International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 7, Issue 1, January - February 2018 ISSN 2278-6856

Lessons learnt from the design and implementation of a gazetteer for monitoring the water physicochemical parameters in Gabon: Spatiotemporal and toponymic approach

Marius MASSALA[1], Franck FAVETTA[2] and Pierre MOUKELI[3]

[1]African Institute Computer, BP 2263, Libreville, Gabon

[2]Claude Bernard University of Lyon LIRIS, Campus LyonTech-la Doua
25 avenue Pierre de Coubertin 69622 Villeurbanne Cedex (France)

[3]African Institute Computer, BP 2263, Libreville, Gabon

Abstract: ***The objective of this paper is to propose a methodology aimed at setting up a gazetteer dedicated to the monitoring of the water physicochemical parameters in Gabon. The project, initiated by the National Center for Oceanographic Data and Information (CNDIO), uses a network of mobile and fixed sensors spread over the Gabonese territory. In this article, a study of foreshadowing has been initiated and the publication of which constitutes the first point of advancement. Two problems are specifically mentioned: the need to create an observation network in the light of the needs of managers and researchers, as well as the potential missions of such a directory in connection with university or government projects initiated for this purpose.***

Keywords: Physicochemical, gazetteer, spatiotemporal, Gabon

1. INTRODUCTION

Facing the emergence of new information and communications technologies, spatialized information is ubiquitous in computer applications, either implicitly or explicitly [23]. Today, various applications and solutions are developed covering broad and varied fields. The environmental domain is not on the margins of this process. At a time when environmental problems are acutely global (as evidenced by the 21st and 22nd Conferences of the Parties 'COP' on Climate Change), it seems appropriate that each part of society should make its contribution to the preservation of the common heritage of planet Earth.

In the area of Geographic Information Systems (GIS) applied to environmental problems, the development of tools for representing spatiotemporal data has made it possible to offer users a particularly efficient means of monitoring. Gabon, as a developing country, is not marginal to this dynamic global initiative. Several measures within the national scope were taken. We will mention, among other things, the design of a National Climate Council, the creation of a Space Observation Agency, as well as national parks, etc. It is within this framework of environmental preservation that the national program for the collection of physicochemical parameters of fresh water falls. This program, carried out by the National Center for Data and Oceanographic Information (CNDIO) of Gabon, aims to know all the parameters that influence the quality of water anywhere on the Gabonese territory.

The work described here is part of the program of the monitoring and the understanding of physicochemical parameters of waters, developed by the Department of Marine Sciences (DESMAR) of the National Center for Scientific Research (CENAREST). Indeed, the National Center for Oceanographic Data and Information (CNDIO), which depends on CENAREST, collects a large quantity of physicochemical data. Thanks to fixed or mobile sensors, several measurements are made on different sites at short intervals. These increasingly large data are stored in disparate and heterogeneous files. Actually, the CNDIO is unable to follow and even to do spatial and statistical analyzes of these data.

Our work lies within the framework of the construction of directories of toponyms. Indeed, the territories on which the sensor arrays are installed generally characterized by fuzzy boundaries. This is the case, for example, with buoys equipped with sensors. When dropped at sea or on a watercourse, these buoys may drift. Under such conditions, they can be located at successive times T_1, T_2, …,T_n, on different geographic objects whose boundaries can be defined in a fuzzy manner. The problem also arises for fixed sensors when they are located at the mouths of streams / rivers or at the border of two or three territories.

Facing such problems, solutions such as catalogs, observatories or the conventional GIS tools developed under the ArcGIS, Map info, etc. platforms are ineffective. Indeed, the databases resulting from such solutions generally answer questions concerning geographic features whose limits are clearly defined. That is why we think that toponyms seem to be a potential trail for such problems.

The objective of this work is to propose the first reflections which should lead to the implementation of a spatiotemporal solution based on toponyms. The gazetteer that we intend to develop must take into account the fuzzy characteristics of the features in the territory, the dynamic character of the sensor networks as well as the nominal variants in the features designation. The aim is to facilitate the sharing of knowledge and know-how between different research organizations. It is within this framework that this article is written.

This article is organized as follows. In the first part, we present the context in which these data were collected as well as their nature. Then, we will describe the scientific problems that arise in the context of the CNDIO. After identifying the problems, we set out in section three a state of the art in this field. The forth section describes the reasons why a solution based on the use of toponyms is recommended. The fifth section of this work presents the proposal of a methodology applicable to the exploitation of these data in the context of Gabon. Finally, we conclude by presenting the missions of the gazetteer through a few cases of use as well as examples of requests.

