EP

# Programme des Nations Unies pour l'environnement

UNEP(DEC)/MED WG.268/Inf.19
30 April 2005

ENGLISH
ORGINAL: ENGLISH

**MEDITERRANEAN ACTION PLAN**

Seventh Meeting of National Focal Points for SPAs

Seville, 31 May - 3 June 2005

# The MedGIS Database

**For reasons of economy, this document will be available in a limited number at the meeting. You are kindly requested to bring your copy to the meeting and not to request additional copies.**

UNEP
RAC/SPA - Tunis, 2005

Mediterranean Action Plan
Regional Activity Centre for Specially Protected Areas (RAC/SPA)
Boulevard de l'Environnement
B.P.337 –1080 Tunis CEDEX
E-mail : car-asp@rac-spa.org

The original version (English) of this document has been prepared for the Regional Activity Centre for Specially Protected Areas (RAC/SPA) by:

Ms. Stefanie Weykam
Expert on GIS applied to
Biological Conservation

# The MedGIS database

## PREFACE

MedGIS project is the result of a common will by MAP-RAC/SPA & UNEP´s Regional Seas Programme. The project has been undertaken to set up a Mediterranean Geographical Information System and a web map service (dynamic atlas) for the Mediterranean countries.

This tool, advocated to be used by all the MAP components, has become a tangible result thanks to the kind contribution of UNEP Regional Seas Programme at Nairobi. The map service is already available in Internet and it provides through queries tables with the information existing in the Database linked to the maps displayed.

This project contributes to address the recommendation of the contracting Parties to the Secretariat "to make the SPA/RAC databases available on the Internet and circulate GIS data and, in collaboration with the CBD Secretariat, make efforts to establish a Mediterranean clearing-house mechanism on marine and coastal biodiversity that will network with the national clearing-house mechanisms and others set up within the framework of the CBD". It provides a tool to managers for integrating every useful information regarding biodiversity, human factors and their geographical location in the Mediterranean; Its further development may allow undertaking prediction analysis on trends concerning biodiversity topics. The project acts as a demonstration one able to be extrapolated to other Regional Seas.

It must be stressed the fact that this tool is intended to facilitate information to all the MAP components, including RACs, MAP Programmes and the Contracting Parties. It is a most needed synchronous tool for a proper management and optimisation of data collected by them through diverse projects and other activities;

MedGIS may aid on: centralisation of information on the Mediterranean environment in relation to spatial parameters; updated compilation of existing information on protected areas, sites of conservation interest, biodiversity, and human factors in the Mediterranean.

It is intended to make it interoperable and in compliance with other international programmes and initiatives, such as the UNEP- World Conservation Monitoring Centre (WCMC), the European Network on Biodiversity Information (ENBI), the Global Biodiversity Information Facility (GBIF), the European Register of Marine Species (ERMS) and the European Directory of Marine Environment Datasets (EDMED).

The collaboration of the Barcelona Convention Parties may strongly enhance the usefulness of a further development of MedGIS for the Mediterranean Region.

## A. Introduction

Many web applications have been launched in recent years which display natural resources in form of maps, including basic map tools, such as zoom in and out, switching on and off different map layers. The objective of the MedGIS project was to go beyond the simple display of information and to show as well how spatial data and its associated attribute data can be queried. While setting up MedGIS both as GIS and Web Map Server we intended to cover as much thematic groups and data types as possible in order to demonstrate the opportunities and functionalities that such a tool can offer to the UNEP-MAP[1] user community.

MedGIS is available at the following Internet address:

**http://www.rac-spa.org/**

Quality and quantity of the data are crucial for any GIS and its suitability as research and decision making tool. The same is true for web mapping applications. Obviously, the MedGIS prototype can neither be considered complete nor fully functional as a tool yet, but we hope to provide a first insight. We describe here the information already included in the prototype version, with special emphasis on data directly derived from projects related to the Mediterranean Action Plan. We also add some reflections on data which are likely to form part of it in the future.

There is still a long way from turning data into information and information into conclusions. Anyway, the future development of MedGIS will very much depend on the contributions, the collaboration and the feedback from different members of the MAP community. As the improvement in data availability and usage will also create new management and research opportunities, it is likely that the requirements and objectives of the users will change and the 'Dynamic Atlas' linked to this GIS Database will always be dynamic in the true sense of the word.

