UNITED NATIONS

EP

UNEP/MED WG.566/4

UNEP

UNITED NATIONS ENVIRONMENT PROGRAMME MEDITERRANEAN ACTION PLAN

10 May 2023
Original: English

Meeting of INFO/RAC National Focal Points

Rome, Italy, 7-8 June 2023

Agenda item 4: **Main elements of the MAP Knowledge Management Strategy: towards a Knowledge Management Platform for the Mediterranean Seaand its linkages with the Digital Trasformation process.**

Main elements of the MAP Knowledge Management Strategy: towards a Knowledge Management Platform for the Mediterranean Sea

Disclaimer: The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

In charge of the activity at INFO/RAC

Lorenza Babbini, Director

Arthur Pasquale, Deputy Director

Annalisa Minelli, Knowledge Management & Digital Trasformation Officer

Document prepared by:

Lorenza Babbini, Annalisa Minelli, Arthur Pasquale.

Main elements of the MAP Knowledge Management Strategy: towards a KM platform for the Mediterranean Sea

Objectives for the strategy

A successful Knowledge Management strategy must start from the complete mapping of existing knowledge to manage. UNEP/MAP collected in more than 40 years a wide variety of data and transformed it into information and knowledge to a point that its management resulted to be an obliged step to better catalog, search and retrieve that data. Evidently, in more than 40 years also methods of analysis and collection techniques changed, and this reflects directly on data quality.

The more consistent the data is and spans over decades, the more a strategic view is necessary. If data management consists in extemporary actions to be taken after necessities, it will not be able to encompass all the possible upcoming situations and this will result in a fundamental impossibility to deal with some specific occurrences, unforeseen at strategical level. On the other hand, a wise strategic view set up from the beginning makes it less probable that emergency situations will occur and require extemporary actions.

The **key point** is to foresee a set, the wider than the possible of use cases and possibilities for knowledge management with the final aim to be prepared to manage knowledge in any situation.

The objectives of UNEP-MAP Knowledge Management Strategy are:

OBJ1.*Individuation of guiding principles underlying the strategy*. These principles are defined to increase accessibility of information across and outside the MAP. The definition is made by taking into consideration some **internationally recognized pillars** for information management and applying them to the UNEP-MAP specific case, starting from the objects handled by the MAP, such as: data, information, knowledge, metadata and infrastructures.

OBJ2.*Set a path to catalogue and harmonize data, information and knowledge across UNEP-MAP*. Cataloguing and harmonizing data, information and knowledge is the first step for a fair knowledge management. In particular, objects must be identified, quantified and qualified to choose the more appropriate management actions and tools. Harmonization also is an important step towards a fair data management since it allows to limit the actions/tools required and allows to define some common rules for specific type of data, information and knowledge ingestion, management and sharing. If harmonization is done following **internationally agreed standards** among specific communities, the handling of products is way easier.

OBJ3.*Fix sharing common practices for UNEP-MAP knowledge* in a way that MAP heritage could be reused more than the possible. Common sharing practices are at the base of potential reuse of knowledge. In particular, if **sharing practices** are the ones agreed among international community, they also improve interoperability among systems.

OBJ4.*Trace a path to ease the application of UNEP-MAP data policy at different levels* in a way that different actors can easily understand **possibilities and limits** for their specific access level.

OBJ5.*Create a unique access point to UNEP-MAP knowledge* to become a **reference to access knowledge** in the wider Mediterranean area.

OBJ6.*Create a cooperation network*, composed by **UNEP-MAP components, Contracting Parties, and relevant stakeholders** in a way that each one of them contributes in some way to implement the strategy, raises observations and constantly work together to improve the strategy itself.

Guiding principles

Recalling Data Policy (UNEP/MED WG.512/4) and Data Management Policy (UNEP/MED WG.470/5) adopted by the MAP, the FAIR data management rises as a pillar for modern data management. The FAIR data management allows for **wise sharing of data** and rely on **basic principles** commonly accepted across international regulations and standards (which effectively put in practice these same principles).

In a period of data overload and over production (each device is potentially able to generate big quantities of data), the data management should be based on stricter rules and rely on new principles and best practices accounting for the volume of data (big data case) related to their value. In fact, **not all the data are valuable** and **not all the data deserves to be preserved over time and managed**.

