

Current trends in the management of groundwater specific geospatial information

Codrina Maria Ilie^{1,*}, and Radu Constantin Gogu¹

¹Technical University of Civil Engineering Bucharest, Groundwater Engineering Research Center, Bucharest, Romania

Abstract. The purpose of this paper is to present the state-of-art of groundwater geospatial information management, highlighting the relevant data model characteristics and technical implementation of the European Directive 2007/2/EC, also known as the INSPIRE Directive. The maturity of the groundwater geodata management systems is of crucial importance for any kind of activity, be it a research project or an operational service of monitoring, protection or exploitation activities. An ineffective and inadequate geodata management system can significantly increase costs or even overthrow the entire activity ([1-3]). Furthermore, following the technological advancement and the extended scientific and operational interdisciplinary connectivity at national and international scale, the interoperability characteristics are becoming increasingly important in the development of groundwater geospatial information management. From paper recordings to digital spreadsheets, from relational database to standardized data models, the manner in which the groundwater data was gathered, stored, processed and visualized has changed significantly over time. Aside from the clear technical progress, the design that captures the natural connections and dependencies between each groundwater feature and phenomena have also evolved. The second part of our paper address the variations that occurred when outlining the different groundwater geospatial information management models, differences that depict the complexity of hydrogeological data.

1 Introduction

The present paper proposes an analysis of the current forms of management and storage for geospatial information used in hydrogeology. The research has a main focus on investigating the relevant data model characteristics and technical implementation guidelines of the European Directive 2007/2/EC, particularly the Hydrogeology Data Scheme - which together with the Geology Data Scheme and Geophysics Data Scheme form the Geology Theme.

Within the paper, a data model is defined as a standardized description of the elements and connections between elements within a universe of discourse, with the specific considered limitations (such as limited types of attributes). The scope of developing and implementing a data model lies in the ability to digitally encapsulate the measurements, observations and calculations of the specific features and phenomena, allowing a clear understanding of that specific domain. Standardization of a data model permits a common language between all parties involved, be they part of the academia, private or

public sector. Furthermore, it allows interoperable sharing of data, be it among different entities, or different countries. The importance of adopting a platform-agnostic, complete, standardized and open data model for Earth Sciences has long been proven and as a consequence, there are data models in different stages of development for most Geosciences.

In the current international context, when proper water resources management is crucial for a sustainable development, it is essential that the groundwater research community takes advantage of the significant progress that has been achieved in Computer Science. Within the last 20 years, various initiatives have been developed following this informational progress, yet another aspect is essential to mention. A crucial breakthrough is the increased collaboration between different specialists towards a common goal. In the highly specialized scientific world of today, taking advantage of the developments in the different adjacent domains translates into highly effective results leading to progress and new research directions.

* Corresponding author: codrina.ilie@utcb.ro

2 Specificity of hydrogeological geodata

Hydrogeological data is collected to understand the physical, chemical and biological characteristics of the groundwater system. As hydrogeology is a geoscience, there are specific parameters significant in storage, processing, visualizing and querying of data:

- location;
- time.

The complexity of hydrogeological geodata management comes from its highly heterogeneous sources, from all points of view:

- science field - geography, geology, hydrology, meteorology, soil science;
- data producers/collectors - national agencies, private companies, research centres ([2]);
- acquisition type - sensors, satellite, modeling etc.

This situation invariably leads to cumbersome processes when integration in an efficient way the necessary geodata in one single system that allows a seamless storage, processing and querying. Moreover, there are concepts with two or more commonly used names, such as hydraulic head versus piezometric head. Although this might not seem of importance, when it regards naming in geodata storage, the difference can generate meaningful errors and misinterpretations.

Furthermore, natural elements do not obey administrative laws and limitations, thus, it is a common situation when data must be transferred across borders. The results are often translated into investments of excessive resources of personal, time and money into understanding, integrating and re-using the data.

Hydrogeology data is obtained through the measurements and quantifications of:

- hydrogeological structures;
- hydrogeological characteristics;
- flow factors;
- physical and chemical parameters of groundwater.

