From data to performance:

What can we learn from production data?

SIXTH FRAMEWORK PROGRAMME

The FineFish development of a tool for data mining and benchmarking

Francesca Margiotta FEAP & Philippe Mack PEPITe

Paris
June 2009

Some background information…

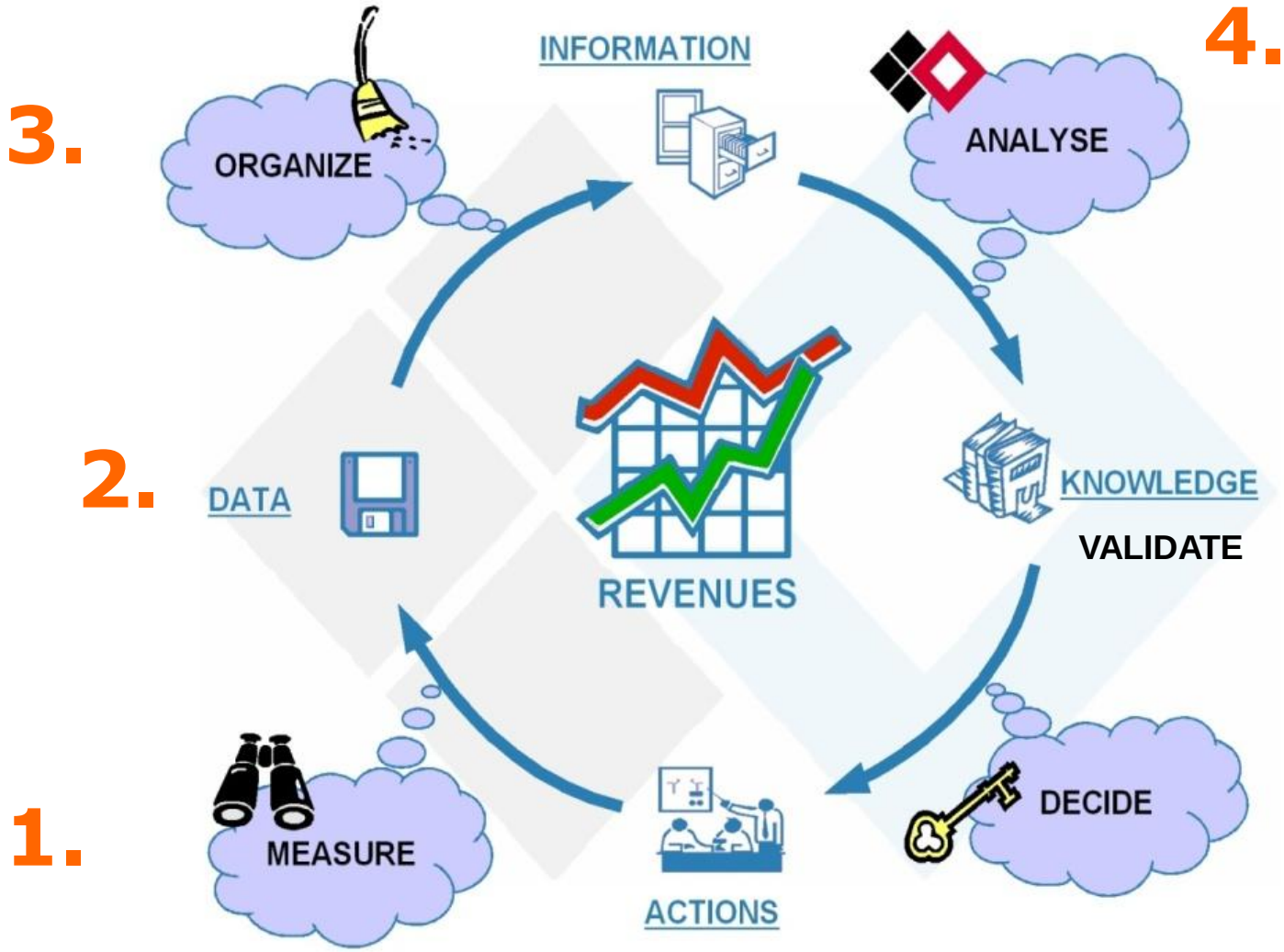
New practical knowledge on how to reduce the incidence of malformations in the major species used in European Aquaculture

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**?

