IOP Conf. Series: Earth and Environmental Science **44** (2016) 032025 doi:10.1088/1755-1315/44/3/032025

Engineering-Geological Data Model – The First Step to Build National Polish Standard for Multilevel Information Management

Grzegorz Ryżyński [1], Tomasz Nałęcz [1]

[1] Polish Geological Institute – National Research Institute, 4 Rakowiecka St., 00-975 Warsaw, Poland

E-mail address: grzegorz.ryzynski@pgi.gov.pl

Abstract. The efficient geological data management in Poland is necessary to support multilevel decision processes for government and local authorities in case of spatial planning, mineral resources and groundwater supply and the rational use of subsurface. Vast amount of geological information gathered in the digital archives and databases of Polish Geological Survey (PGS) is a basic resource for multi-scale national subsurface management. Data integration is the key factor to allow development of GIS and web tools for decision makers, however the main barrier for efficient geological information management is the heterogeneity of data in the resources of the Polish Geological Survey. Engineering-geological database is the first PGS thematic domain applied in the whole data integration plan. The solutions developed within this area will facilitate creation of procedures and standards for multilevel data management in PGS. Twenty years of experience in delivering digital engineering-geological mapping in 1:10 000 scale and archival geotechnical reports acquisition and digitisation allowed gathering of more than 300 thousands engineering-geological boreholes database as well as set of 10 thematic spatial layers (including foundation conditions map, depth to the first groundwater level, bedrock level, geohazards). Historically, the desktop approach was the source form of the geological-engineering data storage, resulting in multiple non-correlated interbase datasets. The need for creation of domain data model emerged and an object-oriented modelling (UML) scheme has been developed. The aim of the aforementioned development was to merge all datasets in one centralised Oracle server and prepare the unified spatial data structure for efficient web presentation and applications development. The presented approach will be the milestone toward creation of the Polish national standard for engineering-geological information management. The paper presents the approach and methodology of data unification, thematic vocabularies harmonisation, assumptions and results of data modelling as well as process of the integration of domain model with enterprise architecture implemented in PGS. Currently, there is no geological data standard in Poland. Lack of guidelines for borehole and spatial data management results in an increasing data dispersion as well as in growing barrier for multilevel data management and implementation of efficient decision support tools. Building the national geological data standard makes geotechnical information accessible to multiple institutions, universities, administration and research organisations and gather their data in the same, unified digital form according to the presented data model. Such approach is compliant with current digital trends and the idea of Spatial Data Infrastructure. Efficient geological data management is essential to support the sustainable development and the economic growth, as they allow implementation of geological information to assist the idea of Smart Cites, deliver information for Building Information Management (BIM) and support modern spatial planning. The engineering-geological domain data model presented in the paper is a scalable solution. Future implementation of developed procedures on other domains of PGS geological data is possible.

1. Introduction

Development of information and communication technology (ICT) has strongly effected the change of planning decision methods. In the era of information society building Spatial Data Infrastructure (SDI) is a background to create efficient IT system [1] supporting multilevel decision support systems (DSS) as well as planning support systems (PSS) [2]. Geological data is one of the crucial elements of SDI used in spatial planning, construction and investment, environment protection and water management, both on regional and local levels. The efficient data management and processing in engineering-geology, like in other earth science domains, needs implementation of data standards, which are common in geology (GeoSciML[3], GroundWaterML [4, 5] or BoreholeML). Implementation of standards is also one of the important goals of geological data integration, especially when heterogenic information resources are under consideration.

During last 20 years' database including over 300 000 shallow engineering-geological boreholes has been gathered in Polish Geological Institute. It was a basic resource to create a set of engineering-geological maps, including foundation conditions, depth to the first groundwater level, bedrock level, geohazards maps and other. Delivering up-to-date resources for DSS and PSS is crucial for the efficient management process of engineering geological information, but it requires exclusion of uncorrelated and diffused data resources having background in desktop approach form the past. Usage of spatial data modelling by the use of UML [6] was the first step to create integrated solution in scope of engineering-geological data description and it allowed the development of the national data standard. Standardization, in general, concentrates on modelling at the abstract level, but when stepping down to the implementation level, technology issues should be considered.