## B. Data nature

The data can be looked at from different points of view. Technically, we distinguish between spatial data and attribute data. Most of the thematic map layers represented in MedGIS consist of spatial data files in vector format – such as the Protected Areas (point) locations, the bathymetric lines (lines) and the No-Take-Zone[2] (area). These spatial data files contain the coordinate pairs that describe the features in space. They may also be related to attribute data in an associated table. Raster images (GeoTIF), such as the seabed model or the satellite images are composed of cells or pixels. Each pixel contains a digital value but there are no associated attribute tables as for vector data files. Regarding the seabed model and the satellite images (colour compositions) in MedGIS, these pixel values are not particularly meaningful and the images are mainly used as backdrop images to provide the user with some kind of ground reference. Nevertheless, there are raster images or 'coverages' which contain digital values that do have a meaning: sea surface temperature maps,

[1] Mediterranean Action Plan http://www.unepmap.org/
[2] No-Take zone below 1000m Depth according to the General Fisheries Commission for the Mediterranean

landuse/landcover maps, digital elevation models and prediction maps, contain digital values that represent temperature, altitude, vegetation class, probability etc. This type of spatial data is not (yet) represented in MedGIS, but very likely to be there in the future, since coverages are the adequate format for representing continuous, gradually changing parameters and many biological and ecological phenomena have this characteristic.

In those cases where a great amount of attribute data for a vector layer is available, the attribute data are best stored and read from a relational database. We chose MySQL[3] – an open source software and a Relational Database Management System based on Standard Query Language (SQL). All information derived from survey projects and impact assessments is stored in this database while the spatial data file which represents the sites only contains some basic attributes on the site itself (name, area, declaration year etc.).

MedGIS includes several query methods which, according to the nature of the query, access either the relatively small attribute tables belonging to vector files or the rather complex relational tables that form the database[4].

We could as well distinguish between data which are stable in time and data subject to changes (time depending data). Most of what we consider 'reference data', such as bathymetric depth, land elevation and administrative boundaries are stable over decades. Information regarding the extension of a site declared as protected area, its name and year of declaration, the country it belongs to as well as the land and marine area composition could be regarded as data (relatively) stable in time. But when it comes to biophysical parameters and natural resources the opposite is true: Water temperature, species abundance and even species presence are subject to change in time (in question of hours, days or years). Surveys and observations are data sampling events – snapshots of the current situation - which are repeated at more or less regular time intervals in order to reflect these changes. Again, data subject to change is best stored in relational databases.

## C. Data categories

MedGIS contains data of a great thematic variety. Disregarding the reference data, we deal with data of three different categories: Natural resources, (anthropogenic) impacts & risks, and biophysical and geophysical data.

### Natural resources

From one side, there are data on species level, data on presence and abundance of species, gathered during surveys and observations. In MedGIS, the registered species are limited to some 130 species belonging to the Specially Protected Area

[3] MySQL http://www.mysql.com/
[4] MySQL may also be converted into a spatial database and replace the vector files with database tables. This will be considered in the future.

(SPA) Protocol[5] Annex 2 and Annex 3 for being especially vulnerable, endangered or of economic interest.

Data on species level were gathered either during surveys and observations carried out with RAC/SPA support or by the responsible bodies of several protected areas. At some sites additional species to the annexes were also registered (nevertheless, any other species registered are dormant in the database and may as well be made accessible in the future).

Data include additional information on the species related to a determined location, such as if a species is rare or common at that location, migrating or breeding and the importance of the site for a particular species. The parameters registered in Standard Data-Entry Forms (SDF)[6] and surveys conducted in Specially Protected Areas of Mediterranean Interest (SPAMI) were different and are not always directly comparable among each other. Anyway, the queries output both groups of data fields. MedMPA[7] project sites did not have standard forms at all, but we were able to reconstruct such a species checklists by analysing the text.

