Proceedings of FRIEND-Water 2014, Montpellier, France, October 2014 (IAHS Publ. 363, 2014).

Architecture of environmental information systems applied to scientific observatories: examples of Carnoulès and MEDYCYSS observatories

J. F. BOYER[1], E. SERVAT[1], C. CASIOT[2] & H. JOURDE[3]

1 IRD, 2 CNRS, 3 UM2, HydroSciences Laboratory, Place Eugène Bataillon, 34095 Montpellier, France

boyer@msem.univ-montp2.fr

Abstract According to the missions of the French OSUs (Observatoires des Sciences de l'Univers), an observatory must provide, on a given territory, a permanent monitoring activity. This mission requires strict organization linked to a relevant monitoring network, a high level of data quality and data availability. The requirements are provided by an Environmental Information System (EIS) such as those implemented for two Observation Systems which are part of OREME OSU: one dedicated to hydro-biogeochemical processes concerning metals and metalloids in surface waters downstream from the Carnoulès mine, the other, dedicated to Multi Scale Observatory of flood dynamics and hydrodynamics in Karst. A specific architecture was chosen to develop an EIS that identifies and provides information mainly around three functions: (i) data centralization in a client/server database following the standards for metadata description, (ii) safe management of scientific information, and (iii) a geo-referenced photo album representing territory, measurement stations, and instrumentation devices used.

Key words observatory; environmental information system; normalized metadata; GIS; document management; database OSU OREME; MEDYCYSS; Carnoulès

INTRODUCTION

French observatories (OSU) are structured around research, training and mainly continuous observation of the environment over long periods, with an obligation to supply long-lasting and certified quality data. This last obligation involves the use of an Environmental Information System (EIS) which is an appropriate tool to operate as requested. Geosciences, and in particular the environmental observatories on water, do not require a real time technology acquisition and data processing except for flood or quality monitoring networks. Nevertheless, they need a specific structure to guarantee data traceability and quality.

STATEMENT OF THE PROBLEM

Essentially EISs organize work between devices and scientists. Devices are expressed in terms of gauging stations, sensors, software, transmission, networks and methodologies, but also as quantity and flow of data to be processed and database size. Data processing has to deal with architectures of operating systems, programming languages, software components and the algorithms used.

The design and architecture have a cost of acquisition, maintenance and operation, and should be compared to the benefits of the EIS: it sometimes leads to reducing the cost of projects. EIS success requires combining easy use and maintenance with requirements in terms of availability and safety of use.

Sensors and monitoring network

An observatory is first a network of environmental sensors organized all over a territory. Measured environmental variables may include hydro-climatic ones, physical and chemical parameters regarding water, groundwater, soil, snow or other physical parameters. A specific organization is coupled to the network to route all information to a reference point. It can be some staff reading measurements on the spot, but also, for example, e.g. an automatic transmitter network via GSM GPRS. The reference point is usually a central database. The time step required for data acquisition, the distance and the price of the sensors determine the choice of communication technology. The real-time transmission is used for predictive modelling. An observatory is often built for a combination of various goals.

Data repository and metadata

The first systems were built around data files organized in a system of directories and subdirectories. This architecture is still used in the case of a deposit of static, heterogeneous and small size data. More generally, the deposit of reference is a centralized relational database.

With time, most observatories' databases grow in size and complexity and have to deal with more and more users. Then, the database has to migrate to a client/server RDBMS architecture to optimize performance and application, as well as their adaptability, availability, safety, reliability and ability to recover.

An EIS must take into account not only the values of the observations, but also their Metadata. Metadata is the descriptive information that explains the measurement attributes (e.g. their names, units, precision, accuracy, as well as the lineage describing how the data was measured, acquired, or computed). It is more particularly (i) identification information of the resource including general information (title, description, reference dates, version, summary, stakeholders), description of the geographic extent of the resource, information about possible uses, legal and security constraints attached to the resource, (ii) spatial representation information, (iii) quality information (usual criteria of quality measurement: geometric and time details, logical consistency, genealogy information, description of the sources and processes applied to sources), and (iv) content description, identification of coordinate reference systems, modalities of presentation, dissemination and maintenance procedures.

