Title :
Fisheries data quality management : toward quality indicators

Authors :
Christophe Tufféry and Stéphanie Mahévas

Abstract :
The quality of model outputs and advices are tributary of the data quality. Providing data quality requires first to analyze methods of data collection, data storage and data extraction, second to propose methods of data validation and finally to develop metrics of data quality. Such an approach of data quality is based on principles of quality management system like those defined by ISO9001:2000 international standard.

The MEQUAPRO project developed within the framework of program SIDEPECHE at the Ifremer, aims at reaching this requirement. Data qualification consists in assigning a quality value to a data according to a range of preset quality values. Qualification is based on a validation by one or more methods (primarily the respect of protocols), which enables to check if a data reaches the preset quality value. The following stage consists in passing from qualification to quality indicators. These indicators make possible to check and follow in time if data and products (estimates from data) achieved a precise qualitative aim, to meet either internal needs of the Ifremer or those for "external customers" (French Ministry for Agriculture and Fishing, European Union, etc). The French process of data collection in board constrained by the European Data Collection Regulation illustrates this approach.

Keywords :

Contact author :
Stéphanie Mahevas : IFREMER /EMH (Ecologie et Modèles pour l'Halieutique) IFREMER - Centre de Nantes rue de l'île d'Yeu BP 21105 FR-44311 Nantes Cedex 03 [Tel:+33(0)2 40 37 41 81 Fax:+33(0)2 40 37 40 75, e-mail: smahevas@ifremer.fr].

Paper :

3. Introduction :
The quality of model outputs and advices are tributary of the data quality. Providing data quality requires first to analyze methods of data collection, data storage and data extraction, second to propose methods of data validation and finally to develop metrics of data quality. These aspects are common to all national and international statistical institutions and programs. Such an approach could be reused at an international scale, especially for the European Institutions concerned by providing statistics for definition and survey of application of the European Fishery Policy.

Among previous works about data quality control procedures or objectives process, some of them already designed step by step procedures to determine the type, quantity and quality of data needed for decisions support. For example, the EPA Guidance for the data quality objectives process (US Environmental Protection Agency, 1994) defined a seven steps process called DQO (Data Quality Objectives), which was a systematic planning tool for establishing criteria for data quality and for developing data collection designs. This step by step procedures includes the following aspects :

 i. state the problem,

ii.   identify the decision,

iii. identify inputs to the decision,

iv.   define the study boundaries,

v.   develop a decision rule,

vi.   specify limits on decision errors,

vii. optimize the design for obtaining data.

By using the DQO process to plan environmental data collection efforts, EPA can assure that the type, quantity and quality of environmental data used in decision making is appropriate for intended application.

The CEC Manual of quality control procedures for validation of oceanographic data (UNESCO, 1993) defined a set of procedures for validating the data for meteorological and oceanographic variables. The validation step for data includes four major aspects :

-   instrumentation checks and calibrations,

-   the documentation of deployment parameters,

-   automatic quality control of data,

-   oceanographic assessment of the results of conditions of the previous aspects.

Till now, and according to the available literature on this subject, no previous similar work has been performed for fisheries data.

According to Council Regulation n° 1543/2000 (European Commission, 2001), the Member States were responsible for collecting both the fishery-dependent data the fishery-independent data. They are obliged to establish standard databases holding biological and economic information. Using these data, the advisory process consists of three stages : the collection of data and information, the analysis forming the basis for the advice and the formulation of advice. This decomposition shows the importance of data quality so that the Community has accepted greater responsibility for the collection of the data needed to operate the common fisheries policy. Whereas data collection had previously been the responsibility of Member States with some ad hoc funding by the Community, the Community is now developing programs to increase cooperation and coordination of data collection, with standardised procedures and criteria.

In 2003, the European Commission communication called "Improving scientific and technical advice for Community fisheries management" (European Commission, 2003) outlines the Commission's view of the needs for and shortfalls in scientific advice in fisheries regarding the proposed new common fisheries policy. In particular this communication insisted on the fact that *"the challenges in European fisheries management have substantially increased demand for up-to-date scientific assessments and advice. Additionally, increased reliance on scientific advice as the principal grounds for identifying and addressing problems has highlighted areas where the advice should be made clearer, more reliable and more credible"*.

