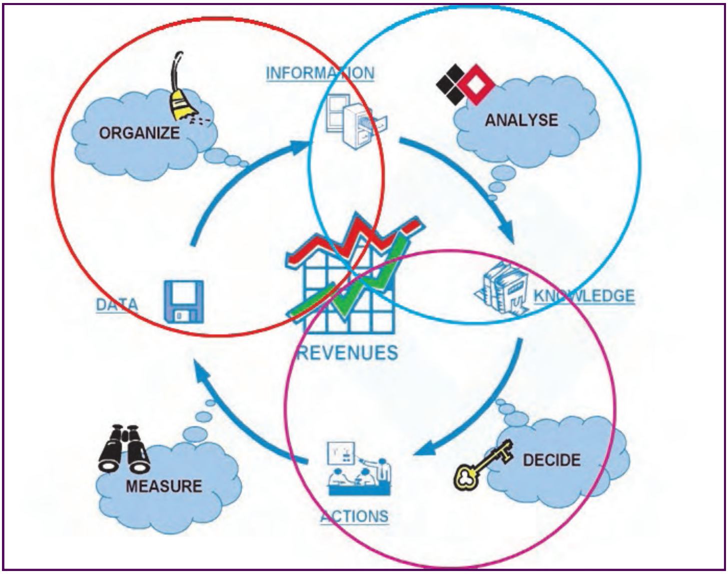
The basic concept is that individual hatcheries would prepare data in a standardised form that is submitted to a database via the internet (a web portal). This data can then be retrieved by the hatchery for benchmarking and monitoring purposes of the hatchery's performance. Obviously, this information is confidential to the hatchery in question and no other operator would be able to access this.



However, the database would thus contain information of high value for analytical purposes if many different hatcheries provide data and which would not breach the conditions of confidentiality. For example, if one wanted to compare the incidence malformations in seabass vs. temperature, or whether lighting regimes have an effect, or whether specific nutritional profiles reduce malformations….

By having the maximum amount of data stored within a single, standardized database one would enable the comparison and benchmarking of data on production methodologies applied in different hatcheries.

The subsequent analysis of the data will then enable the extraction of useful information that can be applied to the improvement of current practices. The main analytical goal is to identify key factors affecting production performance and the underlying causes of the onset of malformations.



When examining the issues of analysis – and recognizing the enormous array of factors that could be causal to malformation incidence – the possibility of applying data mining technology appeared as a highly interesting option for application.

DATA MINING is "the science of **extracting implicit, previously unknown, and useful information from large data sets or databases**". It can also be defined as "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."

The benefits of this approach include the use of a systematic and smart way of looking at data and the application of a methodology that is able to extract - from very large databases - information that is:

- Previously unknown
- Valid
- Understandable
- Useful

There are also a wide range of tools available to the users (i.e. the hatchery) which include visualization of results, statistical analysis of data. A newer tool is the fact that automatic learning can be generated within the system, allowing the application of prediction or forecasting models – which can also give guidance through decision trees within a process.

In summary, the concept proposed is to develop and provide a comprehensive service that will allow better use of the data 'assets' that, in many cases, are already collected at the individual hatchery level, enabling the individual operator to

✧Understand the past

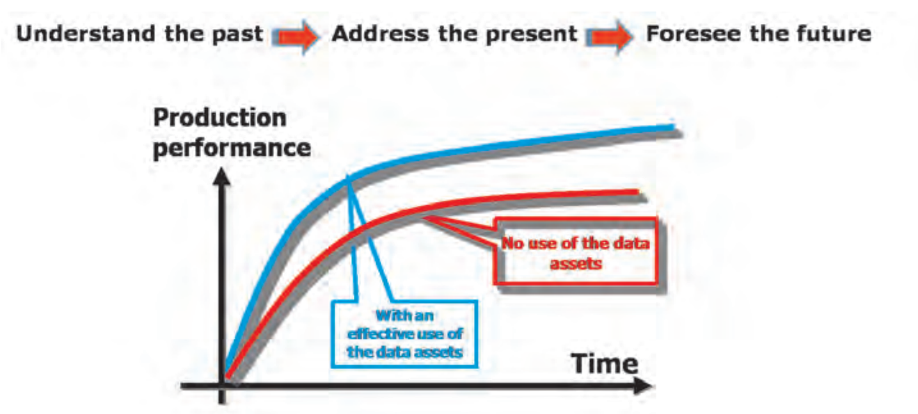
- Explain the behaviour of key performance indicators [KPI] (e.g. malformation rates, growth rates…)
- Transform implicit knowledge into rules of procedure (protocols)
- Identify the past conditions that improved production performance (so as to reproduce these consistently)
- Identify process weaknesses 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