GOALS:

- ◇ systematic collection of hatchery data
- ◇ analysis of these data with regard to incidence of malformation
- ◇ identify key factors affecting production performance and the underlying causes of malformations onset
- ◇extraction of useful information and the improvement of current practices


1.
2.
3.
4.
5.
INFORMATION
ANALYSE
ORGANIZE
DATA
KNOWLEDGE
VALIDATE
REVENUES
MEASURE
DECIDE
ACTIONS

1. Measure:

◇ **Automatic measurements**
(e.g. sensor measuring continuously or periodically)

◇ **Manual measurements**
(periodical or case based)

On site

Off site

Measured parameters:

Data!

POSITION (tank id)
TREATMENTS (chemical treatment, antibiotics…)
MONITORING PARAMETERS (T, pH, salinity, light…)
FOOD (rotifers, algae, different feeds..)
MALFORMATION RATE…

2. Data:

◇ **The measured parameters are stored in form of data in excel sheets that can be later uploaded in the database.**

This files will enable to:

◇ **Keep track of batches**
◇ **Keep track of variations in parameters**

Broodstock tank number: G3

The original excel sheet!

Level of domestication :	wild caught as juveniles	wild caught as adults	domesticated	selected	unknown
		*	*	*	

Generation number :	0	1	2	3	4	5
		*				
Year of spawning						

Natural spawning season for the species in this region: start: november end: february

Date of first spawning (dd.mm.yy): 25/02/2008

Date of end spawning (dd.mm.yy): 04/05/2008

Method for regulating spawning :	Thermal	Photoperiod	Both	None
	*	*		

Water system:	Flow trough	Resirculation
	*	

Daily water exchange (%, volume new water of total resirc volume)

Tank renewal rate (%)

Water source	Sea water	Tidal water	Bore hole
	*		*

Hormone treatment: yes no *

if yes brand:

dosage:

Diet	dry feed	premix	fresh fish	frozen fish	squid	other (specify)
	*				*frozen	
Species						

	Temperature (ºC)	Temperature (ºC)	Temperature (ºC)	delta ºC (max-min)	Salinity (ppt)	Oxygen (mg/L)	pH	NO_2^- (mg/L)	NH_4^+ (mg/L)	NO_3 (mg/L)	Chemical treatment (product)
Maximum	15.0	0.0	15.0	0.0	37	15	8.05	0.025	0.2	0.2	
Minimum	15.0	0.0	15.0	0.0	37	8.5	7.46	0	0	0	
Mean	15.0	0.0	15.0	0.0	37	0	0	0	0	0	

		Morning	Evening	Average	Per day						
Date (dd.mm.yy)	Day	Temperature (ºC)	Temperature (ºC)	Temperature (ºC)	delta ºC (max-min)	Salinity (ppt)	Oxygen (mg/L)	pH	NO_2^- (mg/L)	NH_4^+ (mg/L)	NO_3 (mg/L)	Chemical treatment (product)
01/02/2007	1	15		15.0	0.0	37						
02/02/2007	2	15		15.0	0.0	37						
03/02/2007	3	15		15.0	0.0	37						
04/02/2007	4	15		15.0	0.0	37						
05/02/2007	5	15		15.0	0.0	37						
06/02/2007	6	15		15.0	0.0	37						
07/02/2007	7	15		15.0	0.0	37						
08/02/2007	8	15		15.0	0.0	37						
09/02/2007	9	15		15.0	0.0	37						
10/02/2007	10	15		15.0	0.0	37						
11/02/2007	11	15		15.0	0.0	37						
12/02/2007	12	15		15.0	0.0	37						
13/02/2007	13	15		15.0	0.0	37						
14/02/2007	14	15		15.0	0.0	37						

New Excel sheet to upload data easily!