Data quality control

Sensor data must pass through a set of procedures of quality control before it is used. This process represents a lot of effort and is very time consuming. It relies on specific tools to control required quality. In many cases, the data is first validated by some staff specialized in metrology before being put into the information system.

Data publication and interoperability

Environmental observatories are means of cooperation. The information they contain should be published in formats allowing this collaboration. The INSPIRE European directive (2007/2/EC of 14 March 2007) establishes a set of services available for sharing geographic data and metadata. For this purpose, the ISO 19115 standard is used as a metadata standard for geographic objects. In the case of environmental observatories, all information can be managed as geographic objects. In fact, any measure (hydrological, climatic, physic-chemical etc.) corresponds to a localized geographic point and to a geographical object location: a station (one point), a river (polyline), a watershed or aquifer (polygon), etc.

Moreover, environmental observatories must not only share data, but must also consider the semantic differences (differences in language and terminology used by data producers to describe their observations) and syntax (differences in data description and coding). Many international initiatives exist in the scientific community pushing for the use of languages called "self descriptive" using a standard markup such as eXtensibleMarkupLanguage(XML) (Barry *et al*, 2013). Mention may be made for example of Water Markup Language (WaterML) (Taylor, 2012) and Open Geospatial Consortium (OGC) Observations and Measurements (Cox, 2011).

EXAMPLES OF OBSERVATORIES CARNOULES AND MEDYCYSS (OREME OSU)

OREME OSU Presentation

OREME OSU is an Observatory of Sciences of the Universe recognized by French CNRS and the University Montpellier South of France (UMSF). OSU has three main missions: research, training and natural environment observation.

Based on the training components of UMSF for its educational mission, OREME mobilizes means of observation and scientific expertise on physical, chemical and biological Mediterranean aspects. OREME make links between the scientific communities – ecology, biodiversity, earth sciences – so that they work together with different and complementary EISs. One of the challenges of OREME is to collect, integrate and share heterogeneous data and highlight correlations.

Carnoulès observatory

Mining of metallic sulfide ores leads to the generation of extensive amounts of sulfide-rich waste materials which are generally stored next to the site of ore extraction. These wastes are sensitive to alteration when exposed to water infiltration. They can release acid waters containing high concentrations of iron and sulfate for several centuries. These waters, called Acid Mine Drainage (AMD), also contain high concentrations of metals and metalloids both toxic for biota and humans.

The laboratory HydroSciences (HSM) has undertaken hydrological, geochemical and microbiological observations at the former Pb-Zn Carnoulès mine in southern France. The mine is located on the Cevennes subwatershed of the Rhône River. It operated until 1962 and left 1.5 M tons of wastes remnant from the recovery process that contain sulfides, metals and arsenic. At the base of the tailings dam emerges the source of Reigous Creek which has acid waters (pH 2–4) with extremely high Fe (0.5–1 g/L) and As (100–350 mg/L) content. The Reigous Creek joins the River Amous 1500 m downstream from its source. Downstream from the confluence Reigous-Amous, the quality of Amous River water is greatly degraded. The Observatory of the Carnoulès mine aims to monitor the physico-chemical characteristics of the waters, metals and arsenic concentrations, and the associated communities of microorganisms.

MEDYCYSS observatory

In recent years, an increase in flood damage in the Mediterranean region was observed. In these, the majority of rivers subject to extreme floods flow in part on karst and fractured rocks, typical around the Mediterranean. On these sites, hydrodynamic and hydrochemical variations in different parts of the system (wells, karst system, temporary and permanent springs), and the flow of springs and rivers, are monitored continuously. Figure 1 shows all the observations and experimental sites grouped in the MEDYCYSS Observatory, including three major river systems in interaction with karst: the Lez, Mosson and Coulazou.