In this general context, the Ifremer launched the SIDEPECHE program (*Systèmes d'information et techniques d'observation, économie et diagnostic de l'évolution des ressources et de leurs usages*) which aims to address "a better knowledge, evaluation and promotion of oceans resources enabling their sustainable exploitation" which is the mission of public interest of Ifremer. The SIDEPECHE program is an operational and multidisciplinary fisheries observatory in charge of data acquisition about fisheries resources and their use.

The SIDEPECHE objective is to take into account the whole fisheries system (biology in its ecosystem, fishing activities in its economic contexte), in all its components. To achieve this aim, a network of observers (from Ifremer and sub-contractors observers) was constituted all along the French Atlantic border as well as the North Sea Channel.

The SIDEPECHE program includes data acquisition, more often co-funded by regional, national and European institutions. Seven actions contribute to the constitution of an European observation network under the control of the EU (Data Collection Regulation - DCR).

Five major actions, linked to major thematic domains, correspond to all the activities necessary to a powerful observation system (sampling design and collection, qualification, storage and data restitution) (IFREMER, 2005) :

- Fisheries data

- Observations data at sea

- Landings data

- Survey data

- Economic data

Launched during the Nineties, the first quality management projects were gradually diffused so that they relate more than one quarter of Ifremer manpower. In order to consolidate confidence that users can have in the quality of Ifremer data, products and services, it was necessary to gradually generalize this approach with the whole of the institute activities.

The MEQUAPRO project (MEthodes, QUAlité, PROduits) aims to design and develop a generic and common approach, so that it could be used for the different group of data mentioned above, even if this approach is a very pragmatic one. The MEQUAPRO project is based on principles of quality management system like those defined by ISO9001:2000 international standard.

This project, launched in 2005, is based on a formal engagement of the general direction of Ifremer for all the institute which formalized and widens the needs for the researchers of Ifremer.

For now, one of the main problems of a quality approach for a fisheries observatory, is that there is still no quality objectives defined neither by national authorities nor by international one.

4. Methods

4.1. Data quality management system understanding

For MEQUAPRO project, the first step was to get a global vision of the data quality management system and, after that, decompose all the process.

An original organization was defined adapted to the following items :

- process definition and modelisation,

- working groups for each process to provide quality indicators,

- specific tools developed to facilitate a common and generic approach.

Working groups identified actors and process concerned by data quality management.

A "5 M approach" [1], which consists in reviewing factors which explain a phenomenon, was used.

Input and output data were defined for each process, as well as a comparative analyze allowed to identify positive points and those to be improved and also interactions between process.
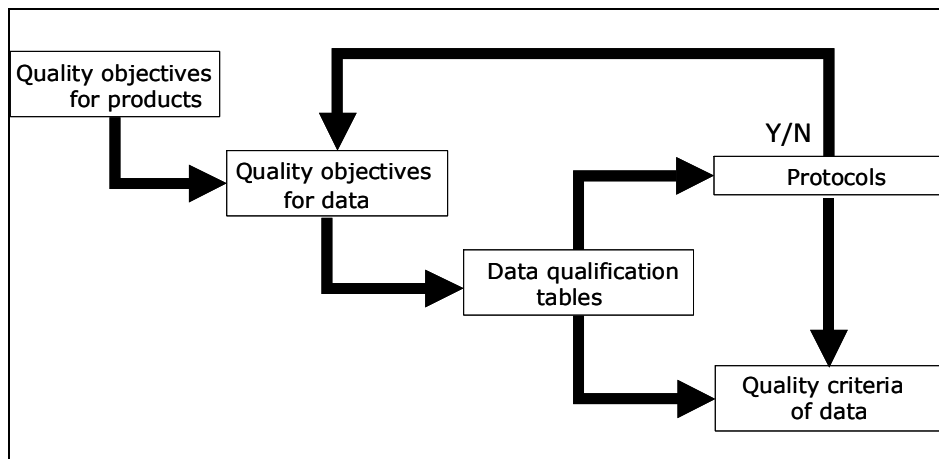
---

[1] Milieu (working environment), Méthode de travail (working method), Matière (raw material), Matériel (products), Main d'oeuvre (manpower)