2. Engineering-geological data base historical background

2.1 The local authority's maps in the scale 1:10000

The very first attempt to create engineering-geological data base has been done in late 1990`s when the digital background for engineering-geological and environmental mapping on local level was introduced for spatial planning [7]. The main idea was to deliver guidelines for visualisation of geological data in the unified and customer friendly manner.

2.2 Engineering-geological databases

2.2.1 Engineering-geological atlases – diffused data

Engineering-geological atlases of biggest Polish urban agglomerations are the largest unique collection of digital data of this type in Poland. They include detailed information obtained from geo-engineering, geotechnical, hydrogeological reports as well as borehole profiles. The cities for which the atlases have been prepared are Warszawa, Katowice, Kraków, Poznań, Wrocław, Trójmiasto, Wałbrzych, Rybnik, Łódź. Also, they help to make decisions related to planning detailed substratum surveys and minimizing the environmental damage as well as preparing prognoses and economic aspects of investments. Layers with data on geological and economic threats can be analysed to prepare maps of risk.

Engineering-geological atlases in the scale of 1:10 000 and borehole databases were prepared on request of the Ministry of Environment. Contractors were selected in a public tender procedure. From 1998 till 2012 such situation resulted in heterogenic methodologies of data gathering and visualisation, resulting in nine separate (desktop - interbase) file databases. The need for integration of these databases emerged. Since 2013 Polish Geological Survey was given, by the Ministry of Environment the task to unify the existing 9 databases and prepare 6 new areas including: Bydgosz, Koszalin, Płock county, Piaseczno county and selected sections of Baltic cliff coast shoreline.

While investments in urban areas are prepared and planned, it is necessary to have a considerable amount of varied information about natural geo-engineering conditions, infrastructure, and ways of land use, ownership relations etc. The possibility of processing the obtained data in various ways is equally important. This is the main purpose of gathering of the engineering-geological data in PSG in form of Engineering Geological Database (EGDB).

The main problems with separate 9 desktop interbase engineering-geological databases were at first: lack of unified dictionaries of lithology, genesis and stratigraphy, accordant to Polish Standards and European ISO/Eurocode classification standards. The second problem was connected with the coordinates of boreholes, as each of archival 9 databases was in different projections. Lastly, the big barrier towards integration and reprocessing of borehole database was the high amount of transcription errors in table fields containing the soil/rock symbol, stratigraphy and genesis ("the human factor"). This situation blocked any advanced and quick reclassification and geo-processing of archival interbase datasets. The quality of data is now constantly improving thanks to implementation of abovementioned dictionaries and use of "data input wizard", which significantly improves data quality by not allowing the data input "by hand". The performed engineering-geological data unification allowed merging all 9 archival datasets and 6 newly gathered datasets into one unified professional geodatabase in Oracle 12 standard. This is a milestone for an advanced GIS geo-processing implementation in the domain of engineering geology in Polish Geological Survey.

This unified dataset has a very practical application. Appropriately prepared geological information allows the assessment of geo-engineering conditions of the subsoil in urban areas for the purposes of spatial planning, for instance when selecting the location of housing estates or mapping out a number of possible routes of linear construction features and underground infrastructure (highways, railways, pipelines, high voltage lines, etc.) Also, the use of GIS-processable engineering-geological data helps to make decisions related to planning of detailed geotechnical site investigations and to minimize the environmental damage. Also engineering-geological data is very useful in preparation of economic aspects of investments prognoses [8].

2.2.2 GEOSTAR – the first step towards local standards

All borehole data in Engineering Geological Database (EGDB) is gathered in the database format of Geostar software. The Geostar is one of the most popular geological software package in Poland. It is used by over 400 private and public parties, including universities, mining plants, geotechnical consultants, research institutes and small and medium geological companies.

The GeoStar package enables creation of documentation cards for borehole profiles, dynamic and hydrogeological soundings, as well as the generation of geological sections of selected boreholes or their straight-line projections. The package has a modular structure, allowing the user to select just the modules needed, including creation of cross section, borehole and sounding logs and documentation maps.