Moving from data collection to improving performance

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

Following consultation with the farm managers and technicians in the hatcheries within FineFish, the main KPI identified was:

MALFORMATION RATE x BATCH

In a continuous improvement context a KPI is a composite of the following:

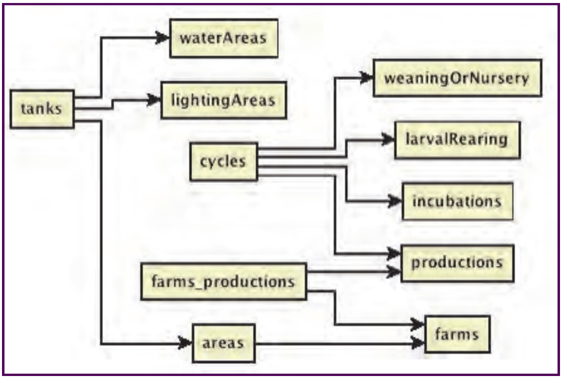
- a **measure of the performance** of specific goals that a business has defined to be of critical importance to their success (e.g. malformation % x batch
- a **target** (or targets) (e.g. the set of a threshold value: > than 10 %)
- an **action** resulting from that measurement and leading to corrective actions if the value is exceeded

An example of a measurable objective for a hatchery would thus be *"Decrease malformation rate per batch from 30% to 20% by 2010".* In this case, 'the malformation rate per batch' is the KPI.

The FineFish Database system — "FindIT"

After a period of observation and interviews in a test hatchery (Ferme Marine de Douhet) and detailed analysis of different hatchery structures and procedures, including interactions of the various areas (water area, tanks, lighting area…) within the hatchery, the logical relationships between these components were used to **design a data model to be implemented in a RDBMS.**

RDBMS is a **relational 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. An example of table structure is given below.



This procedure allowed the modelling of the hatchery production process, an audit of the data collection procedures on site and led to the design of a data model. This design was implemented within a data warehouse (Data warehouse is a repository of an organisation's electronically-stored data. Data warehouses are designed to facilitate reporting and analysis)

While the first model was based on the test farm “FMD” it is adaptable to all farms, because it is a model of the real functional world and its relationships. Knowledge based logic is inherent to the system.

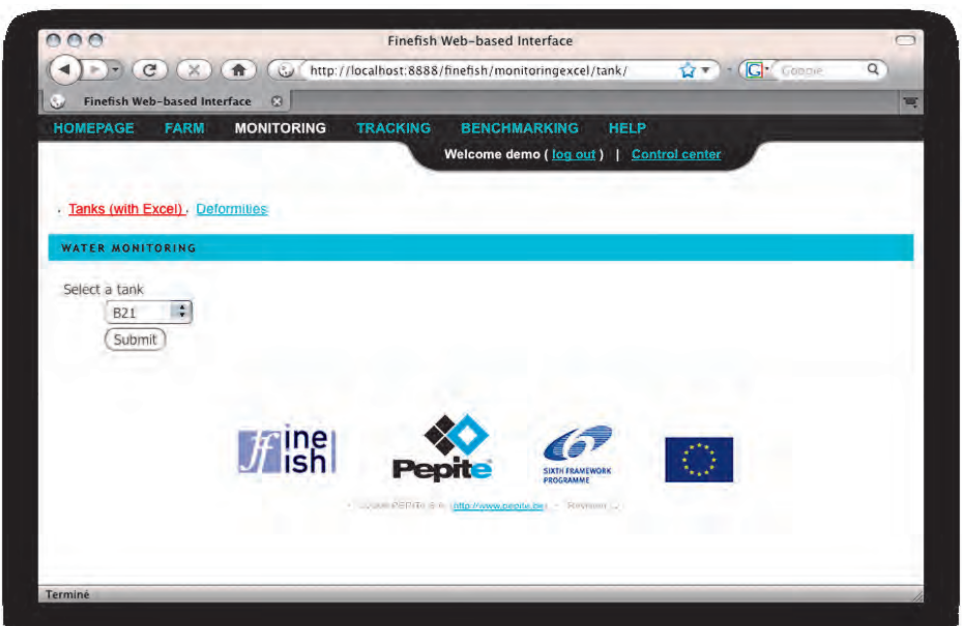
Using ‘FindIT’