4.2. Process approach and data qualification

The MEQUAPRO project were organized according to the process approach and mapping.
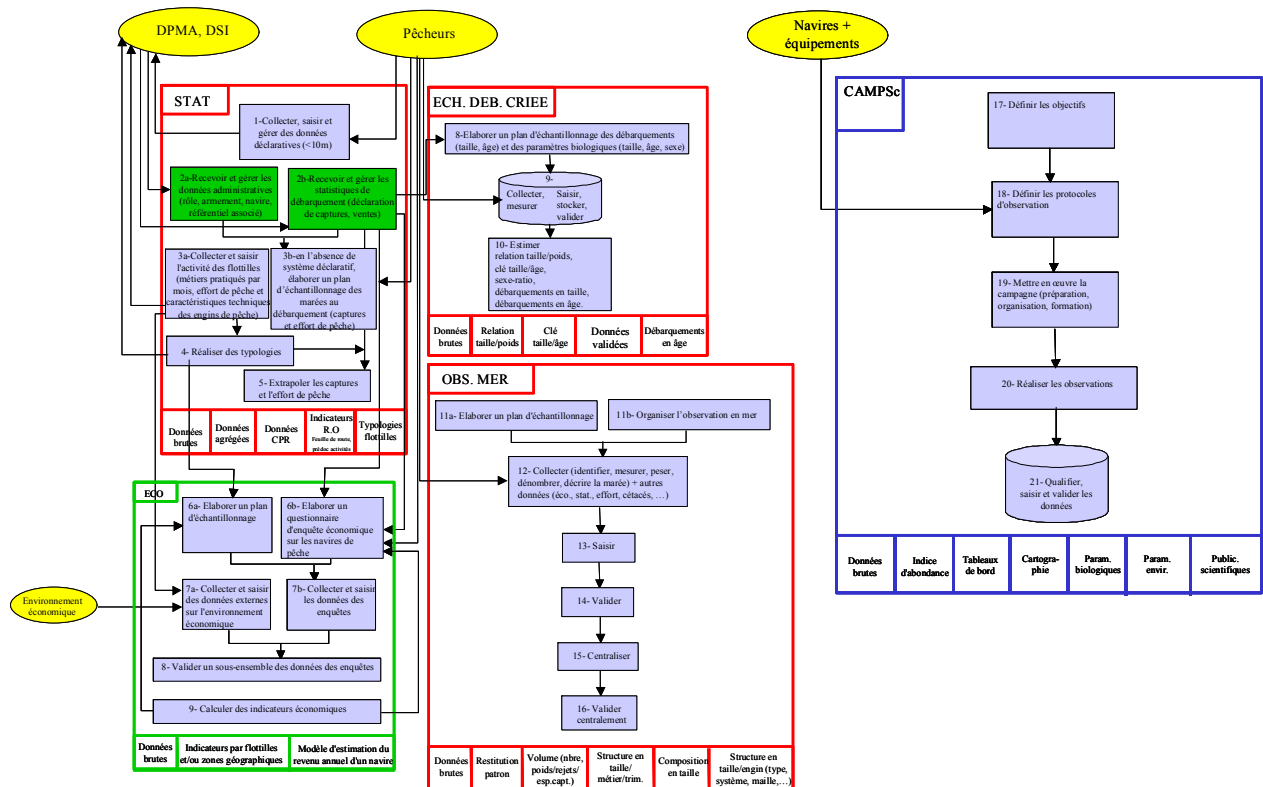
After that, quality objectives for data and products were identified as shown in figure 1.

Figure 1 : General principles of data qualification according to quality objectives for data and products

The next step is to identify quality indicators for data collected, stored and used for scientific studies and provided advices :

- to provide quality indicators,
- to develop methods to estimate the current indicators values,
- to compare the quality objectives.

A data quality indicator is a level of data quality to be achieved for a defined period. For example, it can be a percentage of data provided and/or maintained that have to be under quality limits for the next 12 months.

As it is shown on the figure 1, it's only when quality objectives are defined for products that it is possible to define those for data, and after that, to identify adapted quality indicators. These indicators need qualification tables to check if data provided are following existing protocols or not. If protocols exist and are correctly applied for providing data, it means that data may achieve predefined data quality objectives. Otherwise, other data quality criteria have to be defined.