Considering that it is very likely that the inventories will be repeated in several sites and that the list of species will be amplified, special attention should be paid registering Zero and Null values in order to avoid misinterpretations. Zero (=absence confirmed) is not the same as Null (=not looked for). If new species are added to the Protocol between survey campaigns (for example, invasive species, ecological indicator species or important food-chain species) which have not be counted during previous campaigns but do appear later, could be interpreted as 'new arrivals' while it is actually not known whether they were present in the past or not. Registering Zero values for absence during the surveys and Null values (database default) permits us to distinguish both cases.

There is also data on landscape composition and habitats, generally based on detailed habitat classifications (a total of 92 classes in 4 specification levels), often accompanied by area estimations (percentages at local and national levels), and sometimes accompanied by maps and spatial data files describing the location and extension of these habitats.

Additionally, data on natural resources are derived from quite diverse ecological units, such as wetlands, coastal and marine environments. This involves different expert groups, different gathering methods and different ways to store, analyse and display the information.

The Standard Data-Entry Forms (number of sites included: 6) distinguish between 'marine', 'coastal' and 'other' habitats[8]. The list is based on the Classification of

---

[5] SPA Protocol http://www.rac-spa.org/annex2.htm and http://www.rac-spa.org/annex3.htm

[6] Standard Data-Entry Form http://161.111.161.171/MedGIS/documents/SDF_ENG.pdf

[7] MedMPA: Regional Project of the Development of Marine and Coastal Protected Areas in the Mediterranean Region http://www.medmpa.net/

[8] http://161.111.161.171/MedGIS/documents/other/Guide-marine_Habitat-FSD.pdf and http://161.111.161.171/MedGIS/documents/other/Ref_list_coastal_habitat_eng.pdf

Benthic Marine Habitat Types for the Mediterranean Region[9]. The category which was completed at most sites was ‘marine habitat’, while ‘coastal habitat’ and ‘other habitat’ were often left out. Some SPAMI inventories registered marine and coastal habitats according to the Habitats Directive (Annex I) of the European Commission which distinguishes fewer classes, especially among the marine habitats, while other SPAMI inventories already applied MAP categories, which coincide with those used in SDF. The available MedMPA project inventories available for MedGIS pilot project did not contain listings on habitats at all, although they included habitat descriptions. There are currently efforts made to ‘translate’ the Habitats Directive classes into the new classification system.

Anthropogenic impacts and risks

The survey data on impacts and risks focus on human presence and activities (agriculture, fishery, leisure, industry, pollution) but also include natural hazards, such as erosion, forest fires, inundations, volcanism and earthquakes. The registers generally include the type, often accompanied by an estimation of its magnitude and origin (inside or outside the Protected Area), and a percentage of area affected.

The Standard Data Entry Forms offer a total of 177 classes using 3 levels of specification. Both, habitat and impact surveys allow queries from rather generalised to much specialised classifications. I.e. the category ‘Fishing, hunting and collecting’ includes 8 subcategories, of which 3 are further subdivided. Queries selecting a high level class include all lower subdivisions. Despite of having only 6 sites available for proper habitat and impact queries we included this option for demonstration purposes. Efforts are made to homogenise inventories by using standardized lists and classes and by facilitating thesaurus to pick the adequate keywords in order to fill in the forms correctly.

Those impacts which are of human origin (as well as some of the impacts of natural origin) directly depend on political and socio-economical pressures, related to environmental awareness on one hand and industrial development of the region or country on the other hand. One may also add to this category, all restrictions due to protection and local or national management plans, ports, marine traffic, fish exploitation, oil platforms etc. This information could possibly be provided by other UNEP-MAP Activity Centres.

Biophysical and geophysical data

This a third category which should be considered important due to its influence on all other phenomena: Biophysical and geophysical data include: air and water temperature, wind and current direction and speed, salinity, chlorophyll content and composition, land and water pollution, biomass, depth and elevation, soil type etc. Except for bathymetric depth, the MedGIS prototype does not yet include data of this category. Hence, given the increasing importance of Internet as information source and the opportunities emerging networks provide, there are good chances to retrieve

---

[9] as elaborated by the Meeting of Experts on Marine Habitat Types in the Mediterranean Region (Hyères, France, 18-20 November 1998) and subsequently reviewed by the Fourth Meeting of National Focal Points for Specially Protected Areas (Tunis, 12-14 April 1999)

these thematic maps and to display and query them simultaneously with MedGIS data. Especially promising are remote sensing services which provide satellite images and their derivates (landcover maps, vegetation indices, surface temperature, pollution, climate parameters etc.).