Borehole database which is used by Geostar software has become an informal standard for geotechnical and engineering-geological boreholes gathering and management. The database is in interbase format (*.gdb). It is a desktop file geodatabase. Its structure is presented in a simplified manner on figure 1. Throughout the years of Geostar software development since 1995 the database structure evolved and constantly grown, however the main tables and fields remained the same (tables GS_Boreholes and GS_Profile log). This allows the Geostar databases to be fully backwards compatible. This approach allows to fully use all archival databases from archives of private and public companies and geological authorities (which is often stored on CD/pendrives/hard disks in interbase files). The database structure (see fig. 1) is a set of related tables, containing only one spatial component – the x,y and z coordinates of the borehole. Two main tables are GS_Borehole (it contains the information about localisation, depth, contractor, date of drilling, drilling method, etc.) and GS_Profile Log (includes litho-stratigraphic profile with information about genesis of strata and the reclassification of borehole profile into engineering geological, parameterized units.

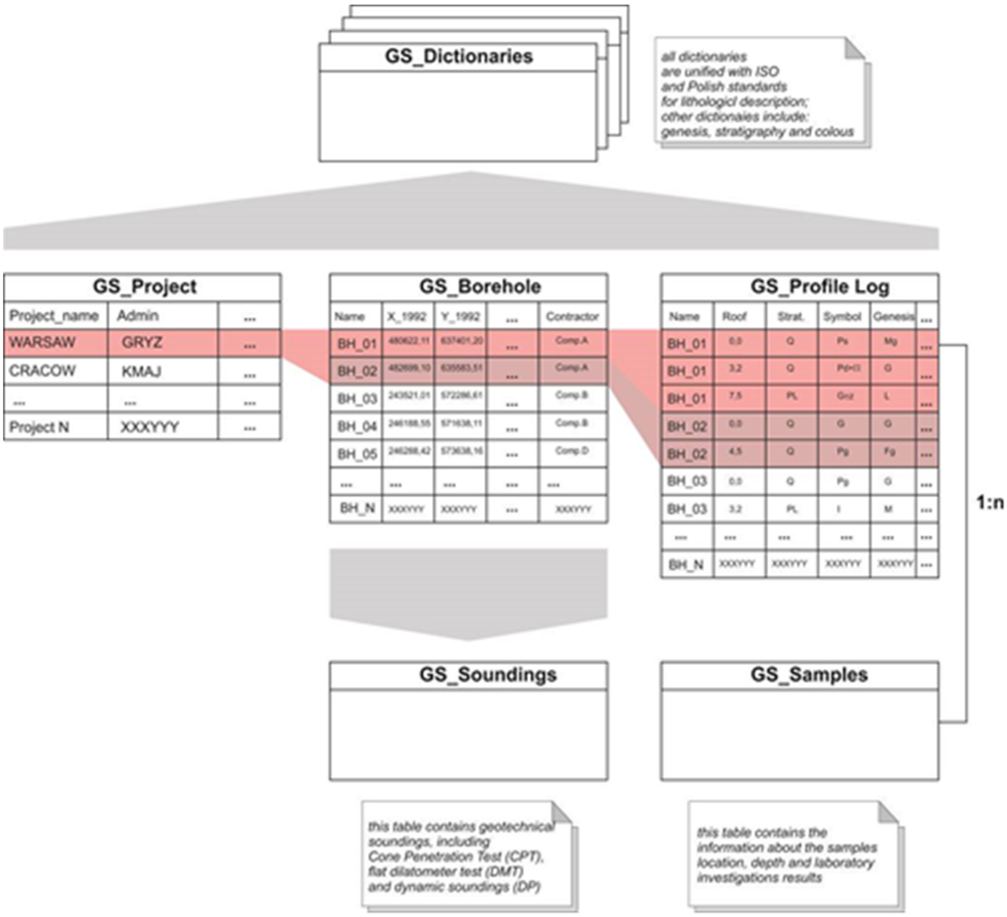

Figure 1. The simplified scheme of GEOSTAR Borehole Data Base structure

Dictionaries used for Geostar software for Engineering Geological Database were developed in accordance with Polish Standards for soil and rock classification as well as with the ISO/Eurocode EU standards. Other tables, such as GS_Soundings and GS_Samples contain optional information about geotechnical soundings profiles accompanying the boreholes and information about the depth, location and amount of samples collected for certain strata and the laboratory testing results.