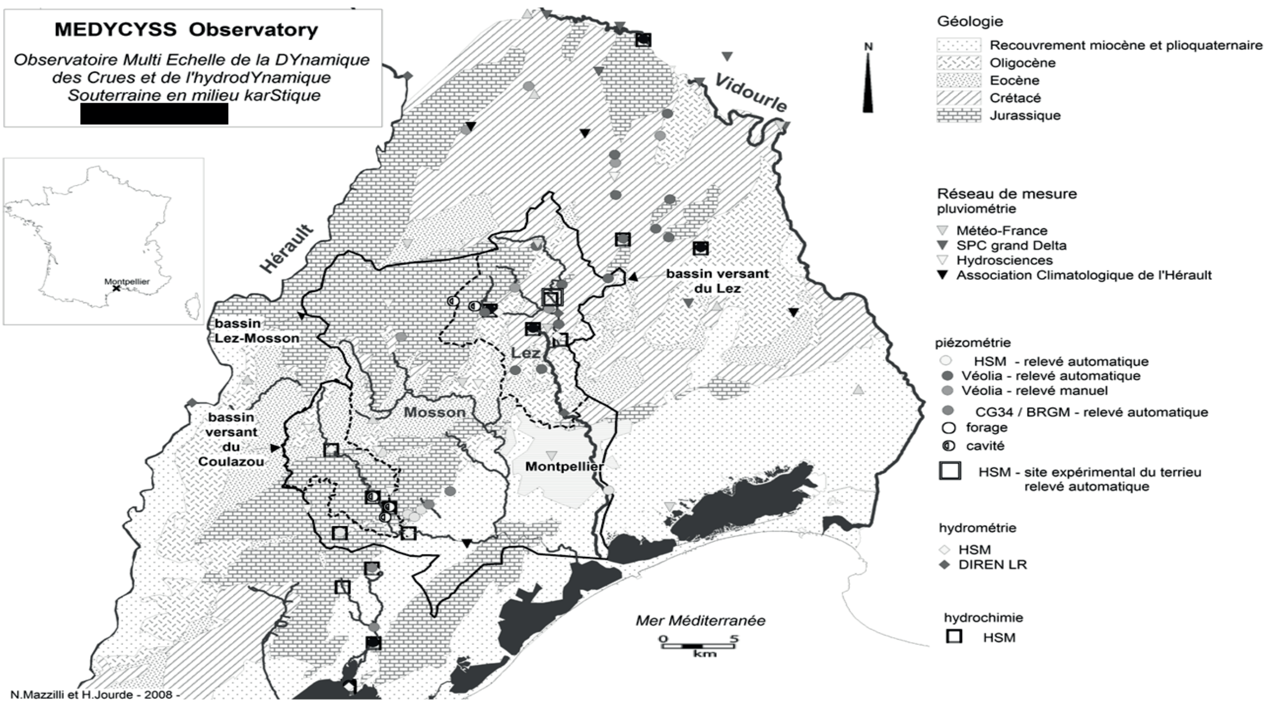

Fig. 1 Location of Lez, Coulazou and Mosson catchments, parts of MEDYCYSS observatory.

Monitoring network of both observatories

Both observatories have stations installed and have data collected by the coordination team of these sites. There is also the benefit of measurements made by other actors in the region, e.g. mainly Meteofrance, Veolia and BRG,M. There are nearly 700 stations and points of observations available across the two observatories (see Table 1) measuring more than 70 types of variable (see Table 2).

Table 1 List of stations for both observatories.

Station types	Number of stations managed by HSM	Total number of stations
Piezometric	36	130
Hydrometric	7	37
Rainfall	9	187
Hydrochemical	37	337
Soil humidity	6	6

Table 2 List of measured variables for both observatories.

Variable	Variable	Variable	Variable
Rain	Uranium	CO3 Carbonate	Reduced Arsenic As(III)
Water level	Zinc	Copper	Stand. Dev. of Bore isotopic ratio
Conductivity	Redox potential	Dysprosium	87sr/86sr stand. deviation
Piezometer	Gadolinium	Total Coliforms	87sr/86sr strontium isotopic ratio
Natural Flow	Sulfur	Potassium	Suspended Particulate Matter
Mean Temperature	Tin	Lithium	TOC Total Organic Carbon
Atmopheric pressure	Thorium	Manganese	Bore isotopic ratio
Dry temperature	Thallium	Sodium	Strontium
Boron	Yitttrium	NO3 Nitrate	O2 Dissolved oxygen
Bromide	pH	Rubidium	Total Dissolved Solids
Cadmium	oxygen 18	SO4 Sulphate	Molybdenum
Cerium	Carbon 13	Iron	Nickel
Cobalt	Deuterium	Vanadium	Lead
Chromium	Aluminium	Silver	Titanium
Cesium	Arsenic	Reduced iron Fe(II)	Tungsten
Fecal Coliforms	Oxided Arsenic As(V)	Phosphorus	Zirconium
HCO3 Bicarbonate	Barium	Sb Antimony	Magnesium
Lanthanum	Calcium	water temperature	Chloride