The FAIR data management theory, firstly introduced by Wilkinson et al. in 2016, suggests that three types of products: data, metadata and infrastructures should be managed in a way that they are:

- **Findable**: data must be indexed from the search engines, they must be identified by a univocal and persistent Uniform Resource Identifier (URI).
- **Accessible**: access procedure to data must be clear, metadata must be standard and identified by a persistent identifier that, eventually, survives data.
- **Interoperable**: data and metadata must be shared using standard formats or vocabularies, allowing for machine reading them in a way that they can interact with other applications/tools for analysis, storage, and processing purposes.
- **Reusable**: data must be extensively described, meet domain-relevant community standards, shared using a license that allows for a wider and real reuse of data (data manipulation and recombination must be allowed).

These **four principles** find a practical application in a set of rules and practical examples detailing the principles put in place by the go-FAIR initiative that are:

For Findability: F1. (Meta)data are assigned a globally unique and persistent identifier; F2. Data are described with rich metadata (defined by R1 below); F3. Metadata clearly and explicitly include the identifier of the data they describe; F4. (Meta)data are registered or indexed in a searchable source.

For Accessibility: A1. (Meta)data are retrievable by their identifier using a standardised communications protocol which is open, free, and universally implementable (A1.1) and it allows for an authentication and authorisation procedure, where necessary (A1.2); A2. Metadata are accessible, even when the data are no longer available.

For Interoperability: I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation; I2. (Meta)data use vocabularies that follow FAIR principles; I3. (Meta)data include qualified references to other (meta)data.

For Reuse: R1. (Meta)data are richly described with a plurality of accurate and relevant attributes: they are released with a clear and accessible data usage license (R1.1), associated with detailed provenance (R1.2), they meet domain-relevant community standards (R1.3).

These rules are fundamental for the implementation of data FAIR principles and their observance for any aspect of data management leads to complete compliance.

From a conceptual point of view, FAIR data management aims, first of all, to increase the value of data, starting from the assumptions that (i) **data is valuable as long as it is usable**, and (ii) **data are at the base of economic growth and they can represent a flywheel, once shared**.

The first assumption, in particular, implies that not only the scientific process leading to dataset building but also data formalization must be rigorous. In fact, **if data are "formatted" in an appropriate manner**, even better if based on standard common practices, and metadata are rigorous (also in this case, based on standards), **their reuse is easy** because they are easy to read (metadata are

understandable and could optionally be read by other machines), and data format is based on rules that are easily applicable and agreed by the scientific community.

The second assumption underlies the intrinsic value of data and, in particular, open data. In fact, data are historically a powerful instrument (at the service of public and private stakeholders) but the frequent limited access to them concentrates the knowledge only in few, wealthy, hands which become principal actors of the progress. **Open data**, conversely, **pushes towards a democratic process** where anyone has potentially the same instruments to emerge it allows anyone to benefit from the same level of completeness (and value) of data. Where anyone could be intended, among others, as small private enterprises, less rich countries' administration, less rich countries' researchers, communities of interests and private citizens.

The principles of FAIR data management (Wilkinson et al., 2016) have increasingly become worldwide shared rules for any data producer/manager and it is commonly accepted to make data FAIR with the aim to open more than the possible the global knowledge to the greater part of humanity. Likewise, it is obvious that **not all the data can be shared for multiple reasons**, which are all equally acceptable: data containing private details of human beings, data putting at risk some rare species of animals, data secreted for sake of state states' security etc. etc. For this reason, the motto at supranational level during the social transition to the idea of "data as a common good" have always been: "as open as possible, as closed as necessary" (Science Europe et al., 2018).