Interbase databases (*.gdb desktop standard) are widespread among geological companies, geotechnical consultants and contractors. Large part of engineering-geological data collected from these parties by Polish Geological Survey for National Geological Archive (obligation by the Mining and Geology Law) has the same *.gdb desktop format and Geostar database structure.
As there is no official standard for engineering-geological borehole data in Poland, Geostar data structure was considered by PGS as a “engineering geological borehole data standard”. A good example of borehole data standardisation can be the AGS format for borehole data developed by British Geological Survey (see http://ags.org.uk/data-format/).

3. Engineering-Geological Data Base (EGDB) – standard creation process

The idea and assumptions were to transfer, after verification, the Geostar database structure (figure 1) into the structure of Central Geological Database (CGDB) on Oracle 12 server. The need for such transfer resulted from the fact, that PGS online borehole data browser (prepared for data visualisation in one, unified web browser window) works only on the data in CBDG structure. The aim for data integration and standardization in PGS is to allow presentation of geological information from different

IOP Conf. Series: Earth and Environmental Science **44** (2016) 032025 doi:10.1088/1755-1315/44/3/032025

domains in one browser. Figure 2 shows the scheme of transferring the borehole data from Geostar structure into centralized CGDB structure. Two main tables GS_Boreholes and GS_Profile_Log are imported with the use of conversion algorithms.

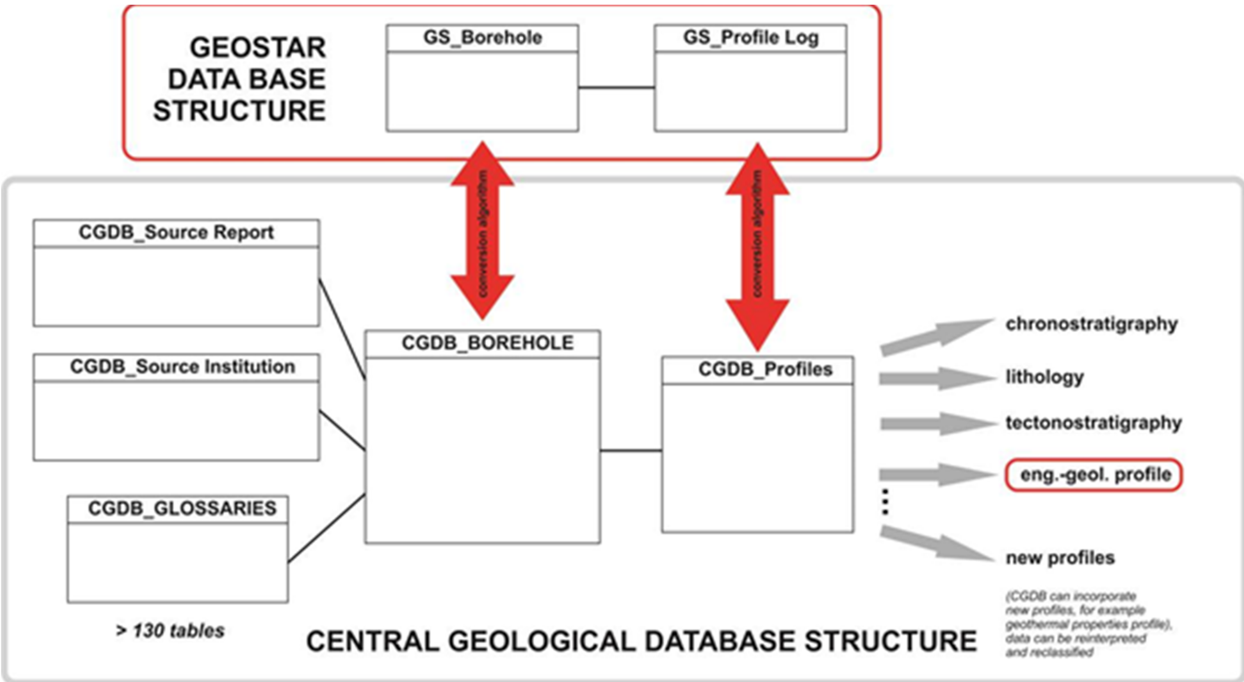

Figure 2. The scheme of import of GEOSTAR Borehole Data Base into Central Geological Data Base (CGDB)

CGDB is the main unified borehole and spatial data repository in PGS and its tables structure was designed to allow to cover and store all domains of geological information. The boreholes can be stored in full scope of depths, ranging from a few meters up to few kilometres with all accompanying descriptions of their profiles.