There are many ways to divide data categories, e.g. primary – secondary ([4]), general – specific ([5]). Nonetheless, worth mentioning in the context of this paper is the classification proposed in 1972 by Castany [5], that categorized hydrogeological data with consideration to what he defined as "mappable data". The author considered two main groups: rock and groundwater. As we will present in the follow section, this division has been used in defining the implementing rules of the INSPIRE Hydrogeology Data Scheme, as well.

Another consistent argument is that in a data structure, there must be a well-defined separation between raw data and processed data, such as point data, interpolated data and interpreted data ([3]). Therefore, beyond the complexity and heterogeneity of hydrogeological geodata, stands the high variability of the processing state of specific datasets.

To best visualize the complexity and the diverseness of geodata in hydrogeology, we will use the concept of hypercube-base visualization model ([6, 7]). As presented in Fig.1., there are four dimensions of the

hydrogeological geodata: (1) location, (2) time, (3) topic, (4) the level of data processing - this dimension refers to whether the data is raw data - collected through various fields campaigns, or in different stages of processing - we will name this dimension - interpretation.

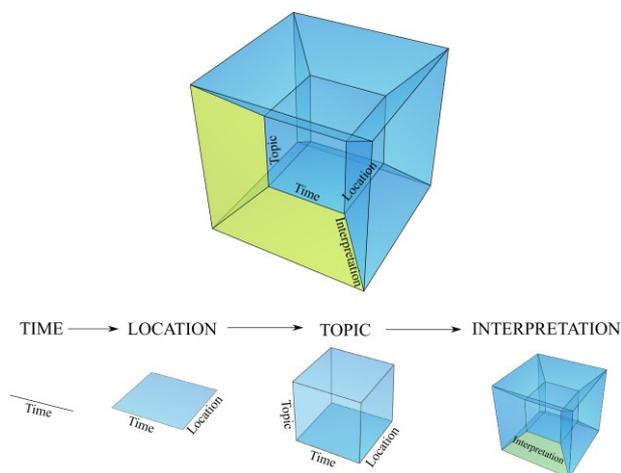


Fig.1. Hypercube-base visualization model of the 4 dimensions of hydrogeological geodata.

3 Groundwater geospatial information management

To reach the current level of development in the geodata management of the hydrogeology specific data, there have been a number of stages delineated, on one side by the intensification of collaboration among scientists in different relevant domains such as geology, hydrology, geography, engineering, geomatics and, on the other side, by the technological progress in collecting, producing, storing, processing, visualization and transferring of data. Once the transition from paper to digital occurred, the way hydrogeological data was structured followed, even though not closely, the informatics advancement, transitioning from structuring data in individual files, to construction of databases to data models.

Even though, in the current paper, the highlight is on data models, to best understand the analysed constructions, we have considered in our analysis, database structures as well. When studying each model, we separated the analysis into the conceptual and logical level of abstraction. The conceptual level refers to understanding the entities and the interactions between entities from a specific universe of discourse, while the logical level relates to data structuring, list of allowed attributes and their domain of extension, classification etc.

With respect to this paper's scope, the authors have selected 3 of the most common data models, with a special emphasis on the INSPIRE Hydrogeology Data Scheme.

- HYGES database schema (2001);
- H2O data model (2013);
- GroundwaterML data model (2017).