2. CONTEXT AND PROBLEM

With the increasing development occurring in Africa, the continent is facing major challenges in providing quality water to people. The problem of accessing drinking water is even more important in peri-urban areas, due to the lack of regulation of the uncontrolled and informal settlement of populations, and also the absence of municipal water services. This absence of public services in those area has as corollary, the use of rainwater, wells, and watercourses surrounding dwellings. It is within this framework that the program for monitoring and understanding the water physicochemical parameters developed by the CNDIO takes place.

2.1 Description of the measuring stations and collected parameters

The water observation and monitoring network set up by the CNDIO is made up of several sensor stations distributed along the littoral and inside the Gabonese territory. Surveillance and data collection campaigns are carried out at irregular intervals. The objective is to verify the consumable nature of the waters, but also to understand the dynamics of the muddy cork. Similarly, these campaigns make it possible to highlight the processes controlling the oxygenation of water and to feed the database which will make it possible, in comparison with future data, to perfectly parameterize the evolution of the estuary in relation to climate change.

Two categories of sensors have been used: mobile sensors and fixed sensors. Today, only the fixed sensors are deployed. Each station measures several physicochemical parameters thanks to dedicated electronic devices. Among the measured parameters, let us mention temperature, salinity, suspended solids content, dissolved oxygen concentration, PH, turbidity, water potential, redox potential, etc. In general, the measurement frequency is calculated on average ten minutes from samples taken one meter below the surface. In the long term, the CNDIO plans to set up drifting mobile sensor stations on rivers. Whatever the sensors used all the information collected is recorded in an Excel spreadsheet.

2.2. Problems

In recent years, a large amount of data has been collected and recorded thanks to the network of sensors distributed on various sites in Gabon. Data collected at irregular frequencies are stored in scattered and heterogeneous Excel files. However, such a storage procedure makes it difficult to monitor, cross-check and analyze these data. As [2] point out, "integrating information available on a territory requires an unavoidable phase of inventory of shareable resources". Therefore, after analyzing the documents available to us, we have identified the following problems:

- Lack of standardization of data from one collection campaign to another;
- Lack of identification information (no identifier, no label of the sensor, not even the reference of the station...) which allow the characterization of the data;
- Excel format table structure not suitable for storing spatial information;
- Ambiguity when collection sites are located in places with identical names. For example, it is difficult to distinguish samples taken from the Ngounié River when it crosses Ngounié province and water points in the same province;
- Absence of a particular statistical analysis of the data and a dynamic follow-up of the evolution of these data. One consequence is that scientists and users cannot get an idea of the quality of the water on the situation of the various sites. This type of information is very useful in the implementation of *e*-administration [7], because more and more *e*-services are using geo-referenced data. For example water quality can impact ecological tourism.

In view of the abovementioned problems, a methodology should be proposed for better monitoring of these data. The establishment of a gazetteer-type solution of typonyms seems to be a potential path. Indeed, the use of several place names as a sensor station, the continuing capture of measures, the lack of standardization in collection procedures, are some of the shortcomings we have identified.

Proposing a solution that gives scientists and decision-makers the ability to place spatiotemporal and toponymic requests is a major challenge. Indeed, as pointed out by [6], a solution of this nature is a way of rethinking the modes of articulation and intervention of the various actors on one hand, and on the other hand, it is an answer to the need of animation, information and communication, based on common reflections organized.

3. STATE OF THE ART

In recent years, the integration of the spatial-temporal dimension in geoscience applications has been the subject of several research studies. As noted by [5], the problem of historical monitoring of environmental phenomena and natural hazards uses spatial and temporal multidimensional data. Several scientific studies have been proposed in various fields (volcanic monitoring, marine pollution, monitoring of air pollutants, management of household waste in urban centers, etc.). In historical risk monitoring, for example, [5] propose a technique of modeling of events (avalanche, flood, etc.). In their proposal, an event has a spatial representation. To characterize the evolutions of phenomena, the temporal dimension has been managed from two points of view: the existence of objects (creation / destruction) and the evolutionary aspect of objects during their lifetime and their change of state.