**D. Data spatial precision**

We may also look at the data from a spatial (geographic feature) point of view: So far, most of the data included in MedGIS are data at point locations (even those which aren't really points, such as Specially Protected Areas). Species presence, habitat type or land elevation, are usually taken at geographic localities, identified by geographic coordinates or place names. (Note that as for human activities and impacts, this 'point' may be a lot more fuzzy). In the case of the MedGIS survey data, this spatial reference is not directly given. Instead, each parameter has been registered to the Protected Area itself and is thus defined by the centroid (or central point) or the site. This has several advantages and disadvantages: the low spatial precision of the data (several kilometres) allows us to freely reveal the presence of all species, regardless of their vulnerability. On the other hand, the low spatial precision makes it difficult to relate the presence of a species to other parameters (such as water depth or temperature) which in turn is necessary for habitat suitability analysis and generating distribution (prediction) maps. Such low precision is also not good enough for management and emergency plans.

Wherever the spatial extension (polygon) of a site was available, we calculated the Maximum Error Distance, estimated as maximum distance from centroid to border following the MaNIS Georeferencing Guidelines[10]. This has been done for 7 of the Spanish SPAMI. The Maximum Error Distance is an indicator for coordinate precision and is usually requested by networking search engines (such as GBIF[11] and OBIS[12]), although it is not mandatory.

Apart from delimitating the extensions of the remaining survey sites, the next task would be to recover and include the original spatial references of the inventories wherever possible. According to the characteristics of the 'subject' and the methods applied, the localisations will vary a lot in spatial precision: For example, whale sightings made at hundreds of meters distance from a boat are less precise than localisations by radiotracking of a tagged whale, which again is less precise than taking the coordinates of a stranded whale using a differential GPS. Amount of data, certainty of taxon determination, counts of individuals and positional precision depend on various factors such as the characteristics of the species itself, its population density, the survey effort and methods applied, among others. It is not adequate to represent all these localisations by points without providing the user with additional information on how this data was obtained. As for maps on species distribution or densities, it is common to display the counted or estimated number of individuals (or species or any other biodiversity indices) in form of regular spatial units, such as grids or map graticules. The spatial precision of the observation

[10] MaNIS Georeferencing Guidelines by John Wieczorek, Univ. of California, Berkeley http://elib.cs.berkeley.edu/manis/GeorefGuide.html
[11] GBIF: Global Biodiversity Information Facility http://www.gbif.org
[12] OBIS: Ocean Biodiversity Information System http://www.iobis.org/Welcome.htm

determines the minimum size of the mesh to which the data may be assigned. There are also cases in which it may not be desired to display data at its original (high) precision. For example, still unpublished data as well as sensitive data may be deliberately represented at a lower positional precision. This is often a better alternative than not showing sensitive or unpublished data at all.

Special attention deserve primary data and synthesised information. While primary data, derived by counting, measuring or observing phenomena, are the basis for research on the subject itself, synthesized information is based on the research results, derived from statistical analysis, prediction models, data interpolation and extrapolation. This synthesised information is then often itself input for new analysis on other subjects in the same or related research fields. In MedGIS, the map layer 'Areas suspect to hold high biodiversity levels' is such an example (see: The Mediterranean Marine Gap Analysis Paper [13]). Naturally, the whole Mediterranean Sea cannot be surveyed in order to identify all biodiversity hot spots. Instead a gap analysis[14] has been conducted aiming at the identification of those areas, which according to certain criteria (spatial heterogeneity of the shelf-slope system), have an elevated probability to hold high biodiversity levels. Such synthesised information allows policy makers to better take nature conservation into account for decision making. Zoning categories and delimitation of areas of high conservation values is another example for synthesised information. However, these data can only be visualised but not yet queried in the MedGIS prototype. We disregarded as well sampling locations and related raw data for queries in the current version, since this may be only of interest to few users. As a tool for decision makers and management plans, this type of information is probably the most important. Certain priority should be given to this type of information in order to provide conclusions and knowledge rather than not interpreted raw data.