The first stage is configuration of the ‘hatchery’, which means providing generic information on the operations, physical structure and materials of the installations, which is achieved through the web-portal (see example below).

- ✧Generic farm information
- ✧ Farm production profile (i.e. species)
- ✧ Farm ‘area’ conditions
e.g. Water parameters, lighting
- ✧ Tanks, ponds (materials, volume, depth...)
- ✧ Production cycles
- ✧ Feeds....(Feed identification, linked to ingredients)

Once this stage is completed, monitoring data can be included, such as the parameters of single units (tanks, ponds)... pH, Temperature, O2, salinity, water flow rates...

This is then supplemented by data on the types of malformation and the number of malformations/batch.



As all livestock managers know, stocks are moved regularly and it is essential to follow their origin. Allowance is therefore made for identification of broodstock and their offspring batches, setting their position (in tanks) and their position in cycle parameters (i.e. nursery, larval rearing...). Data is included on broodstock and batch movements, broodstock food/diet and batch food/diet.

Benchmarking

With the information entered into the database, it is possible to start benchmarking activities, which can be simple or more complicated.

As examples, one can benchmark the farm's structure (i.e. tank volumes, colour, depth) or any of the relevant monitoring data entered (e.g. temperature).

Output

The main outputs are tables that are made in an organised structure. These can range from, for example, a simple report of the daily temperatures for the month or the growth rates per batch using different diets within a defined production cycle. However, all of these outputs depend on a data request or query.

Queries are precise requests for information retrieval within a database and information system and their structure depends on what information is required. Their structure is quite complicated and requires specialist assistance if the hatchery does not employ an IT expert.

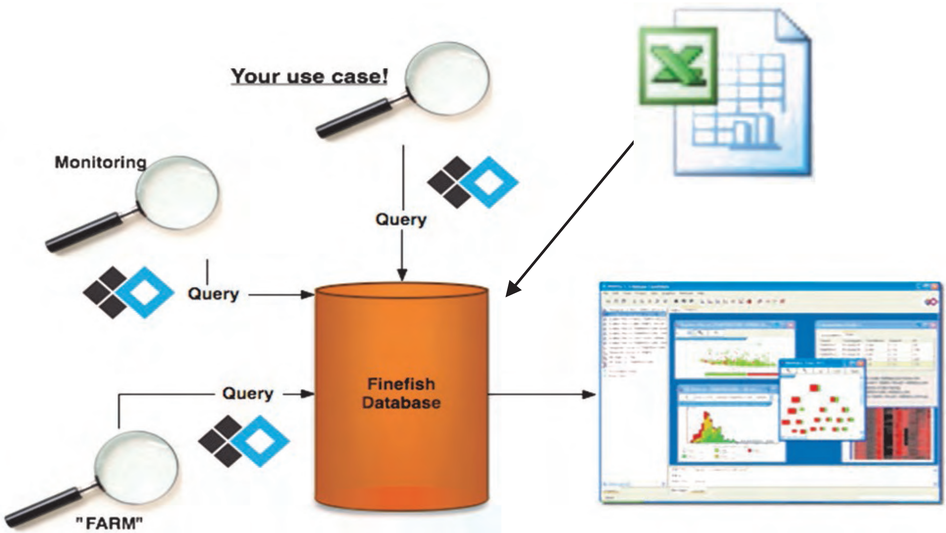
Examples for regular monitoring could be the growth and malformation incidence of a specific batch, requiring access to all stock movements and batch tracking, in addition to malformation assessment and measurement — as indicated below.