For their part, [18] have worked on understanding the mechanisms of groundwater circulation. They propose a modeling of the risks of pollution generated by the anthropic activities on the Loyalty Islands of New Caledonia. From field observations, fluorimetric tracing and piezometry [18], a model was constructed in which it was now possible to reconstruct the trajectories followed by the waters and their speed in their flow. In general, consideration of the spatiotemporal dimension has been the subject of several studies leading to numerous applications. We can cite for example the work of [4], [20], [8] or those of [1], and so on.

Whatever the nature of the work carried out, their applications did not integrate problems related to the fuzzy nature of the component elements of the territory (rivers, mangroves, etc.). Similarly, these authors did not take into account the dynamic nature of a sensor station. Indeed, in the case of a station of mobile sensors (such as buoys in a stream), they may belong to several geographic locations because of their trajectory. From the above, it seems that the proposed solutions appear inoperative.

4. TRACKS OF GAZETTEER AND OBJECTIVES

For a long time geoscience applications were based on the use of catalogs, atlases and observatories. Today a new generation is taking over: that of gazetteers or directories of toponyms. Indeed, toponyms are now used as a central pillar in the modeling phase of this new generation of tools. As a reminder, a gazetteer is a dictionary of remarkable elements of the territory that combines both the place names and their description [3]. This description may include relationships with other entities (objects or phenomena, for example, an island is within the waters) or may take the form of a position (latitude and longitude), or geometric form (line or Polygon) or in the form of attributive descriptors that can be assimilated to key words [16]. The construction of geographic indices may comprise several elements that form the territory (continent, country, city, mountain, river, archaeological site, etc.); and their descriptions are variable (geographical position, depth, area, population, photograph of the place, etc.) [11], [12], [3].

Figure 1. Complex relationship between a geographic location and its names [14].

Several recent scientific works have now included the names of toponyms as complex databases. The complexity of their uses leads progressively to develop ontologies of toponyms [10], [14]. This approach, which consists in developing directory-type solutions, is recent and is increasingly experimented in new geoscience applications [16]. Today, several projects dealing with the gazetteers exist throughout the world and numerous scientific works have been carried out. This is the case for example in Brazil [22], Croatia [10], Tunisia [20], France [3], [25], and China [13]. The use of geographic toponyms directories opens a new era in the development of solutions to assist decision-makers in information management and environmental / geographic knowledge in the field of geoscience [17].

Out of all the works cited above, none of them have yet dealt with problems related to fixed and mobile sensor networks coupling both the fuzzy dimensions within the geographic objects and toponyms. This is the logic behind this work. The National Center for Oceanographic Data and Information (CNDIO) is facing a lack of tools adapted to its context. The objective of the propose solution is to help the researchers in their analysis, monitoring and even forecasting work. Therefore, it seemed appropriate to bring to this research organization, an innovative solution through which actors can observe, simulate, analyze, map, etc., the physicochemical parameters measured on different sites.

As [9] point out, environmental monitoring involves several observation stations. The latter are responsible for monitoring various phenomena to which one or more parameters correspond. Since these sampling stations are located in different places (regions, cities, villages, neighborhoods, etc.), the spatial and toponym dimension is therefore an essential element. All of the data collected by the CNDIO also follow this logic. In the remainder of this article, we outline our proposal. Section 5 presents the main methodological principles (models, use cases, etc.). Then we propose the architecture of the solution followed by some examples of queries.

5. CLASS DIAGRAM AND GAZETTEER ROLE

As pointed out by [6], proposing a model that represents the territory, follows from several points of view: natural territory, planning and administrative divisions. This approach, which takes into account all the aforementioned aspects, is called cognitive [23]. The modeling of a territorial system is therefore a function of the objectives we set ourselves from the beginning. That is why, in owncontext, we have identified at each level the essential elements that must be taken into account. Our proposal is based on a model that supports the implementation of a gazetteer dedicated to this type of problems. Here, the territory is composed of a set of entities interacting with each other: these represent the geographic features. The integration of the data takes into account the modeling of the spatio-temporal aspects of these geographic objects. The same applies to all the physicochemical data collected by each collection station. The UML formalism was used to model our territory and the set of elements that gravitate around geographical objects. The following section presents the conceptual framework of the model followed by a use case.

