The Database for The Processing of Environmental Data

Monika Chuchro¹, Wojciech Sarlej¹

¹ University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, Mickiewicza 30 st. 30-059 Krakow, Poland

chuchro@agh.edu.pl

Abstract. The main purpose of the research was to design a database for collected measurements of the Computer System for Monitoring River Embankments (ISMOP) project. The ISMOP project aimed to conduct research on comprehensive monitoring and forecasting the flood embankments condition. As a part of the project, measurement data were collected continuously, transformed, interpreted and analysed using computer simulations. The assumption of the database was to achieve high availability of data in a short time, allowing quick (in less than 15 minutes) decision about the state of the flood embankment. The warehouse project contains large amount of time series data sets, 2D numerical models, analysis results of flood embankments condition and additional measurements like thermographic simulations. Simple database structure was proposed. Data were loaded into small number of tables, data types were adjusted for high accessibility. Such structure allows to use quick data selection for analysis and visualization and prevents from unnecessary data redundancy. To effectively visualize and analyse sensor's data an additional ArcGIS plugin tool was created. The plugin offers the spatial visualization functionality of sensor placement (2D and 3D views), their selection, distance and area measurements, statistical calculations and spatial analysis. The tool also provides ability of manual modification and exporting data into shape and spreadsheet files for further analysis. For the examples that were carried out, the database structure and the plugin were efficient and results were quickly achievable. Data were presented in a manageable way in form of 2D charts and 3D visualizations.

1. Introduction

The Computer System for Monitoring River Embankment (ISMOP) project was aimed to conduct measured parameters research and forecasting of the flood embankments conditions. Measurements were gathered in a continuous way. Simulations, analysis and interpretation have been conducted in real time. The research was carried out on an experimental flood embankment located in Czernichów, near Krakow by the consortium consisting of AGH University of Science and Technology departments (Computer Science, Hydrogeology and Engineering Geology, Geoinformatics and Applied Computer Science and two companies – NEOSENTIO and SWECO Hydrotechnik Krakow – in co-operation with the Czernichow Community Council [1].

The experimental embankment was built as a two parallel sections 150m long and 4.5m high, connected by a meandering (Fig.1). Inside the experimental flood embankment, more than 1300 sensors were placed to measure temperature, pore pressure, vertical displacements; as well as inclinometers, and two strands of an optical fiber [1,2]. Simultaneously with data collecting from experimental embankment, numerical modelling simulations were performed. Calculations were focused on describing the impact of water pressure on the flood embankment and the impact
of increasing and decreasing reservoir water level on the phenomena that occur within the embankment [3].

![Figure 1. Experimental flood embankment (img. Sonia Bazan, http://losa.tech)](image)

Figure 1. Experimental flood embankment (img. Sonia Bazan, http://losa.tech)

The project proposed a flow of data collected by sensors and numerical modelling, presented in Fig. 2. Measurements were conducted by sensors placed in the experimental embankment with a time step of 15 minutes. Collected data was sent to the database in which they are stored as table. Data cleansing and analysis process works on intermediate table which is a duplicate of the initial table. In addition, the same database keeps numerical modelling, with the same time step, performed in the FLAC 2D program for vertical cross-sections analysis. Results of predictive models as well as point and area models of short-term forecast of the embankment’s state are stored in separate tables.

![Figure 2. Data flow](image)

Figure 2. Data flow [4]

2. Database
The main repository has been created in MS SQL Server 2014 [5]. It has been chosen for several reasons. First of all, free access to the license is guaranteed under MSDN AA program. In the case
of commercialization of the system, the cost of purchasing a commercial license is relatively low and there is no need to migrate schema and data into different database vendor. Furthermore, MS SQL Server has the ability to connect to the ArcGIS Desktop software, which is the main part of the implemented application [6]. Such approach allows users to manually edit the database without SQL language knowledge or managing the DBMS. This solution can be particularly convenient for system operators, for example in local governments, municipalities, etc. The third advantage is the time efficiency of queries, which compared to open source software (PostgreSQL / PostGIS) is definitely better. The disadvantage of the solution is the necessity to put the database on a Windows Operating System server and not Linux [7].