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

As a result, the data gathered through such monitoring provides new information tables as outputs.

This new data can then be used for the application of data mining - using a dedicated IT tool available at the FineFish web portal.

The diagram below summarises the operations. Data (as a worksheet file) is uploaded into the Database. Queries also can be made—which generate new tables. These new tables can then be analysed within PEPITo—the data mining tool



This tool is the proprietary software 'PEPITo' which accomplishes the following advanced tasks:

✧Data validation and filtering

✧Data transformation: FFT (Fast Fourier transform, sampling…)

✧Data visualisation, using

- distribution plots, scatter plots, temporal curves,…

✧Statistical analysis:

- analysis of variance, correlation analysis,…

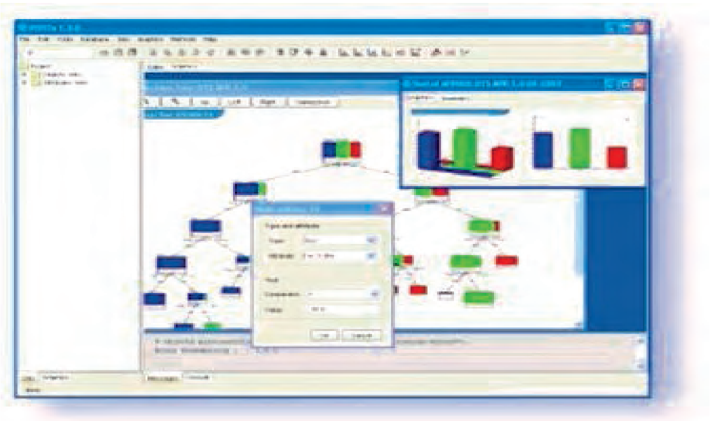
✧Predictive analysis:

- neural networks, decision trees, association rules,…

The core interest in this application is access to a comprehensive data treatment system that provides both predictive and root cause analysis.