* Corresponding author: codrina.ilie@utcb.ro

3.1. Groundwater data models

3.1.1. Hydrogeological Database Concept, 2001

The Hydrogeological Database Concept - HYGES - was published in 2001 ([1]). The primary reason for the database was to store hydrogeological data for the Walloon Region, Belgium. The goals set out were divided between (1) processing and preparation of data for modeling and (2) for cartographic design. A specific characteristic was given by the necessity of applicability of the database in vulnerability assessments for calcareous regions. Furthermore, the database was constructed with high regard to (1) the existing data, in various formats, quality and completeness: tables, paper maps, other geodata formats and (2) to planned field campaigns. At the end of the study, the authors proposed a databased concept that encapsulated 17 groups of layers with relevant characteristics represented ([1]). Even more, within the construction of the concept, a division based on the level of processing was considered: raw data was primary, secondary was the result of processing the raw data. Crucial to the achievement of the initial scope are the two group layers: surface water point and groundwater point. With respect to the logical level, and from a geometrical point of view, the two group layers are represented as points (vector data), with a complex attribute table construction that would allow, through cardinality connections, to store information on: activity type, exact address, accessibility, tests performed, time series, sample information, construction characteristics (in case of) etc. Considering the type of technical solution used in the HYGES database design, the common element that extends throughout the entire construction is the field that stores the unique identifier for each groundwater point.

HYGES database was built using ESRI ArcINFO technology that allowed coupling with Groundwater Modeling System (GMS) and even though, at that time, operations such as attribute transfer were not possible, the authors developed using Arc Macro Language^a an interface that supported data query and transfer. Using the tool, one could query spatial or time relevant characteristics such as flow rates or hydraulic head values and easily create a readable GMS file.

3.1.2. H₂O data model, 2013

H₂O data model has been developed at University of Liege, by prof. Piotr Wojda and prof. Serge Brouyère, as a proposal for a standardized way to store and transfer hydrogeological data, at different levels and scales ([8]). The data model is compliant with international recommendations and standards, developed by organizations such as the Open Geospatial Consortium (OGC), International Organization for Standardization Technical Committee 211 (ISO/TC211) and the

^a The ARC Macro Language (AML) is a proprietary high-level algorithmic language for generating applications in ArcInfo.

European Geospatial Information Working Group. The data model was presented to the scientific community in 2013.

At conceptual level, the H₂O data model is particularly focused on the management of field-acquired data - pumping and tracer tests - as well as on relevant measurement metadata, such as number of piezometers effectively used, tracer concentration, tracer type etc. The conceptual model of the H₂O defines the following hydrogeological entities:

- well;
- groundwater body;
- aquifer;
- geochemistry measurement;
- piezometric head level measurement;
- groundwater extraction volume;
- relationships between the above entities.

The logical model of the H₂O data model has been developed with respect to the international standards, as mentioned above. Thus, UML^b has been selected for the technical layout, while XML^c is the selected language for the geospatial data storage and exchange. The reason for this decision was given by the following delineated scopes of the data model: (1) usability in international context, as well as multi-user and multi-purpose context, (2) capacity of integration with other international standards. With respect to the development structure, the data model divides all considered classes into four different specific packages: AbstractFeatures, GroundwaterFeatures, Hydrogeology and Observations&Measurements.

The AbstractFeatures contains the abstract classes that are common for the entire model. It is a specificity for the logical structure of a Geography Markup Language particularization. The authors defined the two main abstract classes of their model: SamplingFeature and HydrogeologicFeature. The SamplingFeature is part of the Observations&Measurements ([9]) international standard and holds the specialized class of HydrogeologicSamplingFeature that encloses every geospatial element natural or man-made that gives access to groundwater, allowing observations or measurements to be made. The HydrogeologicFeature abstract class is built for all geospatial features that are not used to make observations/measurements on groundwater.

We can deduce that the authors of the H₂O data model have started building their structure from the initial division of Castany, in 1972 - rock Hydrogeology and groundwater GroundwaterFeatures. Afterwards, the development of the model closely respected the initial requirement, to encode measurements data from field

^b UML stands for Unified Modeling Language and it is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems.

^c XML stands for eXtensible Markup Language and it is a software- and hardware-independent tool for storing and transporting data.

campaigns. In the context of data models, a feature is a real-world object of interest for the specific domain.

3.1.3. GroundwaterML 2.0

GroundwaterML 2.0 represents the international standard proposed by the Open Geospatial Consortium for data exchange transfer via web in hydrogeology. The efforts invested in the development of this document are rooted in the first attempt of establishing a standard for data transfer between Canada and USA, starting with 2009. The work has been done within the Open Geospatial Consortium Hydrology Domain Working Group.