For the Carnoulès observatory, analyses focus on the main physico-chemical parameters of water (pH, conductivity, dissolved salts concentration, temperature), the concentrations of iron, sulfate, arsenic, lead and zinc, as well as redox forms of iron and arsenic. Additional analyses (redox potential, dissolved oxygen concentration, concentration of suspended solids, concentrations of Li, B, Al, Ca, V, Cr, Mn, Co, Ni, Cu, Rb, Sr, Mo, Cd, Sb, Cs, Ba, Tl, U) are occasionally measured. Analysis of the diversity of microorganisms (bacteria, archaea and eukaryotes) present in water and sediment is made retrospectively based on physical chemistry of water and running projects. The rainfall data commenced in 2006 as well as piezometric variation, drilled into the stock pile. Flows to the Source Reigous are also measured. However, they are affected by a significant uncertainty associated with the aggressiveness of the environment on measuring instruments.

For MEDYCYSS observatory, the hydrological and hydrogeological Lez basin is a remarkable site because of its vulnerability to the risk of pollution and flooding and its contribution to the Montpellier water supply. The existence of continuous pumping and high volume (6000 m^3 h^{-1}) at Lez spring allows investigation at regional scale of the hydrodynamic characteristics of the karst hydro system (400 km^2, 40 wells), and locally the Terrieu experimental site (2500 m^2, 22 wells).

There is a concern about the water resources change as a result of climatic and anthropogenic forcing, with the Lez aquifer used as a reference location with (i) measurements of the water level of Lez spring since 1946, and (ii) flow measurements implemented since 1974. In addition, the piezometric monitoring network includes 22 regional drill holes, recorded on a monthly time step since 1981, and, for 14 of them, continuously since 2000. The piezometric Terrieu experimental site had been monitored on 21 wells since the 1970s.

Database and metadata definition for both observatories

The developed database model is a model based on a planned Entity/relation which now is known and used by numerous developments of hydro-climatological bases (FRIEND AOC, FRIEND AMHY, HYCOS AOC, and SIEREM). This plan is based on three separate groups of information: (1) meta-data descriptions, (2) time series of organized data, and (3) specific management of relevant scientific literature information (Fig. 2).

The first group is organized around a main table storing information in accordance with the mandatory set of describing data of the ISO 19115 standard and describing the measurement stations as a point on the Earth. Thus, information concerns the location in latitude, longitude and altitude of the station location, the country and the type of installed station. At this point, there are also described and attached information on: photos, hydrographic watershed and river for the hydrometric stations. One can develop this plan by adding as many tables as needed to define other meta-data, e.g. it is possible to add a table containing parameters of dam description if this information is available and interesting to be managed by the system.

The second group of tables allows storage of specifically hydro-meteorological data. These data are linked to a record describing the time series to which they belong. Time series are described in type, time step and source of data by a system of tables so that any new type of chronological series can be described and measurements stored without any transformation of the base plan.

The third group of tables is dedicated to store scientific documents such as scientific papers, reports and PhD documents, in connection with spots and data. This information can be linked to stations or time series to attach the set of data which contribute to the overall dossier. Data access security is implemented and maintained at the database diagram level.

Users must be identified, registered in the database and organized into groups. A group is a set of users with the same rights regarding the same data set. A separate login and password is assigned to each user. Each time series and each document have its own security. There can be different levels for various kinds of users. The first level establishes a free access to information, the second allows an unauthorized user to access the metadata, but not the data, and the third level protects both metadata and data to unauthorized users. For each time series and each of the documents, one of the three levels of protection is defined. For a given time series, the three levels of access can sometimes be defined according to the considered period.

The metadata for these two observatories are defined according to the ISO 19115 standard. This standard defines a core of mandatory information to describe data to meet general needs. But it must be declined by the different users communities to respond more specifically to their needs and requests. These adaptations of the standard are called “Profile”. Within these two observatories, the WMO Core Metadata Profile was used. The WMO profile and its relative standard, ISO 19115, provide a way to extend the use of the data beyond the communities that traditionally use them. It also allows users to access data in different ways depending on criteria such as elements of the profile.

