**UNITED NATIONS
ENVIRONMENT PROGRAMME
MEDITERRANEAN ACTION PLAN**

8 August 2019
Original: English

7th Meeting of the Ecosystem Approach Coordination Group

Athens, Greece, 9 September 2019

Agenda Item 6: IMAP Pilot Info System and Related Quality Assurance Issues; Data Standards and Data Dictionaries; MAP Data Management Policy

IMAP Pilot Info System: Quality Assurance and Quality Controls

Note by the Secretariat

In the framework of the Programme of Work and Budget for 2018–2019 of UN Environment/MAP (Decision IG.23/14), INFO/RAC is leading the work on the development of the "*Info/MAP platform and platform for the implementation of IMAP fully operative and further developed, connected to MAP components' information systems and other relevant regional knowledge platforms, to facilitate access to knowledge for managers and decision-makers, as well as stakeholders and the general public*" (output 1.5.1).

INFO/RAC, in full compliance with the needs of MAP system, and for the implementation of Integrated Monitoring and Assessment Programme of the Mediterranean Sea and Coast and Related Assessment Criteria (IMAP) is going to provide a web software platform for data and information management of all datasets related to marine ecosystem monitoring and assessment under its mandate.

The platform should allow:

- comply with the information requirements in the light of the obligations laid down in the Barcelona Convention;
- evaluate, define and monitor decision-making strategies through achieving the goal of adequacy and usability of the data.

Within the framework of this project, the creation of the IMAP Pilot Info System, including the integrated communication infrastructure, is of particular importance. The system will collect and integrate data from different sources and data providers and provide information to different target user groups.

The aim of this document is to describe the quality controls implemented in IMAP Pilot Info System and the Quality Assurance process that grants their actual application in order to effectively improve the data quality collected and published by the IMAP Pilot Info System.

Table of contents

I.   Quality Assurance and Formal Quality Controls

Data quality is not a product, but it is a process. Quality assurance is the process whose aim is to improve quality and it is based on the definition of what is good and what is not good.

The first step of QA process has been the definition of Data Standards (DSs) and Data Dictionary (DDs) and associated formal QCs for the monitoring modules associated to the selected 11 IMAP Common Indicators. A detailed description of the DS and DD is provided in document *Data Standards (DSs) and Data Dictionaries (DDs) for the selected core-set of Common Indicators*.
In a DS information is aggregated in different tables (represented by excel spreadsheets) and, for each table, several fields with different formats are defined. When a field has to be filled selecting a value included in a predefined list of admissible values, such lists constitute the Data Dictionary associated to DS. Data are compliant to DSs and DDs if and only if each of the following formal quality controls are satisfied:

a) **Format** - every field is compliant to its format, i.e. its value is text, numeric or date according to the required format;
b) **Unique coding** - codes used to identify each row of the spreadsheet are unique, i.e. there are no more than one row with the same code in the spreadsheet where such rows define the associated objects. For example, in the spreadsheet where stations are defined each row has a unique code that identifies the station. In general, there can be spreadsheets where information is linked to the stations and in such spreadsheets codes that identifies the stations are not unique. For example, the spreadsheet where concentration values for different parameters corresponding to one station are represented have the code of such station repeated for each concentration value;
c) **Coherent linking** - codes used to link information that is present in different spreadsheets have to be coherent. For example, in one spreadsheet there is the code of the station such code has to be present in the spreadsheet where stations are defined;
d) **Regular expression** - every field is compliant to specific regular expression when such regular expressions are required. For example, percentage field are to be filled by a numeric values in the format 0-100 or time field have to respect the expression HH:MM:SS;
e) **Admissible values** - every field for which there is a list of admissible values, is filled with one and only one values of such list, i.e. it is compliant to the DD associated to the DS;

Data Dictionary, i.e. the set of lists of admissible values, are represented in two ways. In one case the list has the value followed by its description after the '='. For example, the list of admissible values representing the matrix associated to chemical monitoring is the following one:

W = Water
S = Sediments
B = Biota

In this case, the admissible values to be used in the field 'Matrix' are 'W', 'S' or 'B'.
In the second case, the list is made by a table with more than one column as, for example, the list of contaminants:

ID_Contaminant	Label	CASNumber	Matrix	Mandatory
CAS_309-00-2	Aldrin	309-00-2	Sediments	Y
….	…	…	…	…

In this second case, the field that requires such list have to specifies which column of the list are referring to. For example, the field DeterminHazSubsName that refers to the list of contaminants, has to be filled with one of the values of column 'Label' of the list, while the field DeterminHazSubsID has to be filled with one of the values of column 'ID_Contaminant' of the list. Furthermore, DeterminHazSubsName and DeterminHazSubsID have to filled in a coherent way, so, if

DeterminHazSubsName is filled with value 'Aldrin', in the same row of the spreadsheet, DeterminHazSubsID has to be filled with value 'CAS_309-00-2'.