Date	Cycle	From Building	From Tank	To Tank	To Building	Transfer Day
3/11/2007	C1-1		B11	C110		-2
3/11/2007	C1-1		B13	C110		-2
4/11/2007	C1-1		B11	C109		-2
4/11/2007	C1-1		B13	C109		-2
5/11/2007	C1-1		B11	C108		-2
5/11/2007	C1-1		B13	C108		-2
6/11/2007	C1-1		B11	C107		-2
6/11/2007	C1-1		B13	C107		-2
7/11/2007	C1-1		B11	C106		-2
7/11/2007	C1-1		B13	C106		-2
10/11/2007	C1-1		C106	B106		1
9/11/2007	C1-1		C107	B107		1
8/11/2007	C1-1		C108	B108		1
7/11/2007	C1-1		C109	B109		1
6/11/2007	C1-1		C110	B110		1
5/11/2007	C1-1	S2	B110	B201	S2	32
2/12/2007	C1-1	S2	B106	B204	S2	33

C110

Age	T in the morning	T in the evening	Salinity morning	Salinity evening	Oxygen morning	Oxygen evening	pH morning	pH evening	NO2- (mg.L)	NO3- (mg.l)	NH3+(mg.l)
-2	18.2	18.3	36	36	7.8	8	7.69	7.52			
-1	18.7	20.7	35.8	36	7.5	7.8	7.81	7.89	0	0	0
0	20.6	20.8	36.1	36	7.9	7.7	7.93		0	0	0.036
1	20.4		36.1		8.1		7.96				0.56

3. Organize

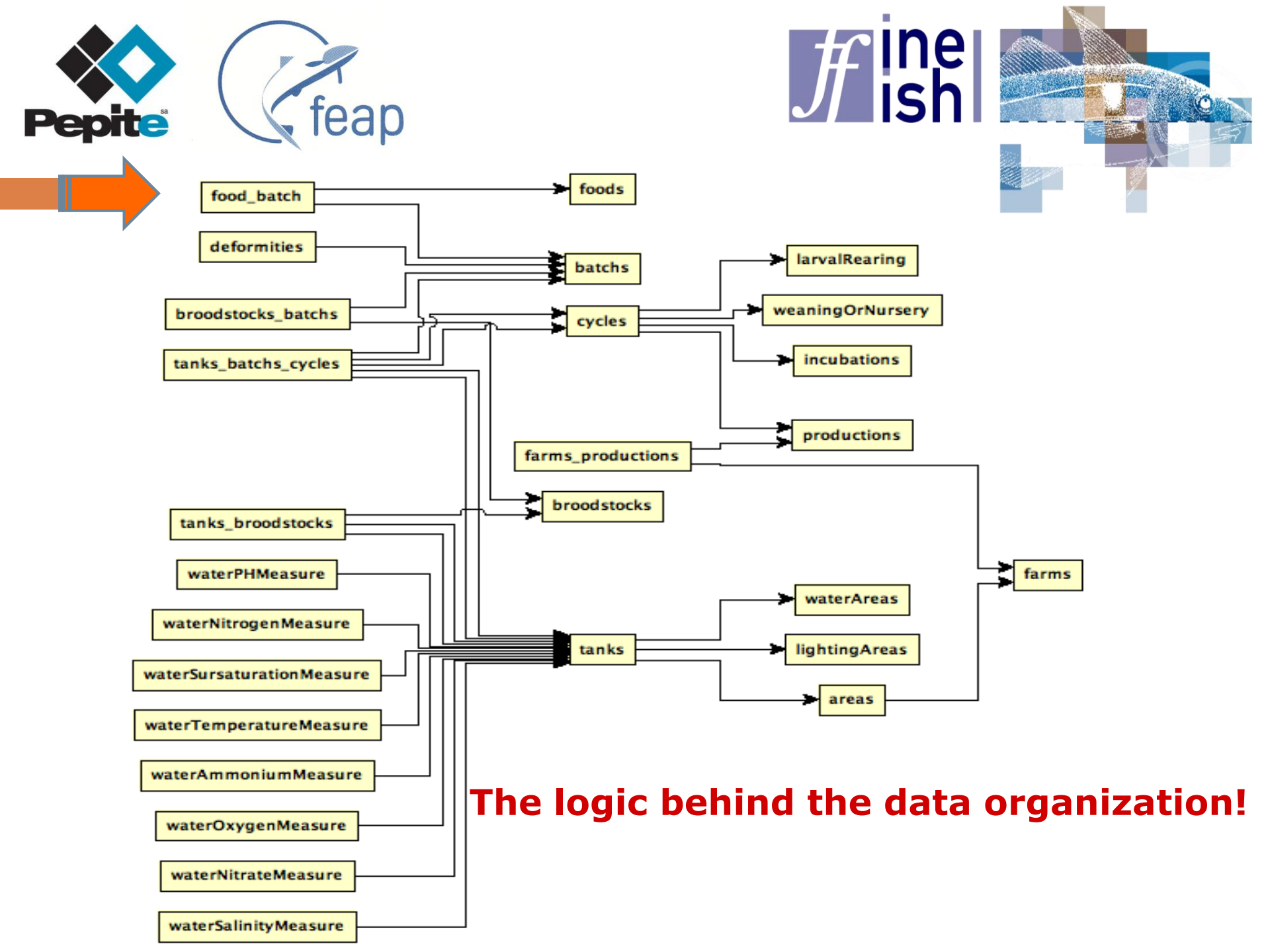
◇ **It is important to organize data in order to be able to quickly recover the useful information.**