Strategy implementation: The Knowledge Management Platform

Evidently, principles must be supported by instruments, otherwise they remain empty concepts. One and the probably most precious ally for the strategy is the **MAP Data Policy** (UNEP/MED IG.25/27, Decision IG.25/10) which defines the terms, and remarks the principles under which data is collected, harmonized and shared among the MAP. Another useful instrument is the **Data Management Plan**. The Data Management Plan is the basic instrument that forces the data manager in a deep data analysis work for what concerns the aspect of management: *which is the format of the dataset? what is/was the data collection process? which is the value of data? who is the responsible for data management? where data should be stored?* these are only some of the questions rising when compiling a new Data Management Plan and the plan itself should represent a sort of identity card for data: as detailed as possible. This plan must evolve with data (*are data collection techniques evolving and influencing data quality? is the volume of data increasing in a way that it influences the storage method?)* and should be updated with a predefined frequency in order to make data FAIR.

To this purpose INFO/RAC performed over the last year a wide work of **data discovery, quantification and qualification, gave input to a metadatation process based on standard common practices, pushes towards data harmonization both for structured and unstructured data** (i.e. geographical layers and documents). The Data Management Plan will be then written accordingly to this thorough work of recognition carried out by the UNEP-MAP Data Management Task force.

The general view of both the Data Policy and the Data Management Task Force is in line with FAIR data management principles and the final product of the present strategy will be the Knowledge Management Platform (KMP).

Objectives

With the Knowledge Management Platform INFO/RAC is working with the vision of creating a unique, standardized, centralized access point for all the relevant data, information and knowledge in the Mediterranean Sea.

According to the **DIKW** pyramidal scheme (**Error! Reference source not found.**), we refer to:

- Data as discrete and objective facts, not related to the context.
- Information as data in context, capable to enlighten the meaning of data.
- Knowledge as information elaborated with experience and expert insights.

Figure 1 - DIKW pyramid scheme.

UNEP-MAP historically deals with many different duties both of reporting (for environmental data) and dissemination (to spread knowledge on Mediterranean Sea at different levels). This implies that UNEP-MAP heritage is quite various, and the products delivered for any necessity could be (only citing some): survey data, elaborated data, maps, best practices and guidelines, leaflets, videos, training material. INFO/RAC **firstly attempted a classification of this material** to better understand which kind and how many different managing measures must be taken into account. From this classification we built an architecture for the Platform, relying on the instruments already in place and looking for new and innovation tools.

Some of the instruments useful for this purpose are already existing in the infoMAP System managed by INFO/RAC: the Data Centre is developed and fully operational to host data from all the RACs, InfoMAPNode reached a sufficiently mature state of evolution to become a reference interface for data cataloging, visualization and map composition, so that the KMP will represent, by the end of the 2023, the unique access point for all the MAP knowledge.

Implementation steps

For 2023 it is in course of delivering the prototype of the KMP that will be articulated in three consecutive steps concerning a **feasibility study**, a **design procedure** and the **definition of methodologies and techniques tied to the design of the KMP**, using innovative instruments.

The three steps are organized in a way that they concern (i) definition of users and requirements, (ii) analysis, design and development of packages composing the prototype, (iii) building of the final prototype and platform testing. The timeline detailing the delivery of the prototype is reported in (Figure 2).

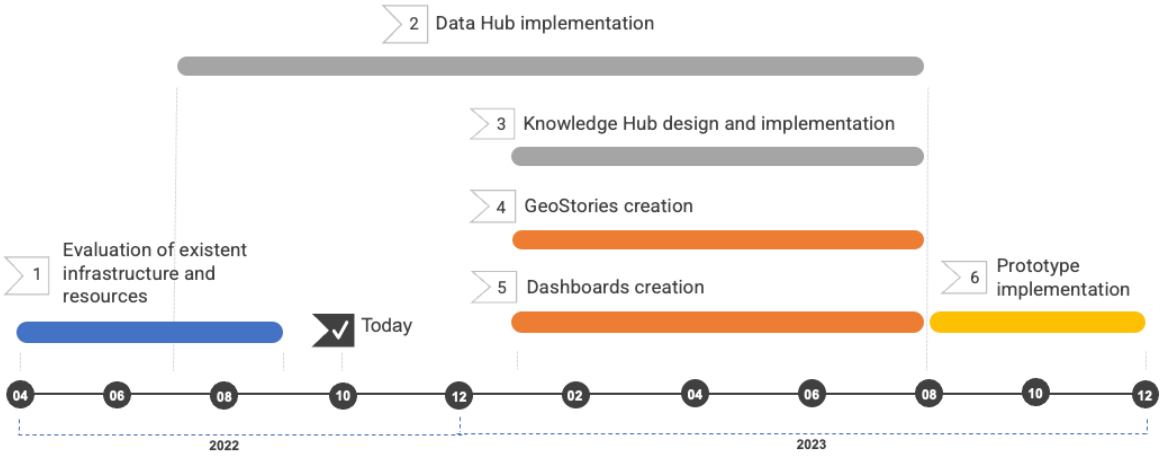

Figure 2 - General roadmap of KMP prototype implementation.