Data sets that are compliant to all the above formal quality controls from a) to e), are to be considered **formally compliant** or of good quality from a formal point of view.

The second step of QA process is the implementation of formal quality controls associated to each DS and DD in order to verify that data sets are formally compliant. In particular, for each DS that corresponds to a monitoring module, the above list of formal quality controls has been defined and implemented in the IMAP Pilot Info System. Each row in the list is represented by:

- **QC_Code**: unique code that identifies the quality control
- **Ecological Objective**: ecological objective to which the DS corresponds
- **Common Indicator**: IMAP common indicator to which the DS corresponds
- **DS/Module**: Data standards or monitoring module to which the quality control applies
- **Spreadsheet**: spreadsheet of the DS to which the quality control applies
- **Field**: field included in the spreadsheet to which the quality control applies
- **Description**: description of the quality control applied to the field

Here follows an example of the quality controls implemented for Data Standard P1- Contaminants in seawater.

QC_Code	Ecological Objective	Common Indicator	DS/Module	Spreadsheet	Field	Description
QC-E09-17-P1-00001	EO9 - Contaminants	Common indicator 17: Concentration of key harmful contaminants measured in the relevant matrix (EO9, related to biota, sediment, seawater)	P1 - Contaminants in seawater	Stations	CountryCode	Value corresponds to country code ISO 3166-1 alpha-2
QC-E09-17-P1-00002	EO9 - Contaminants	Common indicator 17: Concentration of key harmful contaminants measured in the relevant matrix (EO9, related to biota, sediment, seawater)	P1 - Contaminants in seawater	Stations	NationalStationID	Unique value in the spreadsheet

The process for the collection and quality control of data sets is implemented for each data standard by the typical *three hand shaking communication*:

1. Step 1: the user, a contracting party, downloads the Data Standard corresponding to the monitoring module for which he wants to transfer monitoring data

2. Step 2: after filling the Data Standard with monitoring data, the user uploads the file into the system for the data flow which corresponds to the Data Standard used
3. Step 3: The system produces a report of QC formal check validation with the results of formal quality control applied to the file uploaded and if every quality control is passed, the file is considered as 'formally compliant' (*OK!*) otherwise (*Not OK*) the user has to correct the file and upload it again into the system in order pass all the formal quality controls

The following picture schematically represents the Quality Assurance process to improve the quality of data sets transfer to the systems:
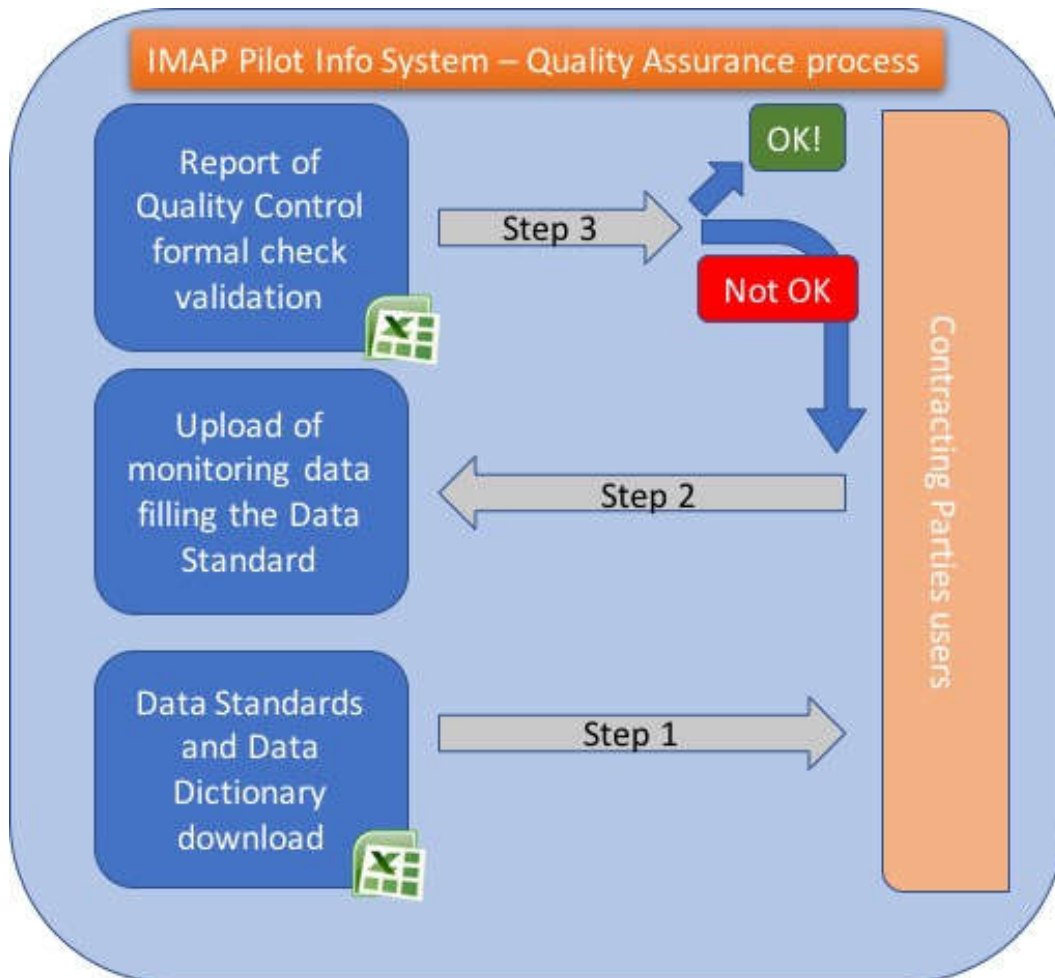
The report of QC formal check validation is produced as an excel file containing the following information for each row that do not pass the quality control check:

- **QC_Code**: unique code that identifies the quality control that has not been passed
- **DS/Module**: Data standards or monitoring module to which the quality control applies
- **Spreadsheet**: spreadsheet of the DS to which the quality control has not been passed
- **Field**: field included in the spreadsheet to which the quality control has not been passed
- **Row**: number of row of the spreadsheet containing the field for which the quality control has not been passed
- **Description**: description of the quality control that has not been passed

The following table contains an example with two rows of DS - P1 -Contaminants in seawater that are not compliant:

QC_Code	DS/Module	Spreadsheet	Field	Row	Description
QC-E09-17-P1-00001	P1 -Contaminants in seawater	Stations	CountryCode	34	Value corresponds to country code ISO 3166-1 alpha-2
QC-E09-17-P1-00002	P1 -Contaminants in seawater	Stations	NationalStationID	45	Unique value in the spreadsheet

In the above example, the user has to correct what has been inserted in rows 34 and 45 for spreadsheet Stations respectively in fields CountryCode and NationalStationID in order to make the file formally compliant. It is also clear that unique coding of each formal Quality Control and production of detailed reports on the results of their applications, are indispensable tools of the Quality Assurance process to improve data quality.

II.   Future developments for Higher Level Quality Controls

Data sets that are formally compliant, i.e. they pass the formal Quality Controls of the previous paragraph, are stored in the Relational Data Base Management System (RDBMS)of the IMAP Pilot Info System that also includes a GIS extension. Once data is collected on such infrastructure and are representative of Mediterranean basin, it is possible to define and apply higher level Quality Controls that regard, for example, the following issues:

1. Check of admissible ranges or maximum or minimum values for parameters based on statistical analysis of monitoring data, on scientific literature reviews or on fixed constraints due to physical or chemical characteristics as for ex. the range 0-14 of pH;
2. Geographical location of monitoring points to be included in Mediterranean Sea;

But an overall Quality Assurance process for data quality also includes additional higher level Quality Controls on different phases of monitoring as:

- Sample collection
- Sample processing
- Analytical determinations

The above elements are directly linked to laboratory data quality and control procedures which consist of accreditation process, use of certified reference material or standardized monitoring protocols and participation of laboratory to proficiency testing.

The IMAP Pilot Info System has been designed to implement in future phases both higher level Quality Controls of type 1) and 2) and also to collect additional information on laboratory data quality and control procedures and application of monitoring protocols. Such information will be included in specific DD and DS and linked to monitoring data in order to apply a categorization for flagging data sets for EO5 and EO9 on the example provided in document UNEP/MED WG.473/9 - *Schemes for Database Quality and Quality Assurance and Quality Control (QA/QC) of Data related to Pollution:*

- Category A. Laboratories/CPs reporting successful Proficiency testing (z-score<2) and/or accreditation for the chemical or parameter analyzed; metadata completed and timely submitted (max2 years delay).
- Category B. Laboratories/CPs reporting Proficiency testing for the chemical or parameter analysed (2<z<3) and/or accreditation; metadata completed and timely submitted (max2 years delay).

- Category C. Laboratories/CPs with no participation in Proficiency testing (for the last 2 years); metadata completed and timely submitted. It also could include scientific literature with full QA reported.
- Category D. Laboratories/CPs with no participation in Proficiency testing (for the latest 5 years); metadata completed but not timely submitted. Also includes scientific literature without QA specifically reported.
- Category E. Laboratories/CPs with gross reporting errors, although metadata might be completed and timely submitted.

III.   Conclusions

The IMAP Pilot Info System has implemented a Quality Assurance process that includes formal Quality Controls for the selected 11 IMAP Common Indicators and it is in the position to further implement higher level quality controls if a) representative monitoring data will be available in the system and b)detailed information on laboratory data quality and control procedures and application of monitoring protocols will be provided by Contracting Parties.