One of the requirements for database design for all data collected during the controlled flood experiments, was simultaneous storage of many types of data with a very large number of records (for example, real data in the form of time series, data from numerical modelling as a multidimensional data tables, results of analyses and embankment state prediction as an individual records, data in the form of graphics). Archiving the results of numerical modelling is a non-trivial matter due to the volume of generated data (as a result of numerical modelling), as well as the time needed to read them. The analysis of archived data had to take place as efficiently as possible and within limited time [4,8].

Therefore, it was necessary to design an appropriate database structure, where SQL queries would be free of time delays, caused by a large number of joins or readings of subsequent tables [8,9]. Preparing database also had to provide the possibility of simplified analysis of information gathered in the form of time series and processed using data mining methods. These analyses were part of the system module (model-driven and data-driven), assessing the risk of the flood embankment damage by matching information (models) stored in the expert system that used real data, analysis and modelling results [6,7]. The comparison of real data coming from sensors, due to project’s requirements, has to be conducted in the fastest possible way, with particular preservation of the unambiguity of the generated results. For the smallest scenario (numerical modelling group for specific external embankment conditions) with number of model nodes equal 50000, a table containing 75 million records has been generated, with a significant impact of processing time factor. Several hundred scenarios have been defined, therefore the simulated data will be a gigantic data set [3, 9, 10]. As a result of tests, a database structure was proposed for real and synthetic data. Computer simulations data and time series were stored in the form of a single summary table, presented in Fig. 3. Such structure allowed to reduce number of tables that corresponded to particular scenarios. Based on the analysis of the distribution of the embankment measurement network, data has been limited to compare only the values in the model nodes whose position corresponds to the position of the sensors. This solution allowed to modify the structure of the database to extract only one table that stores data from all scenarios. This table has been called the node table and is a source for model-driven analysis. Node table contains the following attributes: geographical coordinates, scenario and simulation identifiers with computed value. In addition, the node table also contains a sensor field that stores information about the type of sensor. It allows to use indexation by sensor type, resulting in a significant increase in the search efficiency of the table. Proposed approach resulted in reduced number of calculations and time overhead which is needed to search and read subsequent tables. The identification of models has been implemented with prepared database using aggregated time windows.

The base schema includes a series of tables which structure is identical to text files generated as batch data for the system. The exchange of numerical modelling data can be proceeded without any additional filtering or intermediate transformations. In case of results gathered during experiments on the embankment (sensors’ measurements during flooding), the structure of the table was
reconstructed on the basis of files downloaded from the web application. This solution is time and resource efficient and also enables visualization of data in real time.

Figure 3. Synthetic scenarios data warehouse

Application, presented in the further part of article, also uses data in form of the GIS model (Fig 4). To the database containing all data and geodetic and geological measurements, additional table has been added. New table contains the coordinates of Neosentio’s sensors (grey dots in Fig 3), reference sensors (UT sensors- green dots in Fig 3) and optical fibers (red dots in Fig 3). By this enhancement, user can select any sensor from the above tables and visualize the corresponding data from the measurement database or numerical modelling.

Figure 4. Database visualization system
3. Application

The assumptions of the application included creation of 2D and 3D sensor visualization ArcGIS plugin using data provided by a database that contains archived time series (see Figure 5). Moreover, simple and intuitive user interface had to allow selection of the section type, multiple cross-sections and specific sensors selection based on their identifier and implementation of mechanisms that allow to choose a sensor and n-neighbouring sensors. Due to use the GIS Desktop developed by ESRI, application is simplified. The ArcGIS Desktop 10.5 allows to perform all operations related to data transformation, native connection with a database. Furthermore the ArcGIS has geostatistical and 2D and 3D visualization packages and geoprocessing tools. In combination with the GIS Server, which provide created GIS model share in a WebGIS application, was obtained simple in usage application allowing realization of multiple visualization problems.