The figure 3 presents the generalized workflow of engineering-geological information exchange between PGS and stakeholders. The idea is to disseminate the engineering-geological borehole data from one unified repository (CGDB) via the Geological Information Office – a dedicated department focused on data management and dissemination. Various stakeholders will always get the reference geological data and a standardized report from CGDB, so the developed standard for engineering-geological data will be used as a basis for data exchange. Also, importing the geological information–borehole data from other parties (such as SDI General Director of National Roads and Highways, geological authorities, private companies) will be facilitated due to standard implementation, as data can be imported directly into CBDG structure.

The general management scheme of engineering-geological information in PGS is presented on figure 4. The p-EGDB component (module) is dedicated to borehole data and is managed by the Geostar software. The three levels of borehole data management are present – data gathering on a local server, domain analyses within dedicated Oracle scheme and cross topic analyses after importing engineering-geological data into CGDB structure on oracle server. However, the borehole data is the basis for generation of 1:10 000 scale engineering-geological maps and spatial layers. For presentation and visualization of over 300 000 boreholes over 3000 generics, 1:10 000 scale maps are needed to be generated.

IOP Conf. Series: Earth and Environmental Science **44** (2016) 032025 doi:10.1088/1755-1315/44/3/032025

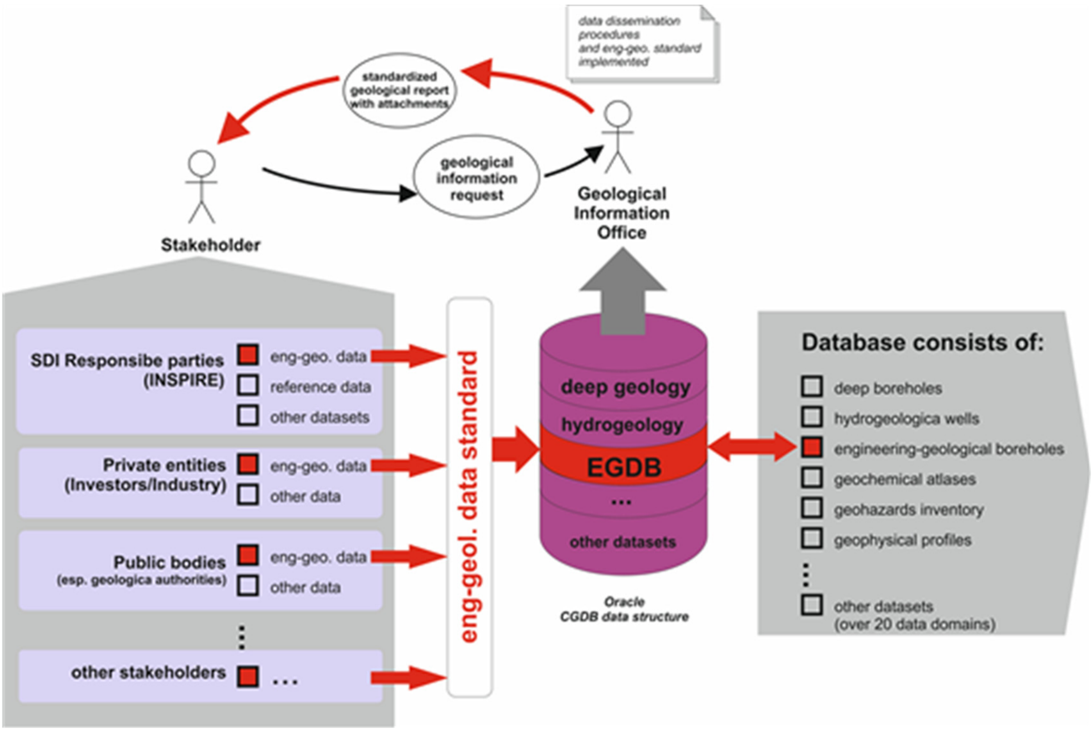

Figure 3. The schema presenting the workflow of the Engineering-Geological Data Base (EGDB) information exchange by the use of data standard

To set up and effectively use automated GIS map production tools the spatial data module m-EGDB was developed. As an effect the complex solution – Engineering Geological Data Management System was developed (see figure 4). The system in its spatial data component uses the ESRI Production Mapping 10.3 tools for management of symbolization and layout of set of 10 thematic engineering-geological maps.