4.3. Transverse approach

The observers network is participating to the certification process. Some of data are directly included into the Allegro software by observers so that it facilitates data qualification by type control and transfer of thematic controls to a computing control.
This approach allows also a combination of indicators and process in the aim to qualify transverse products.

4.4. The intranet and the wiki websites

To achieve requirements and objectives of the MEQUAPRO project, it was decided to develop two different tools, designed as very complementary as shown in the figure 2:

- an intranet website,

- a wiki website.
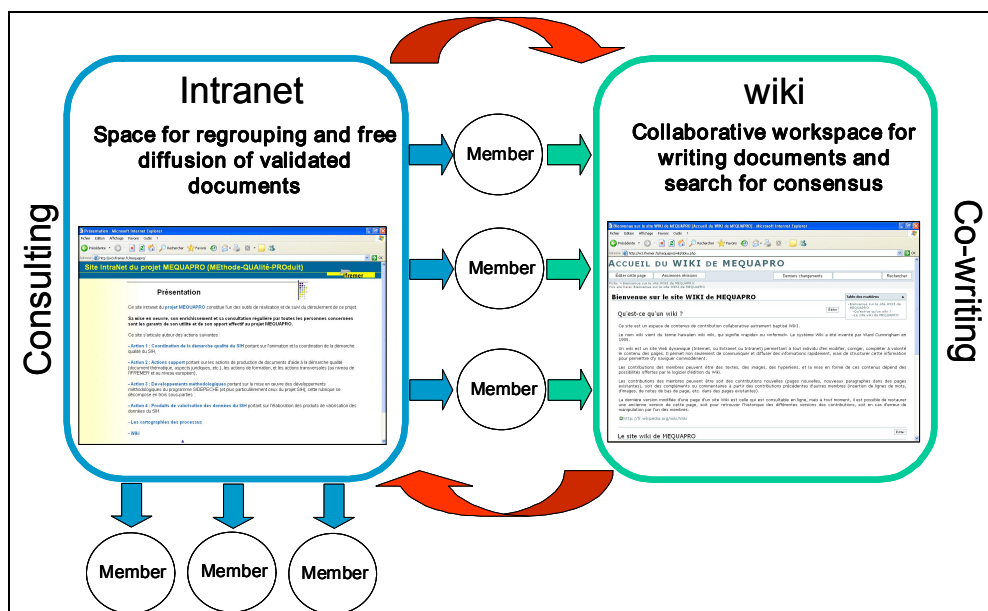
Figure 2 : Relationship and members role

for MEQUAPRO intranet and wiki websites use

4.4.1. The intranet website

The intranet website is dedicated to documents which are validated by members of project team so that other members of the project or from other Ifremer projects may consult and download a wide range of documents.

The intranet website is organised following the four major actions of the SIDEPECHE program mentioned above.

This two different and complementary tools were developed to facilitate the previous tasks (4.1 to 4.3).

4.4.2. The wiki website

A wiki website[2] is a "*type of website that allows users to add, remove, or otherwise edit and change all content very quickly and easily, sometimes without the need for registration[3]. This ease of interaction and operation makes a wiki an effective tool for collaborative writing. The term wiki can also refer to the collaborative software itself (wiki engine) that facilitates the operation of such a website*". In our case, we used Dokuwiki software[4].

Using a wiki, users provide HTML pages. Each change is stored allowing members to come back to a previous version if changes are not validated by the group, and historic changes are accessible to all members of the group. The wiki system provides users with an easy way to monitor the constantly changing state of the wiki, as well as a place to discuss and resolve problems with some of contents. Users may just consult pages written by others or they may become authors of new pages as well as they may modify or add some complements to existing pages.

5. Data

Ifremer provides a great number of data about fishes and fishing activities.

In the aim to identify if data may be validated, a qualification table model was defined and contains the following items :

-   Data definition

-   Origin

-   Population

-   Description

-   Domains of use

-   Possible validation methods   : measurement, estimation, calculation (including reference to a protocol number)

-   Qualification : current state and possible state in the future (qualification depending on the validation method used)

-   Possible values (a number code) and eventually a string corresponding the protocol number

0	unqualified
1	good
2	out of statistics
3	doubtful
4	false
5	corrected
8	incomplete
9	missing protocol reference

6. First results and perspectives

According to the first step about the data quality management system understanding, Ifremer researchers are working, for 12 months now, in workshops, one for each of group of data mentioned above.