5.1. Conceptual Model

As noted in [2], the use of an object formalism will allow the interactions between concepts to be managed through the relations between classes. In the model given Figure 2, we have defined generic classes taking into account the spatio-temporal dimensions. The model revolves around the *Toponym, Geographic Objects, Linear_station and Static_station classes.*

- The *Geographic object* class is a generic class. It has a general character on all the territorial spaces on which the stations for collecting physicochemical parameters will be located. Geographic objects are identified in three forms: polygons (where all the segments are known, such as provinces, departments, towns and villages, and lakes), lines (all kind of water courses except lakes), and finally points (in which we will have fixed stations and certain sites whose geometry is so small that we will assimilate them to punctual objects). Surface objects are characterized by their fixed and fuzzy boundaries. The fuzzy aspect has been taken into account due to the current configurations of the major world agglomerations. These large cities are generally made up of the small municipalities joining them. Thus, in the collective consciousness, these small towns are generally considered to be part of the great metropolises. This is the case of Villeurbanne and Lyon and Paris and Nanterre in France. African cities are not on the margins of this pattern. In Gabon, for example, we have the cities of Libreville (capital of the country), Owendo and Akanda (small surrounding municipalities). The definition of a fuzzy boundary can be used to solve the problems that may arise when searching for information. An illustration of this problem is represented in the diagram below.

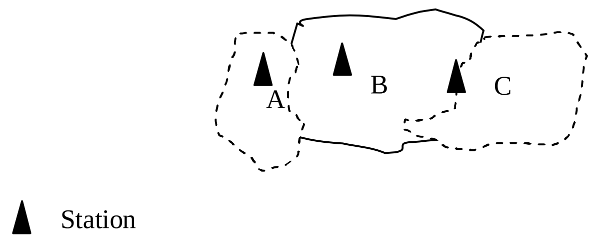

Figure 3. Illustration of a situation where the fuzzy boundary of an object (B) is the union of objects (A, B and C).

During the collecting campaigns, it was found that a station which is located, for example, in the municipality of Owendo at the river "Nomba" can be counted as part of Libreville since the municipality of Owendo is part of greater Libreville. The same applies to measurements taken north of Libreville, in the municipality of Akanda. In view of the above, we can define the fuzzy boundaries of Libreville as the union of all the municipalities adjacent to it.

- The *toponym* class will describe the set of names that geographic objects will carry. During the implementation, a table will be created which allows to collect all the names of the places with their synonyms. Linear and surface objects were identified by their names. The sites and stations that populate the geographic objects are identified by the different names of the geographic feature on which they are located. For example, if a site is located in Owendo, it will carry not only the Owendo but also the Libreville toponym since the fuzzy boundary of Libreville encompasses Owendo.

From the above, it follows that a site or a station can belong to several places at the same time. Such an approach makes it possible to solve the problems posed by the drifting stations which may belong to instants given to several rivers and even to several geographic objects of surface types. Figure 4 illustrates such a problem.

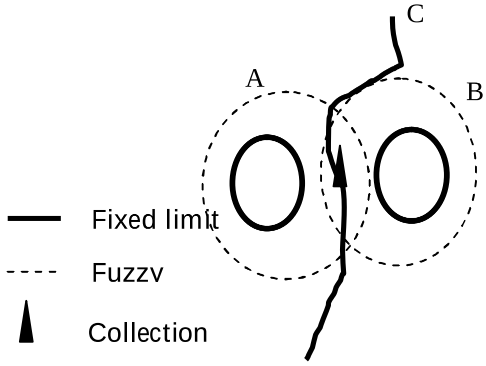

Figure 4. Example of a fixed station located on a watercourse (C) crossing two geographic objects (A and B) with fuzzy boundaries

- The *Linearstation* and *staticstation* classes represent the stations of the mobile and fixed sensors. They are composed of a set of dedicated sensors. These stations make recordings of the physicochemical parameters at a

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 7, Issue 1, January - February 2018 **ISSN 2278-6856**