Examples of Data and Information Graphic plots within PEPITo

Examples of Decision Tree creation

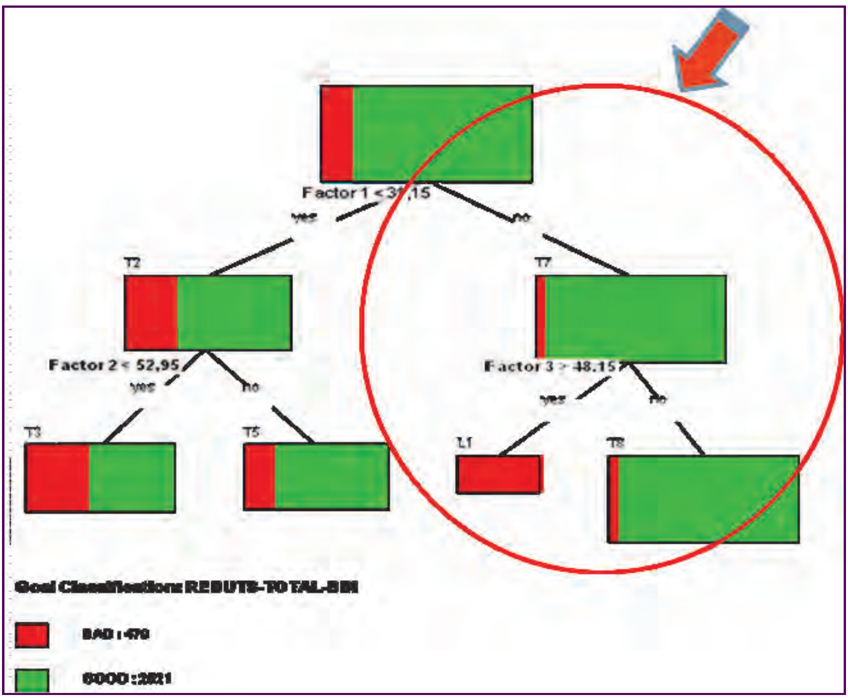


Once enough data is stored in the system (a combination of the database and the newly created data information tables), it is then able to apply predictive analysis tools (such as decision tree models) to detect the root causes of malformation rate. Evidently, without adequate data entry no analysis can be achieved.

Possible analysis could be:

- ✧to detect in an individual farm the parameters that explain a malformation rate drift between two production cycles (in this case, the conclusion would probably be **specific to the hatchery**)
- ✧ to detect in the complete set of data within the database why the malformation rate is higher in a range of farms' production (e.g. in the Mediterranean on seabream); in this case, one can expect that the conclusion(s) would be broader, and that improvement actions could be applied to **every hatchery** within a range of conditions.

Consequently, the benefits of providing data to such a system are applicable to both the data provider and the sector as a whole.



This example of a decision tree shows the influence of 3 factors within a process, where decisions have to be taken so as to obtain the highest ratio of "good" (green) vs. "bad" (red). The absence of factor 3 is indicative of the best result (vs. factor 2). This type of predictive analysis should be able to help production and process planning

Current situation

5 hatcheries have configured their structural inputs and are registering their production cycle data and 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 needs and requirements
- Identify possible bugs in the database and correct these in order to enable and facilitate the correct and easy input of data
- Modify the database and data organisation so as to enable the creation of a range of specific queries to be analysed through the data mining tool.

Identified needs

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

"we want to be able to keep track of different light regimes in tanks over time"

"we want to keep track in time of the different treatments reserved to fish in hatchery tanks. "

All of these are simple to achieve with the current version of FindIT.

✧ The database maintains records of all fish movements within a hatchery/farm as well as the different conditions of each tank over time

✧ For each batch of fish produced within a hatchery, one can extract historical data, over time, for

- Location/position within the farm (tank/pond id)
- Treatments (e.g. chemical treatment, antibiotics…)
- Monitored parameters (T°, pH, light…)
- Feeding regimes(rotifers, algae, different feeds..)

It is therefore possible to integrate all of the variable factors that are believed to have an influence on malformations as well as FCR and growth performance, linking all of the fixed operations of the hatchery to operational variables.

Potential and future of 'FindIT'

The function and expectations of using this tool are the following:

- To discover unexpected correlations between parameters
- To benchmark production patterns within different hatcheries and identify the best practices (BMP)
- To verify *ad hoc* knowledge with historical data so as to detect abnormal situations early within a production cycle
- To expand the system to other KPIs (identify new KPI on the basis of analysis)
- To include genetic information in the system (relating to broodstock)
- To automate the creation of reports
- To automate the import of data into the system which is collected using other proprietary production monitoring software

It has not been possible to complete the full development of 'FindIT' within the FineFish project, partially because it was an unexpected project component but also because it requires more work on rendering the application to be user-friendly and applicable to a wide variety of situations. At the present, data is being entered by Finefish partners and it is hope to extend this to RTD institutes and companies operating within the Larvanet COST action (see http://www.larvanet.org/)

It is evident that the complexity of preparing automated queries that provide the anticipated answers needs thorough investigation and preparation. Easy data uploading and analytical procedures are the first priorities in the development of this aspect.

The FEAP, Pepite and other FineFish project partners are examining the means of achieving this so as to progress 'FindIT' to being able to answer the expectations developed within FineFish .

It is clear that this experience is a fine example of a new cross-cutting and innovative approach to monitoring and forecasting.

Companies or research institutions who wish to know more about the progress of 'FindIT' and/or who would be prepared to provide data should contact the FEAP Secretariat at secretariat@feap.info.

FindIT's components and operations are structured to reflect real-