---

[2] The word wiki is a shorter form of wiki wiki which is from the native language of Hawaii, where it is commonly used as an adjective to denote something "quick" or "fast".

[3] This definition was found on the famous Wikipedia website: http://en.wikipedia.org/wiki/Wiki

[4] 2004-2006 © Andreas Gohr. Sea Dokuwiki Website : http://www.splitbrain.org/go/dokuwiki.
The DokuWiki engine is licensed under GNU General Public License Version 2.

The first results are the following

- Process mapping organized by thematic data groups,

- 5M approach results,

- A glossary,

- Qualification tables for each data,

- Protocols for each raw data,

- Colours codes and associated values for qualification tables.

A glossary was produced which contains more than 220 terms and their definitions, proposed, update and completed by wiki users. For some of the terms, the FAO Fisheries glossary was used [5]. A glossary was considered as a condition to enable the MEQUAPRO project to de a success because researchers use many terms in different meanings, so that it appeared necessary to obtain a consensus about terms used for this project.

Data quality objectives are not so simple to be achieved because of its lack both from Ifremer researchers and from European institutions. The next Data Collection Regulation (DCR) will define some quality objectives which have to be defined for each set of data and planned on a calendar.

For now, data quality objectives definition is on progress for the MEQUAPRO project.

A common qualification table model was designed for sea observations data by one of the group of data mentioned above. About 30 qualification tables were fulfilled in a Word format and were translated into the wiki format.

---

[5] http://www.fao.org/fi/glossary/

Figure 5 : Example of a qualification table on the wiki website

Protocol tables are available by data and by groups of protocols as shown in the figure 6.

For data qualification, different conditions to attribute a value are possible as well as for test of thematic coherency for data acquired.

0	default value
1	respect of a protocol
2	(possible value at the time of ex-post control, corresponding to 1B level)
3	non-observed protocols or non-confident in the value given par the expert or measurement tool
4	(possible value at the time of ex-post control, corresponding to 1B level)
5	(possible value at the time of ex-post control, corresponding to 1B level)
8	incomplete
9	no data

-

A list of protocols necessary to validate data acquisition was also provided and included into the wiki website.
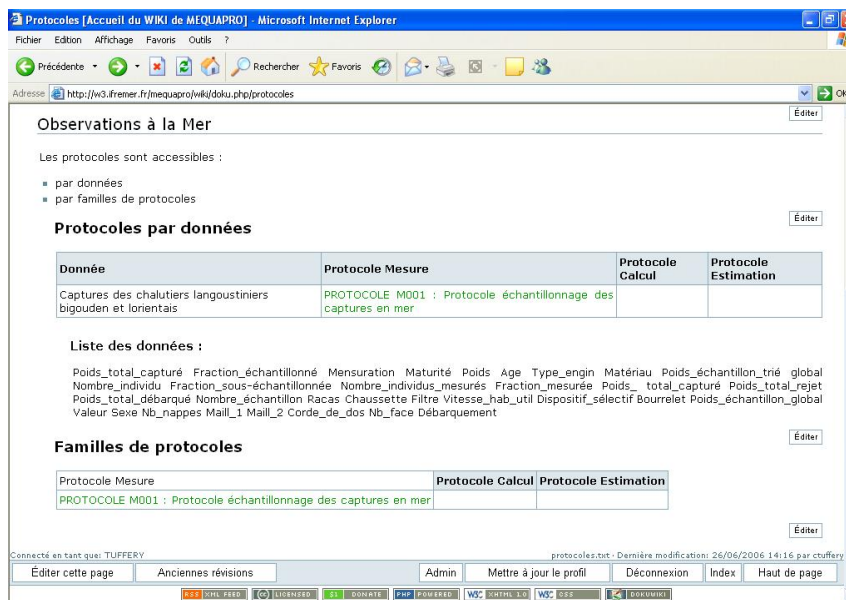