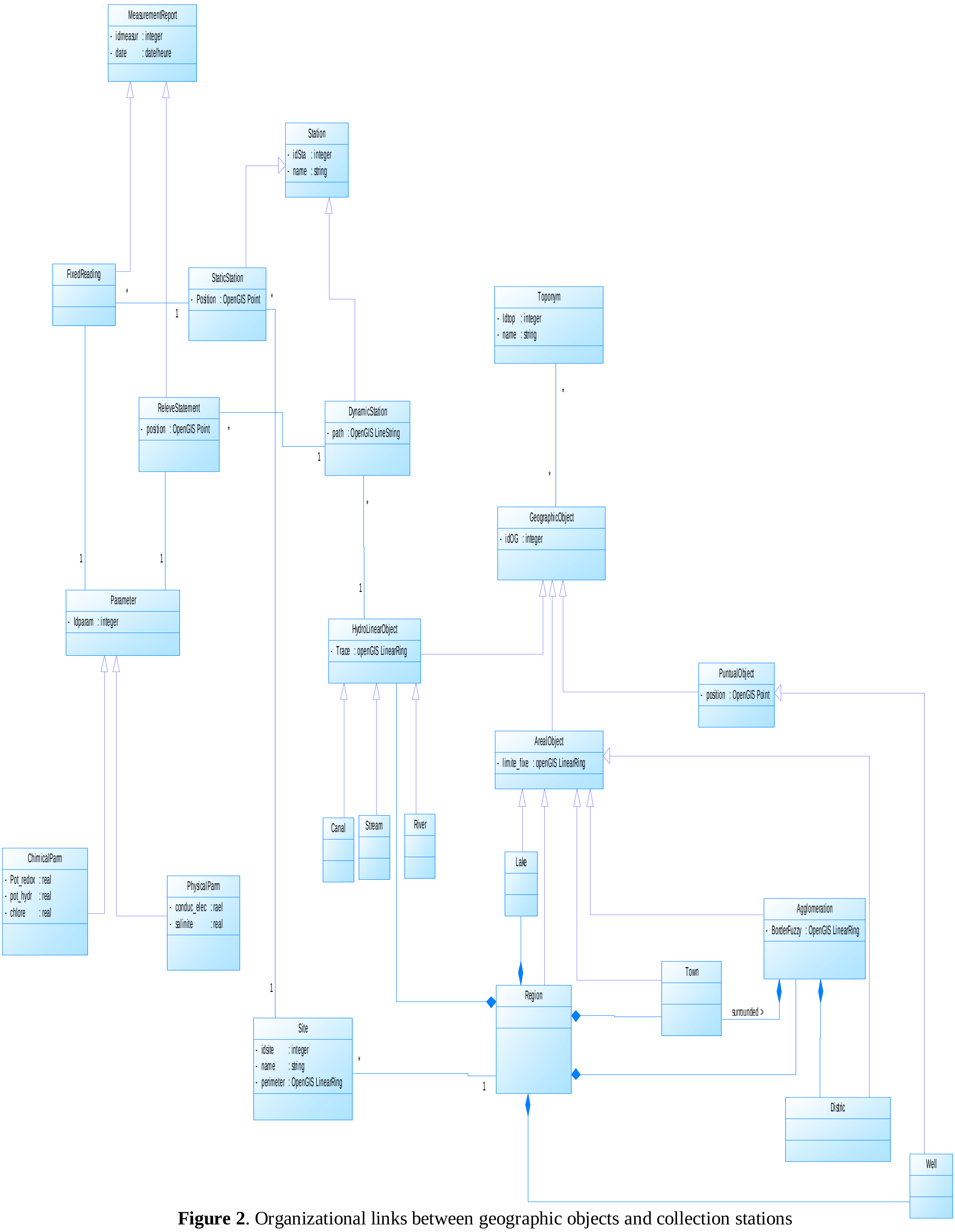

Figure 2. Organizational links between geographic objects and collection stations

- given date and according to a quantum of time previously defined on sites. A site is a point geographic location, located on a hydrographic entity (rivers, lakes, canals). To properly locate the sites, the geographic position through the GPS coordinates is important. In the model, only fixed sensors make recordings on sites. This is why we have created a direct link to the site and static_station classes. However, for mobile sensors, the geographic position is dynamic as a function of time. From this state of fact, it follows that a mobile station discretizes a dynamic continuous process. As a result, the sensors of these mobile stations measure a dimension of this process. If we call *X* this process, X_{ij} represents a record for one of the sensors of a station at a given instant.
We will consider *i*, as the number of sensors, and *j* is the recording time. Thus, the process *X* can be represented in the form of a matrix. With online recordings at different times and in columns, the different types of sensors of the station. With the rows, as the different recordings at successive times, and columns as the different types of sensors of the station.

$$X = \begin{pmatrix} X_{11} & X_{12} & X_{13} \ldots\ldots & \ldots\ldots X_{1n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ X_{m1} & X_{m2} & X_{m3} \ldots\ldots & \ldots X_{mn} \end{pmatrix}$$

As for the quantum of time, we denote as *T*. And between 2 records, we will say that $T_{n+1} = T_n + T$.