Data is stored in form of tables and the relationship among the data is also stored in form of tables.

The organization reflects the structure and the relationships that can be found in the real world.

The farm structure

The production process


food_batch
foods
deformities
batchs
broodstocks_batchs
cycles
larvalRearing
weaningOrNursery
tanks_batchs_cycles
incubations
productions
farms_productions
broodstocks
tanks_broodstocks
farms
waterPHMeasure
waterAreas
waterNitrogenMeasure
tanks
lightingAreas
waterSursaturationMeasure
areas
waterTemperatureMeasure
waterAmmoniumMeasure
waterOxygenMeasure
waterNitrateMeasure
waterSalinityMeasure
The logic behind the data organization!

4. Analyse:

◇ Reporting and descriptive analysis –> monitoring of Key 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**

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 → **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**

4. Analyse:

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."

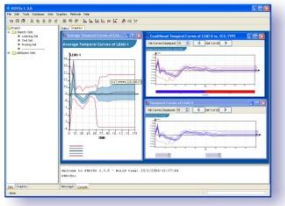
Methodology that aims to extract information from large databases, that is:

- **Previously unknown**
- **Valid**
- **Comprehensible**
- **Useful**

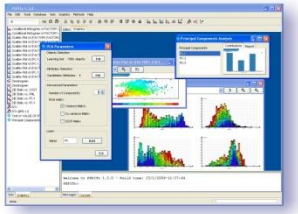
Wide range of tools

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

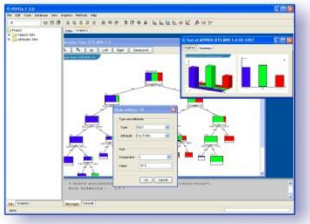
Data Mining
The Technologies



- **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,…

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

5. Validate:

The new knowledge, meaning information extracted from the data, needs to be validated by experts!

Results ⟷ **Expertise**

VALIDATION

↓

Decisions

↓

Actions

↓

Improvement of performance!

Pepite Database
Benefits of a common platform

Having all data available stored in a single, standardized database will enable:

The sharing of data and knowledge;

The comparison and benchmarking of data on production methodologies;

The extraction of useful information and the improvement of current practices.