Lessons Learned – Organising and managing multiple and complex environmental datasets

The challenge

The area of Prespa, like any natural landscape, is a living system, which is constantly evolving as a result of numerous processes taking place in the immediate and wider area, related to both natural and anthropogenic parameters.

Since 1991, the SPP has systematically and effectively contributed to the protection of the natural and cultural heritage throughout the Prespa basin, and beyond, with a wide programme of conservation activities that also promote the harmonious coexistence of people and nature. Its work aims to achieve the sound management of natural resources, sustainable development and conservation of species and habitats under threat, as well as raising public awareness about conservation and biodiversity issues.

In order to realise its goal and vision, for over 30 years the SPP has collected, and needs to manage, a substantial amount of data related to the abiotic and biotic characteristics of the wider area, which are parameters and indicators relating to the area's species and habitats, the life and well-being of people, and the development of the landscape. The SPP also collects data on threats and pressures, which may disturb the balance and good functioning of the ecologically important area of Prespa, as well as spatial and cartographic data on the measurements, observations and management actions that comprise its work.

The data stored in the database managed by the SPP include meteorological and climate data, physicochemical and biological parameters of the waters of the area and data on the water level of the Lesser and Great Prespa lakes, as well as corresponding elements of the transboundary basin, altimetric and bathymetric data. Also included are fauna and flora monitoring data, spatial data derived from SPP management practices, such as areas of reed cutting, grazing, creation of fire zones, etc., cartographic material for the area, geospatial data on the natural and anthropogenic environment, environmental and administrative information and topographical data, amongst others. Lastly, evidence of human activities linked to threats and pressures on wildlife is sought, stored and analysed.

Some of the aforementioned data comes from monitoring data measuring stations in the area. This data is updated at regular intervals, as well as in exceptional circumstances when necessary. Other data are recorded by the SPP staff after detailed planning of measurements, observations or actions, at which point the database is updated with the new information. Finally, the database is supplemented at regular intervals with data obtained from public services, either through direct correspondence or through searches on the websites of organisations such as the Ministry of Environment and Energy, Hellenic Land Registry, Energy Regulatory Authority and the Army Geographical Service, and others.

All these many parameters, the variety of the characteristics being measured and the long time-series make the process of gathering, organising, updating, storing, managing and ultimately analysing these data a challenging and laborious undertaking.

The solution

The key to successfully managing large volumes of data is to maintain a comprehensive, properly organised database. This means that it must include all the measured data, be properly structured and prioritized, and include distinct collections of related data stored electronically or digitally using the appropriate software that allows us to analyse the data in each field of knowledge.

In all cases, care should be taken to organise, update and manage the data well, to avoid incorrect entries, to ensure the continuity of the time series and to allow further processing and analysis of the data held in the database.

Lessons learned

The most important conclusions drawn from the SPP's experience in long-term data recording and organisation are noted below.

- The first step in successfully implementing a database is to identify and assess the identity, characteristics and type of information to be collected, as well as the purpose for which it is being collected. This allows the choice of appropriate software and a set of applications that will serve the purpose for which they are intended.
- The database should be organised in such a way that it has a logical arrangement of the different data elements, ensuring that, regardless of the position occupied by particular data in the database, they can be related to other data with the help of an index. This will help ensure that each time information is extracted it will have a specific meaningful content and value for its intended recipient. In addition, logical organisation of the data allows for easier oversight, thus reducing the possibility of repetitions or any discontinuities. Furthermore, a well-formed structure makes it easy to insert new records, and to search for and retrieve subsets of data, through queries, for further processing. Lastly, a well-organised database also facilitates the extraction of reports and presentation of the results of data processing.
- Record fields should have a defined data type, be it text, date or time, number or some other type, in order to avoid errors and inconsistencies and make data processing easier. A typical example is the definition of a field as a date field, from which further analysis can be done at month level, etc. Or defining a field as a number field based on which operations such as addition, numerical ordering, finding the minimum / maximum recorded value, etc. can easily be done. If by mistake data that does not correspond to the selected predefined format (e.g. text character) is inputted it will be easily noticed, since it will not be possible to further process the data.
- Another important factor is the use of macros, with which most database operations can be performed automatically, thus saving working time. Macros group and automate database tasks. You can set criteria or restrictions and run queries. For example, how many threats are detected at a given distance from the protected object, how much land was managed in an area range in the previous year etc.
- The basic guideline for the organisation of a database should be its ease of use, by selecting appropriate software with easy-to-use and a simple data entry environment and interface between the user and the system, so that time-consuming procedures are not required to manage and process its contents and so as to give flexibility in both usage and modification.
- The data should be regularly examined to identify failures or errors in the records.
- Finally, a properly planned update of the database is necessary, in order to identify and correct problems related to the operation of the measuring stations. If, for example, the operation of an automatic measuring station (e.g. a meteorological station) stops or is disrupted, the regular and planned updating of the data from that station will allow us to detect the failure in its operation in time and avoid the loss of large amounts of data.