- The *parameter* class is a generic class that inherits the "physical" and "chemical" classes. These two classes make it possible to collect all the physical and chemical data collected on the sites for each station For the sake of simplicity, we did not want to make appear in the model all the properties of these two classes, because they are very numerous. In the parameter class, we also introduced the temporal dimension by inserting the "*date*" attribute. Indeed, the collections are realized according to sequences well determined at a given period.
After this brief presentation of the class diagram, we illustrate through a case of use the different functionalities that can be encountered in the future solution.

5.2. Directory roles and sample queries

The gazetteer we are proposing will include a set of concepts and functionalities. We describe through use cases (Figure 5). In consultation with all the actors of the CNDIO, we identified the different shareholders of the community who could be the future users of the solution. These are the administrator, the researchers, the data collecting engineers and the users. Depending on the user's status and the access rights associated with it, the proposed documents may differ.

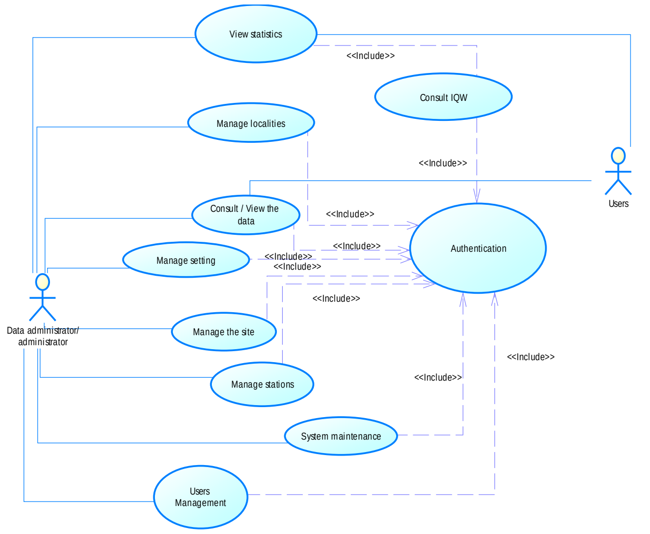

Figure 5. Use cases to the interaction between the different levels of users and some of the functionalities of the system.

The use case outlines the various possible situations that users may encounter: administrator, researchers, collector and users. Regarding the role of the administrator, it will manage the users, maintain all the system parameters that will allow the solution to function properly. For his part, the researcher manages the functionalities relating to the stations, the sites, and some parameters relating to the various analyzes and indicators that are carried out of data. Among these indicators, we will have, for example, the Water Quality Index (WQI). This index will make it possible to determine, through color codes, the quality of the water at the collection sites. The same applies to the statistical calculations performed on the data. The latter will ultimately determine whether water is conducive to bathing, drinking, protecting aquatic life or simply protecting the body of water from anthropization. A researcher can be authenticated as a collector and also capture the data collected in the field. The last role is that of any user. He can consult on the website all the information relating to the water quality in a specific region.

Currently, the solution is under development. It is a web solution in which we will find the gazetteer and the different functionalities that we have been presented above. The database is implemented under the PostgreSQL / GIS Databases Management System (DBMS). The search for information in the solution will have to take into account a certain level of complexity: for example, the system must be able to answer the following query:

“I am looking for the collecting stations that are located in Libreville (places) based on the river" Guégué "(rivers) which produced the data on date t1 (date) and which have a quality index of the water suitable for swimming “.

6. CONCLUSION AND PERSPECTIVES

In this paper, we have presented the procedure leading to the implementation of gazetteer. The purpose of this project is to meet the needs of the CNDIO in terms of monitoring the water physicochemical parameters in Gabon. The developed directory integrates the spatio-temporal and toponym dimensions. We proposed a conceptual model that takes into account the following points:

- Insertion of a fuzzy boundary in the description of geographic objects;
- The proposal of a dictionary of names of places and their synonyms for all of Gabon. This gazetteer will describe not only the geographic objects but also help the solution when searching for information;
- Implementation of extensive spatial analysis capabilities dedicated to monitoring water quality;
- The integration of biological parameters when the stations are equipped with the technologies to collect the related data.

This work is currently being finalized, notably through the development of the tool web accessible to the general public as an e-service [8], and the deployment of the solution. As [7] point out, beyond this particular example, the reflection that we have carried out in this publication constitutes a real methodology. It can be for our part usable in other contexts. This is why, in the implementation, we can still identify several points that need to be further explored.