A specific intranet administration allows all authorized persons to update, add and delete all types of information maintained in the IES. This interface, made of a set of internet forms, requires administrators to meet the standards of metadata (no validation if lack of information) and exempts knowledge of the physical structure of the database to be able to work on it. As the administration is made through the website itself, it also enables collaborative and decentralized administration.

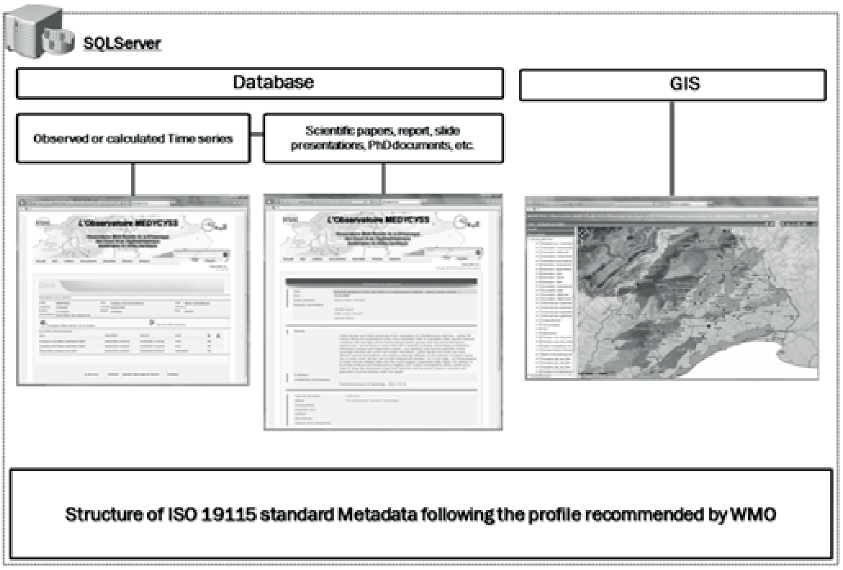

Fig. 2 General principles of Observatory information system.

Data publication and exchange for both observatories

As for the data administration, the data publication and exchange is operated through the web site. Creating a website is an imperative for visibility, management and control of observatory internal and external communication. This site will allow the project management in terms of animation and it is the interface for any user to access information: the data (time series, GIS layers, technical and scientific documents etc.) and communication information in general (adverts, events, news). The web site provides a secure access to the data and provides to the administrator a special interface to manage the information of the EIS. A special intranet part of the web site is dedicated to this administration allowing the administrator to update, add, remove data and obliges him to follow the ISO 19115 norm for the metadata.

CONCLUSIONS

One of the primary missions of environmental observatories is the collection, management and publication of the data. This mission requires adopting principles of organization and information formatting under constraints linked with a need for interoperability between observatories. The elaboration of the network, the strict definition of the structured database in coordination with modeling rules and the adoption of a metadata standard widely accepted by the community are the minimum set of principles to be implemented in such a working framework.

The examples above demonstrate the application of these principles in the construction of a simple EIS dedicated to a scientific community rather than to the general public. The application of these principles has allowed creating these EISs that centralize all kinds of data (from semi-automatic field measurements to results of laboratory analyses) with a common definition of metadata and common procedures for data publication and distribution.

Acknowledgements The authors gratefully acknowledge the efforts of the OREME-OSU committees for developing the vision of the Mediterranean Observatories Network.

REFERENCES

Barry, D. K. & Dick, D. (2013). Web Services, Service-Oriented Architectures, and Cloud Computing: The Savvy Manager's Guide. *Morgan Kaufmann,* edn 2, 248 pp.

Bailly Comte, V., Jourde, H., Roesch, A., Pistre, S. & Batiot Guilhe, C. (2008). Time series analyses for karst/river interactions assessment: Case of the Coulazou River (southern France), *Journal of Hydrol.* 349, 98–114.

Cox, S. (2011). Observations and Measurements - XML Implementation, Open Geospatial Consortium OGC 10-025r1, 77 pp.

Taylor, P. (2012). OGC WaterML 2.0: Part 1 – Timeseries, *Open Geospatial Consortium: OGC-10-126r3,* 137 pp.