Phase 0: During this phase a set of necessary and preparatory activities are foreseen. Such as the definition of different types of users and their requirements, the study of usefulness of data with respect to users' potential interests, and the exemplification of some use cases for data and users. These actions are aimed to ease the definition and characterization of data fluxes to better implement the products foreseen in the following phase: Dashboards, Geostories and Knowledge Hub.

Phase 1a: During this parallel phase Dashboards are being created. Dashboards are flexible instruments to query data from users that can directly interact with the tool receiving a summary of georeferenced data, time series data, textual data, structured or unstructured data in a manner which is coherent with the user requirements previously defined. Dashboards are also capable of representing data using specific visualization techniques, filter data with respect to themes of interest for the user, eventually download filtered data.

Phase 1b: During this parallel phase Geostories are being created. Geostories are multimedia products composed by interactive maps, descriptive parts, multimedia objects (such as videos, or photos), integrated websites, aiming to introduce, also to a less experienced user, UNEP-MAP environmental themes of interest, projects carried out and results obtained. Key characteristics of an effective Geostory are: intuitiveness and ease of navigation through the story, suitability of contents, variety of tools and integrable multimedia objects.

Phase 1c: During this parallel phase the Knowledge Hub (KH), an appropriate system managing the MAP documental heritage, is being put in place. The Knowledge Hub will collect both structured and unstructured data, documents and multimedia products, metadata and tables coming from different information sources with the aim to enable the user to search products by their content. Compulsory information in input to this phase are: characterization and definition of information sources, definition of use cases for the KMP, description of potential requests from the users, and the type of attended response from the platform. Thanks to these preliminary actions, it is possible to identify the information and knowledge retrieval models (IR and KR respectively) fitting better for the building of the KH. All these aspects involve semantic management of knowledge.

Phase 2: During this phase (which follows parallel Phases 1a, b and c) the prototype of KMP is being implemented and will be completed with a homogenized visual interface where both the parts (Data Hub and Knowledge Hub) function together as one. A first prototype is foreseen to be subject to receive feedback from an expert user group in a first testing phase that will be finalized with the final release of the proptotype.

Foreseen functional levels

The Knowledge Management Platform usage is based on the definition of potential users and their requirements. To this purpose, we defined some users' profiles that could interact with the KMP:

- *Platform administrator*: the platform administrator is INFO/RAC. It takes care of data protection and system security. The administrator can, equally, upload and delete all the data of the platform; it can view and download all the data and set access privileges for each group of registered and non-registered user.
- *Contracting Parties*: the contracting party user is able to feed the platform with data from its Country, so this kind of user has some granted privileges, such as to upload, edit and delete its own material from the platform, while he will not be able to edit or delete other users' material. He can also access restricted data (in view and download mode) under specific circumstances that require that privilege. The users from the same Country can also be represented as a group, inside which different access levels could coexist after the competence of the user.
- *UNEP-MAP components*: similarly to what happens for Contracting Parties, Regional Activity Centers composing the MAP system can upload and delete their data from the platform. They often participate in data harmonization and rework, so they can access (view and download) data under higher privileges than the unregistered users. Nevertheless, they cannot edit or delete other authors' material.
- *Stakeholders and Researchers*: this profile identifies a group of or single users that have the necessity (for different reasons) to access MAP data. For example, a researcher that needs specific data for research purposes, or an enterprise which collaborates whit UNEP-MAP for data handling. Ideally the profile belongs to a registered user that cannot upload and edit/delete data to/from the platform but can view and download also restricted data, where necessary.
- *Anonymous users*: the group of Anonymous users are non-registered users navigating the platform with no specific duty with respect to the platform itself. They represent users who are not identified, although they have the possibility to search, view and in some cases download metadata and data which are publicly available.