GroundwaterML 2.0 is a particularization of the Geography Markup Language - GML - version 3.2.1, especially designed for hydrogeology. In the development of GWML 2.0 other relevant standards have been used, such as the OGC Observations and Measurements and the International Union of Geological Standards GeoSciML 4 ([10]). At conceptual level, GWML 2.0 defines the following main hydrogeological features:

- hydrogeological units: these are defined as earth material containers for subsurface fluids and can be of 2 types: (1) aquifer-related units - the distinction is made by the different hydrogeological properties of the rock material and (2) groundwater basins - the distinction is made by the different flow regimes;
- fluid bodies: distinct bodies of fluid that fill the voids - it can be water or gas;
- voids: are considered to be the spaces inside the earth material or inside a hydrogeological unit (such as an aquifer) that can be filled with fluid;
- fluid flow: represents the characteristics of the process of groundwater flow;
- wells: can be water well, springs or monitoring sites.

3.2. INSPIRE - Infrastructure for Spatial Information in the European Community - Geology Theme

In 2007, the European Parliament issued the European Directive 2007/2/CE that aims to create a spatial data infrastructure within the European Union that would allow access to information relevant to protection and sustainable management of the environment within the European Union member states. Full implementation is scheduled for 2021^d. It is the only legislative act which has incorporated technical aspects, with motivation rooted into developing a completely interoperable infrastructure that would allow seamless interactions among any possible participant within the European Union or any other country that implements the INSPIRE rules, such as Switzerland or the Western

Balkan Region^e. In that sense, the directive provides common Implementing Rules, that have been adopted as European Commission Decisions or Regulations and are binding entirely.

The directive addresses 34 spatial data themes that are considered to be relevant to the environment, such as topography, meteorology, cadastre, postal addresses etc., among which hydrogeology is enclosed as well. Relevant to the current paper is the Data Specification on Geology - Technical Implementation ([11]) published in 2013, as the 5th version of the document. The document is divided into 3 components:

- Application Schema Geology;
- Application Schema Hydrogeology;
- Application Schema Geophysics.

Comparable to the division Castany defined in 1972, the INSPIRE technical team developed the data model based on the two fundamental elements that interact: rock and groundwater ([11]). The two systems and their interaction complete the hydrogeological system. The data model aims at encoding relevant information of on the two systems and create the logical connections between them.

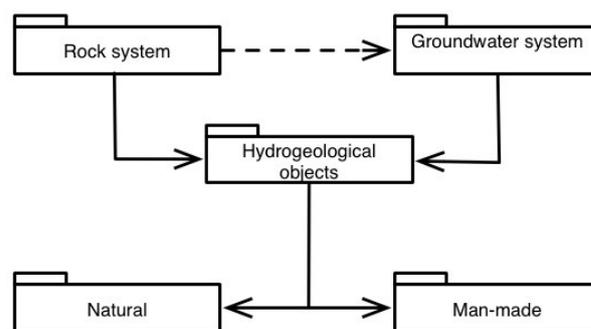


Fig.2. The INSPIRE hydrogeological system concept.

The rock system is considered invariable in time and it encodes information related to the geological structure. The data model is constructed based on a main class HydrogeologicalUnit that is a generalization of 4 subclasses: aquifer, aquitard, aquiclude and aquifer system. The hydrogeological unit is defined as a part of the lithosphere with distinctive parameters for water storage and conduction. In its own right, the HydrogeologicalUnit is a particularization of the GeologicalUnit, the main class of the Geology schema application. Thus, the connections with the Geology application schema is created. Furthermore, HydrogeologicalUnit has a logical link to the GeologicStructure, that is defined as a configuration of matter in the Earth based on describable inhomogeneity, a pattern, or fracture in an earth material. Moreover, it is independent of the material that is the substrate for the structure.