What all these data have in common is a spatial relation and an effect on the natural resources in the Mediterranean Region. It is also noteworthy that most non-European countries are already represented in the prototype. But, while MedGIS at the present stage contains mainly data gathered inside Specially Protected Areas, it is quite obvious that these sites cannot be regarded as environmentally isolated from the rest of the Mediterranean Region.

Many of the potential users are involved in research and/or planning and are interested in explaining the past and the presence or predicting the future, modelling different scenarios in order to develop plans which allow achieving a sustainable use of the Mediterranean region and its rich natural resources. They are simultaneously potential users and contributors to such an Information System. Thus, it would be convenient to determine which are the necessities of the wider MAP frame community, what information is already available or could be made available and which the gaps to be filled are, both on biological issues and related fields.

[13] http://161.111.161.171/MedGIS/documents/other/GapAnalysis_Paper.pdf
[14] http://www.gap.uidaho.edu/about/what_is_gap_analysis.htm

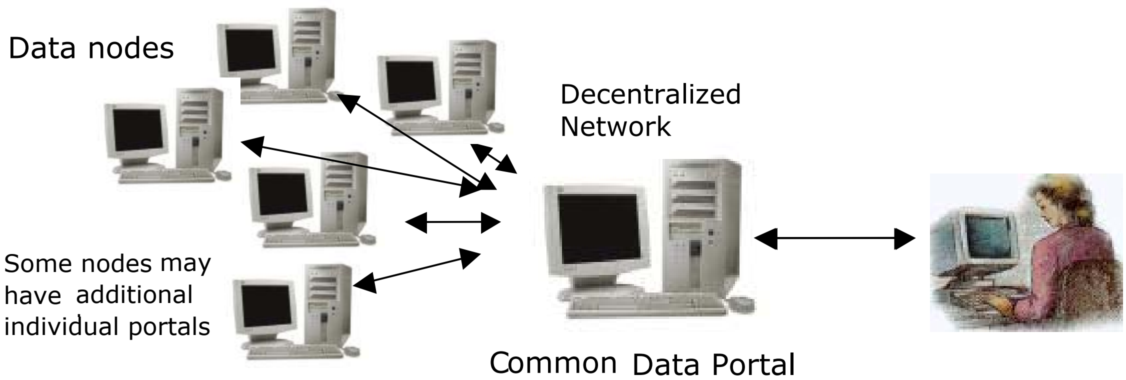
**E. Metadata and their importance**

Metadata are data about the data. They allow a user to find adequate data for his/her purpose and to find out who did the survey, who processed the satellite image, which methods have been used, year, scale…etc. Incomplete or missing metadata are a great problem. Tracing back the core metadata years after a project has been closed is a very hard work. Additionally, without metadata, proper credit can not be given to the data contributors.

**F. Possible further developments**

It may appear somewhat risky to combine sets of so diverse data themes. An adequate integration requires certainly a very high degree of coordination between the different centres and working groups and homogenisation of data gathering and exchange. Great efforts are already underway to coordinate and achieve synergy. But, this enormous information diversity might call for a subdivision of MedGIS. The MedGIS user interface may become soon overloaded and too complex for fast and effective use, especially when the addressed users are also quite diverse. One might argue that it is more reasonable to focus on data quality and quantity rather than data variety.