Specific attention is paid to the **communities of interests**, which are communities of people interested in some way and with different objectives into UNEP-MAP knowledge. Some examples could be fishery committees, scuba diving associations, citizen science associations and many others. It is already clear from this first raw exemplification that the user type in this case is not evident, so permissions will be established evaluating from time to time user requirements and UNEP-MAP availability and interest in the specific matter.

General architecture

The KMP is designed in a way that back-end architecture clearly emerges in the front-end (Figure 3). Starting from data, information and knowledge MAP heritage, considering possible uses of this material, INFO/RAC depicted an architecture based on three main parts: one part dealing with geographical data (where for geographical is intended whatever information that could be geo-localized), the **Data Hub**, one part dealing with documents, the **Knowledge Hub** and one part dealing with interaction among UNEP-MAP users, the **Knowledge Exchange** (Figure 3). While the first two sections (Data Hub and Knowledge Hub) are part of the prototype that will be delivered in December 2023, the Knowledge Exchange is foreseen to be implemented in 2024.

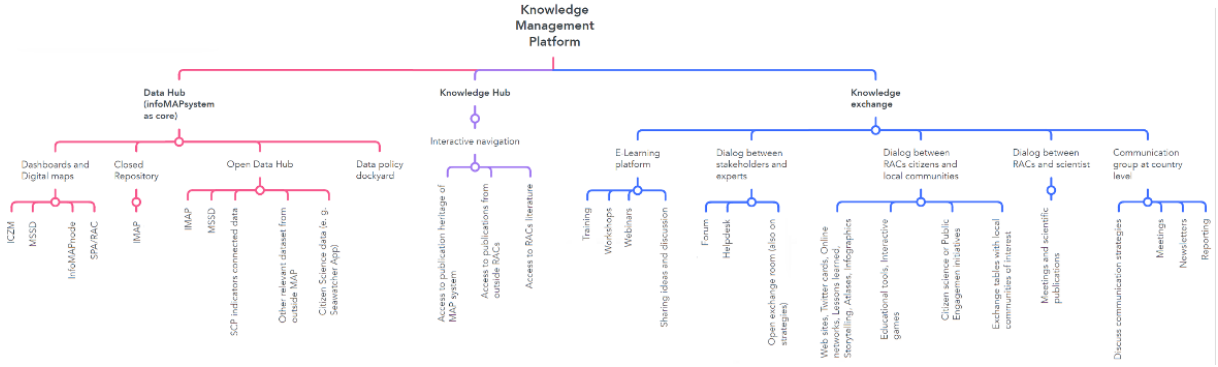

Figure 3 - General architecture of the final KMP.

Basically, the tripartite model depicted in Figure 3 is being encapsulated into a Content Management System (CMS) guiding users' navigation through data and documents by using a set of isntruments as the ones described in the previous sections (GIS viewers, Geostories, Dashboards, Kowledge networks, etc.). Qualitatively, the system is designed to support a traffic of at least 500 single users per month, but the hythoesized infrastructure is sufficiently flexible to support eventual modifications.

Source data flows

While KMP foresees to become a reference hub for data, information and knowledge in the wider Mediterranean area, this also includes data from outside the UNEP-MAP system. Nevertheless, before concentrating any effort in the research, metadatation, collection and organization of external data, it is crucial to complete the integration of the available data within the UNEP-MAP system, pertaining to different relevant environmental themes (biodiversity, coast and hydrography, pollution, marine litter, socio-economic sector, etc.). A characterization of data sources is therefore compulsory and this was performed over the last year through the valuable contribution of the MAP Data Management Task force.