Without standardization and unification of engineering-geological borehole data there would be no possibility for application of any cartographic production system. Final step of engineering-geological data management is the online presentation and dissemination of web services and downloadable pdf's with cartographic products and borehole logs generated by Geostar software (this task is currently in progress – see atlasy.pgi.gov.pl). Also, providing the unified dictionaries online and dissemination is essential for improvement of the quality of the engineering-geological and geotechnical data collected by the private and public companies.

4. Results and Discussions

The presented solution allowed to achieve a number of important objectives for the engineering-geological resources management. Firstly, database structures from several independent past projects of engineering-geological atlases have been integrated with GEOSTAR solutions that are the most popular on Polish market. Secondly, engineering-geological data model has been developed and integrated with the borehole model of the Central Geological Data Base, which finally allowed the efficient borehole information management.

Thirdly, a comprehensive system for borehole management covering the whole process form the gathering data in the field to publication of spatial information on maps or web browsers was developed. Fourthly, a set of IT procedures and solutions has been produced which, in the future, will be applied in other domains as part of the implementation in PGI enterprise architecture.

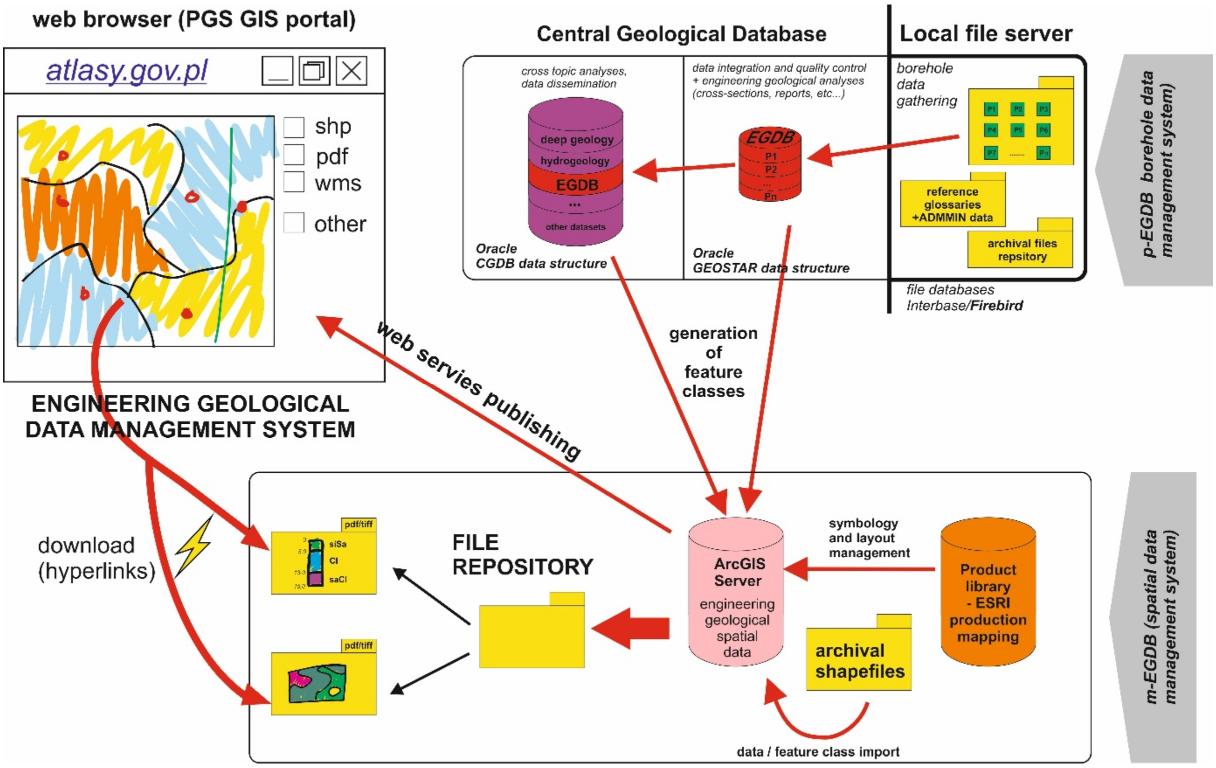

Figure 4. The schema of engineering-geological data acquisition, processing and dissemination integrated process

Nevertheless, the most important achievement of the project is preparation of IT infrastructure that enables the dissemination of the data model as the national engineering-geological standard by using UML modelling tools among variety of users. According to the adopted assumptions it is possible to import the model into any user's application and to finally implement it as a homogeneous data structure. As a result, the geological companies as well as public administration will be able to automatically transfer borehole results to EGDB while maintaining compatibility with the structures of the central PGI database.

5. Conclusions

Building and deployment of engineering-geological data standard is a crucial step towards implementation of data integration process in Polish Geological Institute. Procedures developed in that process are scalable and ready to be implemented to other domains as the best practice. Delivering standard on national level will increase the awareness on the role and importance of engineering-geological data, especially in decision support process. It will also allow PGI to disseminate the