^d <http://inspire.ec.europa.eu/about-inspire/563> last accessed January 2019

^e <https://www.lantmateriet.se/sv/Om-Lantmateriet/Samverkan-med-andra/impuls/about-the-impuls-project/> last accessed December 2018

The groundwater system represents the second basic element and it is considered variable in time, with the main defined class GroundWaterBody. Groundwater body is defined as a distinct volume of water within an aquifer or system of aquifers, which is hydraulically isolated from nearby groundwater bodies ([12]). They are described as hydraulically continuous entities, defined based on flow or abstraction. Groundwater bodies are inextricably linked to surface water bodies ([12]). Through this class, the hydrogeology schema application is connected to the Management, restriction or Regulation Zone class defined in the Area Management Restriction Regulation Zones and Reporting units. The linked feature is the WFDGroundWaterBody representing the feature defined in the Water Framework Directive: "a distinct volume of groundwater within an aquifer or aquifers" ([12]).

The third core class of elements is represented by HydrogeologicalObjects. This class encodes data on any elements that comes in contact with groundwater, be it natural, such as a spring or man-made, such as a well. The classification is clear: HydrogeologicalObjectManMade and HydrogeologicalObjectNatural, with the mention that the first represents a generalization of the class ActiveWell. The class creates connections with two other data models within INSPIRE: geology, though Borehole where the relation is an association - an active well is within a borehole, and with EnvironmentalMonitoringFacilities - when the active well represents a monitoring point.

With consideration to the logical level of development, the Hydrogeology INSPIRE schema application defines four logical diagrams^f to enclose all classes and relations:

- HydrogeologyCoreView - as the name states, it offers a complete picture of the entire data model: defined entities and connections;
- HydrogeologySystem - as mentioned, the hydrogeological system is given by the interaction of the rock system and the groundwater system. Therefore, this diagram presents all encoded connections. Additionally, the relations with the ActiveWell class are represented. Furthermore, within the diagram, proprieties of groundwater are considered:
 - PiezometricState - the geometry could be represented through point or coverage;
 - ConditionOfgroundWaterValue;
 - WaterTypeValue.
- HydrogeologyAquiferSystem represents all entities and proprieties for the rock system:
 - QuantityValue;
 - AquiferTypeValue;
 - AquiferMediaTypeValue;

- AHydroGeochemicalRockTypeValue.
- HydrogeologyObjects with encoded proprieties:
 - NaturalObjectTypeValue;
 - StatusCodeTypeValue;
 - WaterPersistenceValue;
 - ActiveWellTypeValue .

Each encoded propriety has an already defined and validated code list that is readily available on the official website^g. Two main aspects must be stated: (1) The INSPIRE Directive is a legislative act that binds 28 European states to implement all of its requirements by 2021. Thus, with respect to the hydrogeology schema application, the same as for the geology one, the scope was to construct the framework to represent information on hydrogeological maps of scale 1:50000 or smaller (national and regional) and (2) The data model has no defined elements for measurements on quality and chemical characteristics. Furthermore, the data model does not support time series measurements of groundwater level within groundwater wells. When such data is available, the usage of WaterML 2.0 standard is highly encouraged. WaterML 2.0 is a data exchange standard for encoding hydrological time series developed by the Open Geospatial Consortium - OGC^h.

4 Conclusions

The data models selected list is not complete, the selection has been guided by several parameters, including: time and language of development, scope, usage and known implementations. The direction of geodata management in geosciences is clearly steered towards using non-proprietary, modeling languages, compliant with international standards that would allow a seamless integration, usage and visualization of geodata. Among existing methods of digital encoding of geographic information, in recent years, one of the most prevailing method proved to be the application of markup languagesⁱ. The motivation lies in a number of reasons:

- The Open Geospatial Consortium, one of the most significant standardization organizations worldwide, is using XML as the modeling language for the developed standards;
- INSPIRE implementation rules are written in XML;
- markup languages are non-proprietary, open and platform agnostic;
- there are tools that are especially developed to be able to ingest data schema written in XML, such as the

^f A logical diagram is a structural element of Unified Modeling Language that allows the construction of connections between defined entities, with respect to the conceptual model. A logical diagram offers a standardised vizualization of the conceptual model.