The application allows users to perform multiple data tasks and analyses like pre-processing, real data and scenario visualizations, real data and scenario comparisons, real data predictions, as well as simple results displaying. Sample visualization of 2D sensors measurements inside experimental flood embankment has been presented in Figure 6. 3D view has been shown in Figure 7. Each sensor type and corresponding scenarios are marked in different colour (Fig 3). Application also allows to choose multiple layers for visualization: scenarios, simulation, analysis results and real data can be chosen from right menu (Fig 5). User can also choose intersection axis for displaying 2D graphs (Fig 6). Charts can be exported into files and separate databases for further analysis. Every time the source data is updated the tool automatically redraws visualizations. As per requirements, the plugin creates charts in very quick and effective manner. Entire simulation (from beginning till end) for a section takes around 1 second which fits into analyst’s needs.

![Image of sample measurement visualization in ArcGIS plugin.](image-url)
4. Conclusion
The Computer System for Monitoring River Embankment (ISMOP) project due to the need of storing a huge amount of data required the tool allowing quick data access, data modification and data
recording with simple visualization for multidimensional time series data. In this paper were presented database project with visualization plugin.

Database is free of time delays, caused by a large number of joins or readings of subsequent tables, through project with limitless tables number. The exchange of numerical modelling data can be proceeded without any additional filtering or intermediate transformations. In case of results gathered during experiments on the embankment (sensors’ measurements during flooding), the structure of the table was reconstructed on the basis of files downloaded from the web application. This solution is time and resource efficient and also enables visualization of data in real time. Created application allows multiple types of visualizations for real data, numerical models, results of analyses also for superimposed data types. The amount of time needed for chart type and data selection, connection with database and plotting for each flood embankment section is lower than 1s.

Acknowledgment(s)
The research presented in this paper was partially supported by the National Center for Research and Development (NCBiR) under Grant No. PBS1/B9/18/2013 and by the AGH grant no. 11.11.140.613.

References
[1] K. Sekuła, A. Borecka, D. Kessler, and P. Majerski, “Smart levee in Poland. Full-scale monitoring experimental study of levees by different methods,” Computer Science, vol. 18(4), pp. 357–384, 2017.
[2] J. Stanisz, K. Korzec, A. Borecka, “ISMOP Project (IT System of Levee Monitoring as an example of integrated monitoring of levee,“ Geology, Geophysics & Environment, vol.41(1), pp. 137–139, 2015.
[3] M. Dwornik, A. Pięta, J. Bała, “Numerical modelling of temperature and pore pressure distribution in the embankment during flooding processes,” SGEM 2015 : Section Hydrogeology, Engineering Geology and Geotechnics : 15th international multidisciplinary scientific geoconference: conference proceedings, vol.2 (Section Hydrogeology, Engineering Geology and Geotechnics), pp. 479–486, 2015.
[4] M. Lupa, A. Leśniak, ”Geoportal jako narzędzie wspomagające prace geologiczno-inżynierskie na przykładzie projektu ISMOP”, Roczniki Geomatyki, vol.12(4), pp.411-416, 2014.
[5] Microsoft: Microsoft SQL Server 2014 (2014), https://docs.microsoft.com/pl-pl/sql/2014-tocto/books-online-for-sql-server-2014?view=sql-server-2014.
[6] ArcGis desktop: http://desktop.arcgis.com/en/documentation/.
[7] Y. Igarashi, K. Nagata, T. Kuwata, T. Omori, Y. Nakanishi-Ohno, M. Okada, “Three levels of data-driven science”, Journal of Physics: Conference Series, vol.699, 012001, 2016.
[8] B. Delipetrev, A. Jonoski D. P. Solomatine, "Development of a web application for water resources based on open source software", Computers and Geosciences, vol.62, pp. 35-42, 2014.
[9] M. Habrat, M. Lupa, M. Chuchro, A. Leśniak, "A decision support system for EmergencyFlood Embankment Stability", Procedia Computer Science, vol.51, pp.2957-2961, 2015.
[10] M. Lupa, M. Chuchro, A. Piórkowski, A. Pięta, A. Leśniak, “A proposal of hybrid spatial indexing for addressing the measurement points in monitoring sensor networks”, Beyond Databases, Architectures and Structures : 11th international conference, BDAS 2015, proceedings, Switzerland : Springer International Publishing, cop., Communications in Computer and Information Science, pp. 437–447, 2015.