Figure 6 : Protocol tables by data and by groups of protocols

The figure 7 presents an example of a protocol (Protocol of measurement M001) defined for sea observations data. It is completed by a report produced after meetings with observers which are very useful documents describing how observers are actually working when they are on boats.
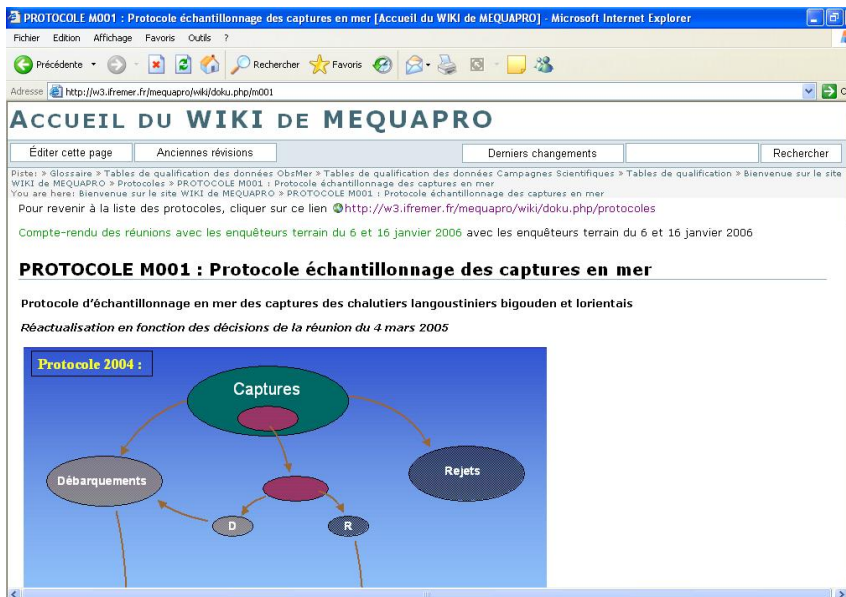

Figure 7 : Example of a protocol description on the wiki website

The next result attended is progress toward a standardisation of protocols.

One of the results of using the transverse approach mentioned above, are :

- an inquiry with observers,

- exchanges with ISIH project[6],

- Allegro software?

---

[6] Computing project of the SIDEPECHE program

One of the main problems is that SIDEPECHE data are provided by inquirers using fishermen declaration for the majority of them. It means that such data have to be validated using new and specific methods adapted to these data. In this mean, an investigation was launched near the observers to improve the knowledge of the data they obtained from fishermen.

The MEQUAPRO wiki website allows users to modify or update tables and also to add new tables or to delete any of them.

One of the main results already achieved by the MEQUAPRO project is a global and common framework for all contributions from Ifremer researchers in the aim to formalize how they work by now, both from theoretical and practical points of views. This work allowed to produce a map of all process concerned by the SIH project, as shown in figure 8.
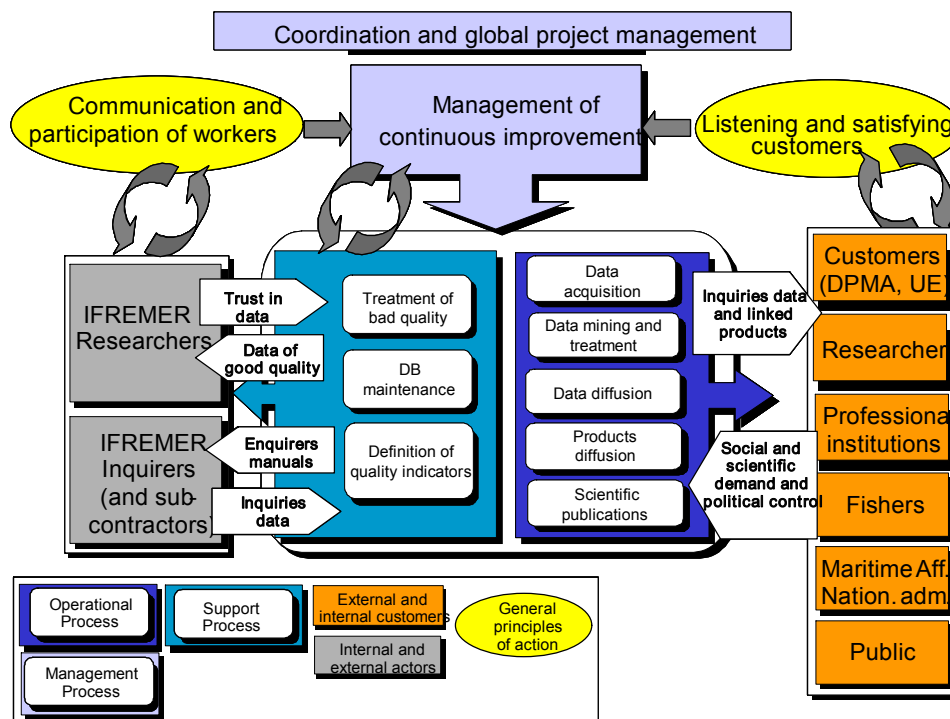
Figure 8: General process map for data involved