geological information among the number of scientific partners and carry on innovative scientific research projects focused on development of new methods of data processing and visualisation [9]. It should also strongly influence data exchange and growth of digital information resources. Easy access to data and customer friendly applications will certainly encourage decision makers to use the information on the daily basis. On the other hand, growing amount of standardised geological information should positively influence development of new technologies supporting ideas of Building Information Management (BIM) or Smart Cities, where geological resources will be an important component that will facilitate the investment process. All the above mentioned factors will influence the development of national SDI as the basic digital infrastructure supporting multilevel decision processes.

Acknowledgement(s)
Acknowledgments to ESRI Poland, Soft Project and TAXUS IT for cooperation in integration of engineering-geological database of Polish Geological Institute project.

References

[1] Nałęcz T., Michalak J., 2011. The INSPIRE conceptual model for the theme geology as an introduction to description and schematization of hydrogeological phenomena, *Annals of Geomatics,* Vol. 9, 4(48): 117-128, Warsaw
[2] CGI (Commission for the Management and Application of Geoscience Information), 2010: GeoSciML Cookbook - How to Map Data to GeoSciML Version 2.1
[3] http://ngwd-bdnes.cits.rncan.gc.ca/service/api_ngwds/en/gwml.html
[4] Boisvert E., Brodaric B., 2008. GroundWater Markup Language Specification v. 1.0
[5] Majer E., Sokołowska M., Ryżyński G., 2013. Identification of geological risk in investing process, Proceedings of XXVIII Nationwide Construction Designer Workshop, p. 305- Wisła, Poland
[6] Han S.Y., Kim T.J., 1989. Can expert systems help with planning? *Journal of the American Planning Association* 55(3): 296-307
[7] Bażyński J., Drągowski A., Frankowski Z., Kaczyński R., Nałęcz T., 1999. The guidance of engineering-geological conditions maps creation in scale 1:10 000 and higher for the needs of spatial planning in local authorities. Ministry of Environment, Warsaw (http://www.pgi.gov.pl/pl/component/docman/doc_download/246-instrukcja-sporzdzania-mapy-warunkow-geologiczno-inynierskich-w-skali-110-000)
[8] Jankowski S., Chreczka A., Szymański Z., Ryżyński G., 2014. Modelling engineering-geological layers with K-NN and neuronetworks, Neural Networks and artificial intelligence, 8th International Conference ICNNAI-2014, Brest Belarus, Proceedings Series: *Communications in Computer and Information Science*, Vol. 440, Springer
[9] ISO/IEC 19501:2005, Information Technology – Open Distributed Processing – Unified Modelling Language (UML) Version 1.4.2, International Organization for Standardization (ISO), Geneva, Switzerland