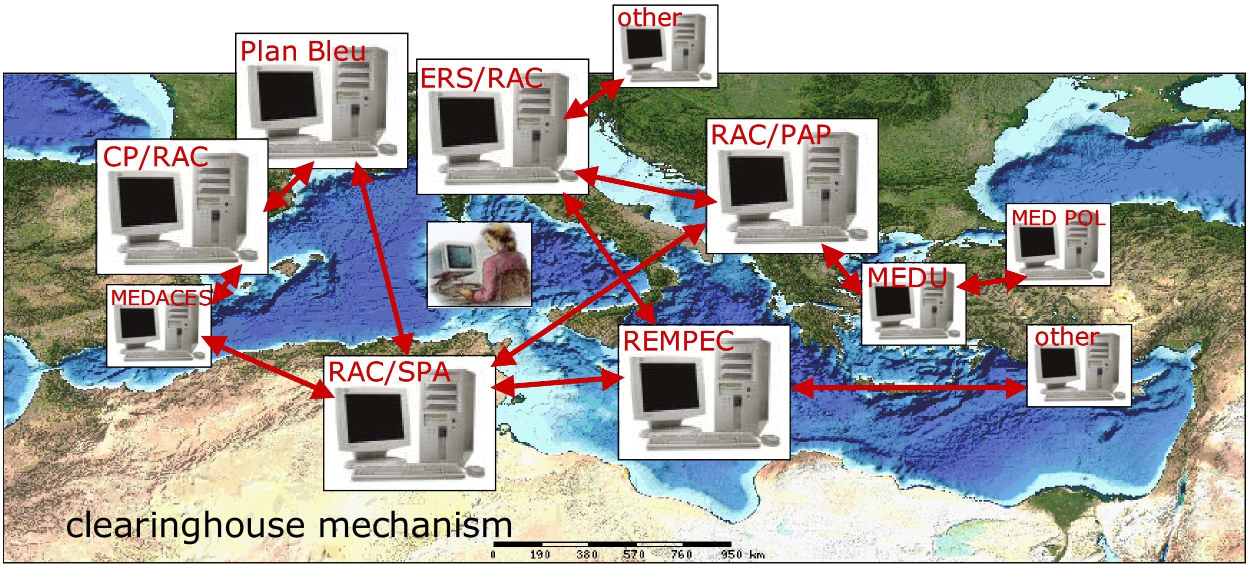
Most international networks, specialise on few data themes, have one common data access portal and many data nodes which are connected to this portal, providing standardised data. The main advantages of a distributed network are: the control and responsibility remain at the providing data nodes and data maintenance and updates are easier. At the same time, the user does not need to visit several sites in order to obtain the desired information, since all data are simultaneously available through the common portal.




On the other hand, a service which combines all factors relevant for protection and conservation of habitats and biodiversity is indeed very important. For researches and decision makers, it does not suffice to know which species are present at a determined site. They also need to know what affects them in one way or the other. More precisely, we need to be able to combine the knowledge of experts of different disciplines. The possibility to integrate teams of various disciplines into one network, able to act and react in real-time, is a great challenge and may well be worth such a coordination effort. But still, the problem not to loose the overview about such a huge variety of data themes remains.

A practical approach would be to evolve into a network formed by several MAP Regional Activity Centres on which data are shared but yet have various portals instead of one that tries to suit them all. According to their expertise, each of these portals would serve different data sets, and some of the data are obtained from other centres. Each portal would thus simultaneously be server (own data) and client (external data).

This network type would preserve all of the advantages of the distributed system with one common portal, except that the user has to choose –according to his/her interests- which portal to visit, since the available information will not be identical. Additionally, it permits that each centre presents to a maximum their own data and expertises, hence counting on contributions form other centres of the UNEP-MAP frame. The services not necessarily have to be designed and financed exclusively for and by UNEP-MAP, nor are the members obliged to evolve in concordance. They may simultaneously satisfy other national, European or global projects and initiatives as long as all implement international (OGC[15]) standards.




To get an idea of how such a network would look like, we may imagine the following case: REMPEC holds an important data base on oil spills and other toxic accidents in the Mediterranean Sea. This database contains detailed information on ship accidents, including the chemical compounds, the cause of the accident, measurements taken, the flag of the ship, who came to aid etc. The experts at REMPEC require as well some information on sensitive areas: They need immediate access to maps which represent SPAs and find out about the shore line structure and the littoral habitats potentially affected by any accident in order to make adequate decisions and interventions. The required data may actually be included and queryable in MedGIS and may be retrieved via Internet. Thus, instead of trying to fit everything into one portal interface and then let the user struggle his way through a complex set of options and selection lists, REMPEC would maintain its own portal and additionally integrate just the necessary external data on protected areas. Other nodes on the network may wish as well to draw the locations of oil spill accidents on top of their thematic maps. But they are probably mainly interested in those accidents

[15] Open Geospatial Consortium http://www.opengeospatial.org/

which have happened recently. Or, they might need rather generic risk assessment maps or models predicting the probability of occurrence or the movement of pollutants, rather than to know the exact cause of the accident or the flag of the ship.