For what concerns geographical data, **InfoMAPNode** represents the core of the KMP Data Hub. In this perspective, it has been upgraded, the user interface has been refreshed, and new functionalities have been introduced. The principal two data fluxes coming from UNEP-MAP data heritage are composed by: (i) layers uploaded in the underlying instance of Geoserver and (ii) layers being fetched by a certain number of remote services. Only a small part of available data is stored in local Geoserver instance behind InfoMAPNode, following the non-duplication of efforts and resources principle. This implies that the availability of layers is strongly related to the persistence of the data source in particular, and the FAIR data management principles application by these sources in general. In fact, the persistence of the identifier (of the remote service, in this specific case) is one of the pilllars of the Findability of a resource. The application of this specific principle grants the stability of InfoMAPNode and other eventual mirroring infrastructures.

For what concerns data flow in general, data present in considered sources is usually collected by contracting parties or stakeholders commissioned from contracting parties, data is then harmonized and sometimes reworked by the RAC or the CP, and subsequently data is shared. RACs are in charge of data quality assessment and control.

Data available in infoMAPNode are not currently under embargo, nevertheless INFO/RAC foresees the possibility to put embargo on specific data sensitive in terms of privacy or involved in legal issues.

For what concerns documental heritage, a fragmented reality of tools from RACs and UNEP are currently available with a non-homogeneous refinement level of data management. The whole UNEP-MAP documental heritage is currently not harmonized and rarely shared via standard and interoperable services. The Knowledge Hub from KMP aims to be a common access point to all these tools and resources available online by harmonizing this heritage and, following the non-duplication of effort principle, linking these documents to the repository where they reside. Documents will be analyzed and

widely metadated via semantic knowledge instruments, also taking advantage of ontologies, and will be made available through a user friendly and intuitive interface.

Technical assets

For the development of the Data Hub, a first core is already present, and it is represented by InfoMAPNode. InfoMAPNode, the platform managed by INFO/RAC collecting geographical data from UNEP-MAP and outside the MAP, relies on GeoNode for the front-end and Geoserver for the back-end. The whole system has been recently upgraded and the actions taken on the InfoMAPNode involving the Knowledge Management Platform building are focused on:

- Integration of all data coming from UNEP-MAP (both via upload in the backend Geoserver and via link to remote services).
- Make metadata INSPIRE compliant for all the data in the platform.
- Review licenses for all the layers.
- Restyling users and requirements defining the right permission for any individuated group, modify the old interface of InfoMAPNode to create a unique visual identity for the Data Hub and the KMP.
- Create guided path for users to ease navigation.
- Create Dashboards to view and analyze data accordingly to data type and themes.
- Create Geostories as narrative instruments to spread UNEP-MAP knowledge and work on relevant themes.

All the data present in the Knowledge Hub is shared following rules defined and agreed by the international community, following OGC standards for geographical data.

For the development of the Knowledge Hub, a brand-new component, based on semantic knowledge instruments is being built which would be able to harmonize and connect knowledge sparse among the different document sources present. This tool is ideally able to efficiently harvest information and metadata from different sources, index and classify documents based on keywords, filter documents based on tags with the aim to achieve:

- A simple architecture, easy to manage, and rapid in information and knowledge retrieval operations.
- Intuitively use both from back-end and front-end.
- Possibility to define major permissions for certain type of users.
- Interaction with the Data Hub and common visual identity with the whole KMP.
- Connection to other catalogs and platforms existent in the Mediterranean Sea.

All the documents present in the Knowledge Hub are identified univocally by a Digital Object Identifier (DOI).

Towards a complete MAP Knowledge Strategy for the Mediterranean Sea

In conclusion, UNEP-MAP knowledge management strategy needs to trace a clear way to manage data, metadata, knowledge, and infrastructures in the MAP BC system. The strategy is based on internationally recognized principles, and it is intended to be updated after eventual emerging application strategies of these principles, with the final aim to grow interoperability with other systems (internal or external to the MAP), promote reuse of MAP products and promote the UNEP-MAP work in general. In this perspective, the writing of an appropriate Data Management Plan is essential because it grants the durability of a FAIR data management. For this purpose, continuous work from INFO/RAC side, with the collaboration of the MAP components, is required. Actions in this sense could be (i) data, metadata, infrastructure harmonization following standards; (ii) application of individuated best practices on new products; (iii) deep comprehension and reasoning of the pillar documents of UNEP-MAP Data Policy and Knowledge Management Strategy to ensure the right application of the delineated principles.