in Ifremer SIH (Fisheries Information System)

A major difficulty of the MEQUAPRO project is that it led people to work differently, because of the wiki website. This tool, rather new for the majority of Ifremer researchers, provides an original way of working together, to reach an online consensus.

Nevertheless, the first of the MEQUAPRO project are meeting some of the objectives of the ICES Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS), especially gathering national standard operation procedure for sampling commercial fisheries, as well as gathering survey manuals and standard operation procedures of abundance surveys.

The 2002 ICES' Strategic Plan specified two main goals for which data quality improvement is expected :

- *Modernise technologies and sampling designs for collecting, measuring, and enumerating marine organisms, and improve the precision and accuracy of resource surveys (Goal 1. Understand the physical, chemical, and biological functioning of marine ecosystems)*

- *Develop quality assurance protocols to enhance confidence in scientific advice (Goal 4. Advise on the sustainable use of living marine resources and protection of the marine environment).*

The ICES' Action Plan, 2003–2007 states that "ICES must ensure that all the components of the advisory process work together effectively" so that "ICES will apply quality assurance procedures to its advisory process… Further evaluate and implement quality control procedures for the Advisory Committees, particularly for ACFM and the Stock Assessment Working Groups. [ACFM/ ACME/ACE]".

Attended future results of the Mequapro project will try to contribute to all these objectives mentioned above.

When the first results for sea observations data are validated, it could be possible to enlarge this approach to other groups of data and process.

It would be also interesting to verify whether this approach has something in common with other national or international approaches in other domains (demographic statistics, geological, meteorological, cartographical data, etc.).

7. References :

European Commission, 2000. Council Regulation (EC) No 1543/2000 of 29 June 2000 establishing a Community framework for the collection and management of the fisheries data needed to conduct the common fisheries policy (OJ L 176, 15.7.2000, p. 1).

European Commission, 2001. EU Data Sampling Regulation. COMMISSION REGULATION (EC) No 1639/2001 of 25 July 2001, establishing the minimum and extended Community programmes for the collection of data in the fisheries sector and laying down detailed rules for the application of Council Regulation (EC) No 1543/2000.

European Commission, 2003. Improving scientific and technical advice for Community fisheries management. Commission from the European Commission, Official Journal of the European Union 2003/C 47/06, 12 pp.

ICES. 2006. Report of the Planning Group on Commercial Catch, Discards and Biological Sampling (PGCCDBS), 28 February–3 March 2006, Rostock, Germany. ICES CM 2006/ACFM:18. 62 pp.

ICES, 2005. Report of the Study Group on Quality Assurance. 27 pp.

ICES, 2002. Strategic Plan Adopted by the national Delegates of ICES on 1 October 2001.

ICES, 2002. Action Plan, 2003–2007, adopted by Council on 28 October 2002.

IFREMER, 2005. Ressources Halieutiques, Exploitation Durable et Valorisation. Dossier préparé pour la revue du Thème 4. Comité de Direction élargi du 17-05-2005.

UNESCO, 1993. Manual of quality control procedures for validation of oceanographic data / IOC; CEC. Paris : 436 p. ill. 30 cm. - Intergovernmental oceanographic commission. manual and guides, n° 26, 1993.

US Environmental Protection Agency, 2000. Guidance for the data quality objectives process / Washington DC. Pagination multiple. ill. 30 cm.