Any centre with expertise on a particular field (for example: oil spill monitoring or satellite image processing and vegetation mapping, coastline changes, socio-economic indices and human activities …) may provide other nodes with high quality and reliable, homogenised information, avoiding thus loss in time and resources as well as data duplication. The benefit of data sharing would be an amplified one, since part of the data could be provided via Internet, automatically and always actualised. Ideally, each node would contribute with synthesised information (rather than huge amounts of primary data) to a better understanding of the marine and coastal environment.

There are also many working groups – often specialised on a particular species or taxonomic group (cetaceans, sea turtles, seaweed etc.) which could significantly contribute to and benefit from a network. Those groups belong to local or international NGOs, university departments, Framework-projects of the European Commission etc., gathering data on dolphin sightings, algae blooms, bird census and similar. Some may not be prepared yet to set up their own service but yet wish that their data were hosted by another centre which passively maintains their database server, while the responsible authors still keep control over their data. Being contributor gives these centres also international recognition. Each Centre would maintain full control and responsibility on their data and thus data quality and actualisation would be considerably improved. Besides, data and metadata losses during data transfer can be reduced or avoided. As a side effect, it will assure a better data registration, dissemination, data storage and update since responsibilities are assigned thematically to one or few centres.

Perhaps the main advantage of such a network is to bring the different expert groups with their survey and research results closer together. An improvement in data availability and usage creates new management and research opportunities. In other words - the whole is more than the sum of the parts! The achievements of all of them would increase suitability and visibility – both within the MAP frame and to other user communities and the public.

Many may be unaccustomed to share and exchange data, knowledge, resources and technology, standards and guidelines, etc. especially via Internet. But design of the data portal, software and operating system do not matter when it comes to sewing the different nodes together. Open-source software makes it easy for developing countries and low budged NGO's to catch up and step out with their results. Hence it requires that the members get together regularly in order to present the data they could possibly serve to the network and to specify their needs on external information.

MedGIS also seeks to become interoperable and meet the standards of initiatives such as GBIF, OBIS and NatureGIS[16] and other networks related to the Clearing

[16] European thematic network for Protected Areas/Nature Preservation and Geographical Information http://www.gisig.it/nature-gis/

House Mechanism of the CBD[17]. GBIF, which focuses on data on species level, offers coordination, financial and technical support, so that already existing data can be adapted and made available to a broad user community. UNEP has been a member of GBIF as associated organisation since 2001.

Another important association is OBIS, which in turn is also member of GBIF. OBIS is focussing on marine environment conservation, an area which coincides with great parts of MedGIS. OBIS' FishBase[18] is perhaps the most important database on fish and certainly the biggest of it's kind that is already online. MedGIS could benefit as well as contribute to these global networks without loosing its own character.

These and other initiatives and international projects are already developing metadata standards and software to facilitate metadata recording. GBIF deals mainly with data on species level and would not be suitable for data on habitats or impacts or human activities. OBIS may be more adequate for most data derived in marine environments. Since these initiatives have all the same objective – make their data available on Internet – great effort is undertaken to achieve interoperability and their standards are based on the same ISO norms. They then may or may not add additional optional parameters to improve suitability within their network. The OBIS schema for metadata is a relatively simple data model to represent taxon occurrence records and is very similar to the GBIF DarwinCore[19] model (in fact, both are compatible and can be browsed by the same communication protocol. It is perhaps a good starting point for a future development in the UNEP-MAP framework. Another interesting metadata profile, especially for protected areas, is the NatureGIS metadata profile[20] .

The participation of the Parties in the development of MedGIS at national level would strongly augment the usefulness of this CHM tool.

---

17 Convention on Biological Diversity. Convention text: http://www.biodiv.org/convention/articles.asp
18 Global Information System on Fishes http://www.fishbase.org/home.htm
19 DarwinCore v.2: http://tsadev.speciesanalyst.net/documentation/ow.asp?DarwinCoreV2
20 NatureGIS metadata: http://www.naturegis.net/metadata/naturegisprofile.htm