References

[1] Barde J: Pooling of data and knowledge for integrated coastal zone management, Application to the SYSCOLAG project. PhD thesis in Computer science, University of Montpellier II, 2005, 287p.

[2] Bedel O, 2011. The Gazetteer as a tool for access to information, Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 335.

[3] Kayleen, E., Guitton, J., Libourel, T., 2014. Collaborative geomatics Atlas for the environment and sustainable management of fisheries resources in West Africa, Mauritania case. Proceedings of the Colloquium INFORSID 2014, Lyon

[4] Bogdan M, Davoine JP, Gensel J, Martin H: GenGHIS. A spatio-temporal modeling tool for the historical monitoring of natural hazards. Information Systems Engineering Review, 2005, volume 10, pp.35-58.

[5] Camara A., Passouant M., Cheylan J. P: Conception of an observatory for Territorial Development, International Journal of Geomatics, 2010, Volume 20, Number 3, pp. 331-361

[6] Desconnets J C, Libourel T, CLERC s: Catalogue to disseminate environmental resources. Proceedings of the Conference INFORSID, 2007, Perros-Guirec, pp. 344-361.

[7] Gant Jon P.: “Electronic Government for Developing Countries”; International Telecommunication Union; report; August 2008; https://www.itu.int/ITU-D/cyb/app/docs/e-gov_for_dev_countries-report.pdf

[8] Gutierrez C., Sankar S. Metadata Temporal Temps-Reel. Journal of Information Science and technology: Engineering of Information Systems, Volume 12, Number 2, 2007, pp. 97-119.

[9] Hecimovi Z: Standardization of the geographical names in Croatia. Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 334-334

[10] Jack P: between wars and conflicts: the planet under tension, international Festival of Geography, Saint die-des-Vosges, SIG the letter, Number 10, 2008.

[11] Janowicz K., Schwarz M., Wilkes m: implementation and Evaluation of a Semantics-based User Interface for Web gazetteers, Visual Interfaces to the Social and the Semantic Web (VISSW) Workshop in conjunction with the International Conference on Intelligent user Interfaces, Sanibel Island, Florida, Volume 443, 2009

[12] JI W: Chinese national Geographical Names data base management system. Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 336-337

[13] Mangeas Morgan, Largouet Christine, Touraivane, Lille Didier: The EAUNET system: an information system integrating a dynamical modelling of pollution hazards to support the fresh water management. International Journal of Geomatics, 2005, Volume 10, Number 4, pp. 11-34

[14] Massala M., Paegelow M., 2014, the contribution of toponymic directories in the dissemination of geographical information: the example of Central African countries. In: Faiz S. and Turki S.Y. (eds.), Geographic Information in Africa, state of the places and new approaches. University Publishing Centre, La Manouba, Tunisia; P. 127-156

[15] Massala M: Intelligent Search for geographical information from toponyms, metadata and an ontology: Application to the forests of the Congo basin. PhD thesis in Landscaping Geography, University of Toulouse II Le Mirail, 2013, 213p.

[16] Monk A: The territory as a complex system: an operative concept for landscaping and geography. Geographic Space Magazine, vol 2, Volume 35, pp. 115-132, 2006

[17] Nasr M., Devilz M., Masmoudy Y., Chauhan W: Towards a database of geographical names in Tunisia: Case of the topographic sheet of Sfax number 107. Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 377-385

[18] Salas Hugo Alatrista: Extraction of spatio-temporal relationships from environmental data and health. PhD thesis, University of Montpellier II – Sciences and Techniques of Languedoc, 2013, 161p.

[19] Santos C.J.B. D: A new approach to the collecting of geographic names in Brazil. Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 313

[20] Sede M-H., Monk A, systems and territorial databases, concepts and tools for rational management of Territories, GIS and Development, International Journal of Geomatics, Volume 11, Number. 3-4, 2001, p. 333-358

[21] Servigne Sylvie., Thérèse Libourel. Spatio-temporal information systems. Volume 10 (RSI serial ISI), pp.136, 2005, 2-7462-1267-6.

[22] Syszkowska: National Register of geographical Names as part of modern, consistent and reference spatial data infrastructure of Poland. Act of the 25th International Cartographic Conference, Paris, 03-08 July 2011, pp. 337-343