^g <http://inspire.ec.europa.eu/codelist> last accessed December 2018

^h <http://www.opengeospatial.org/> last accessed December 2018

ⁱ In computer text processing, a markup language is a system for annotating a document in a way that is syntactically distinguishable from the text.

Hale - the Humboldt Alignment Tool13 - developed within an FP7 program.

The development of the three analyzed structures spanned over a period of 15 years (2001 - 2017), a time in which the technology and the globalization of efforts in geoscience have greatly intensified with significant results. These changes are visible in our segment of analysis, as well. If between 2001 - 2006, the focus was on the development of data base structures that could be coupled with specific platforms - GIS platforms - for storage, processing and querying hydrogeology specific geodata, as the technology advanced, storage capacity and web-based data transfer developed, so did the ways of encoding geodata evolved. The focal point shifted towards a uniformity in the way of coding geospatial data, regardless of the geoscience domain in question. Hydrogeology followed the same path. We believe that it is safe to assume that an exhaustive data model for hydrogeological geodata management is still at project level, especially when considering implementation prototypes. Nonetheless, in the context of the current Information Age, for a valid data integration and interoperability level, the necessity of using common standards to store, process, transfer and visualize geodata or compliant data models is, without a doubt, essential.

References

1. Gogu, R., Carabin, G., Hallet, V., Peters, V. & Dassargues, A. Gis-based hydrogeological databases and groundwater modelling. *Hydrogeol. J.* **9**, 555–569, DOI: DOI10.1007/s10040-001-0167-3 (2001)
2. Wojda, P., Brouyère, S., Dassargues, A., *Geospatial Information in Hydrogeological Studies*, **chap. Groundwater**, John Wiley & Sons, Ltd., (2006)
3. Boisvert, E. & Brodaric, B., GroundWater Markup Language (GWML) - enabling groundwater data interoperability in spatial data infrastructures. *Journal of Hydroinformatics* **14**, 93–107, DOI: 10.2166/hydro.2011.172 (2012)
4. Gogu, C. R. Habilitation thesis, (2012)
5. Castany, G., *Prospectiunea si exploatarea apelor subterane*, Editura Tehnica (1972)
6. Andrienko, N., Andrienko, G., Gatalsky, P., Visual data exploration using space-time cube. In *Proceedings of the 21st International Cartographic Conference (Durban, South Africa)*, **1981–1983** International Cartographic Association, (2003)
7. Enescu, I. I. et al. Hypercube-based visualization architecture for web-based environmental geospatial information systems. *The Cartogr. J.* **52**, 137–148, DOI: 10.1080/00087041.2015.1119469 (2015), <http://dx.doi.org/10.1080/00087041.2015.1119469>
8. Wojda, P., Brouyere, S., An object-oriented hydrogeological data model for groundwater projects. *Environmental Modelling and Software* **43**, 109–123, DOI: 10.1016/j.envsoft.2013.01.015 (2013)
9. Cox, S. Observations and measurements - xml implementation, (2011)
10. Sen, M., Duffy, T., Geosciml:Development of a generic geoscience markup language, *Comput.&Geosci.*, **31**, 1095–1103, DOI: <http://dx.doi.org/10.1016/j.cageo.2004.12.003> (2005)
11. Geology, I. T. W. G. D2.8.ii.4 data specification on geology – technical guidelines, (2013)
12. J. Vogt, A., A., S. Peedell., Implementing the GIS elements in the water framework directive. *Tech. Rep.*, Joint Research Centre (2002)
13. Paolo Villa, M., A., G., Roderic Molina. Data harmonisation in the context of the European spatial data infrastructure: The Humboldt project framework and scenarios. *Earth Obs. Glob. Chang. (EOGC)* **179** – **191**, DOI: 10.1007/978-3-642-32714-8_12 (2013)