Groundwater data management system

Kamil Nešetřil[1], Jan Šembera[1]

Abstract

We have developed a system to manage groundwater data (geology, hydrology and geochemistry). The system is flexible, readily extensible and adaptable. It can load data in different formats. Loading and cleaning the data is performed by an ETL tool Pentaho Data Integration (aka. Kettle). The coordinates are converted by GeoKettle (spatial fork of Kettle). The database PostgreSQL uses our own data model. Groundwater specialists can visualize data in EnviroInsite desktop software. A non-specialist can visualize data in a web application. Non-point spatial data is stored in a spatial database (PostGIS) and served via a map server (GeoServer) as a map service to our web application or to any geographic information systems (GIS). We have successfully used tools for business intelligence and spatial data infrastructure.

1. Introduction and requirements

In order to increase efficiency of hydrogeologists' work it is necessary to properly manage data. It is not a trivial task, because groundwater data consist of data on geology, hydrology and geochemistry. Crucial is the data management of primary measurements and observations obtained from points (e.g. wells, boreholes, sampling objects). Managing spatial interpretations of data (e.g. extent of aquifer, contours of solute concentration etc.) and storing them in a structured form is of lesser importance. The crucial requirement is the capability to readily add imports of new data formats (mostly semi-structured data) and add exports to software for analysis and modelling. Data needs to be accessible, in some extent, also via web client.

There are many environmental data management systems (EDMS) designed to handle groundwater data. Some of them are EQuIS from EarthSoft (earthsoft.com), SiteFX from EarthFX (earthfx.com), GW-Base from ribeka (ribeka.com), EnviroData from Geotech Computer Systems [1], Oasis-montaj from GeoSoft (geosoft.com), HydroManager from Schlumberger Water Services (swstechnology.com) or ESdat (esdat.net). Deployment of those solutions takes some effort because different countries have different data formats and different spatial reference systems. Loading or exporting of specific data formats is not available in existing EDMS and programming new ones is not a trivial task. We developed a new solution (Figure 1) that makes creation of new data imports, exports and conversion of coordinates as flexible and user-friendly as possible.

2. Technical solution

2.1. Database

It is reasonable to use an existing data model for the newly developed information system. It makes it possible to store all groundwater data in a fixed form and exchange it easily. We have reviewed available data exchange standards and data models. The only truly international standard is Ground Water Markup Language (GWML) [2] from Open Geospatial Consortium (OGC). It is an application schema of Geography Markup Language (GML). GML is significant for European users because it is widely adopted to an EU directive INSPIRE (INfrastructure for SPatial InfoRmation in Europe). Another data model is Hg2O [3] that is planned to be incorporated into GWML.

[1] Technical University of Liberec, 461 17 Liberec, Czech Republic, kamil.nesetril@tul.cz

ESRI Groundwater Data Model has been implemented in Arc Hydro Groundwater [4] that is an extension of ArcGIS. The last ones are Data Model of National Groundwater Information System [5] and H+. Because primary intention of the developed system is not the interoperability based on aforementioned standards, none of those data models was used. Besides other things some are not suitable for needs of groundwater practitioner (GWML), some are too concise or not sufficiently documented. All data models, data exchange formats and data models of EDMS were reviewed and used as an inspiration for the developed data model.

Visualization of hydrogeological data on desktop computer can be easily performed with EnviroInsite from HydroAnalysis, Inc. (enviroinsite.com) – low priced software in .NET to display maps (including localized tables and graphs), technical documentation of boreholes, geological cross-sections, 3D geological models and interpolation in 2D and 3D. Our data model is based on existing data model of EnviroInsite. Therefore the database and visualisation software has consistent data structure. That reduces need for non-unique data transformation and it does not confuse users.

The original data model of EnviroInsite (9 tables) was extended to 28 tables because the EnviroInsite data model contains data relevant for visualisation only. Original tables were extended by additional fields and the model was further normalized. It contains data on:

- observation objects (wells, boreholes, sampling points with coordinates and detailed description),
- characterisation of geological layers (description of boring logs and it’s interpretation – stratigraphy),
- technical construction of wells (casing, screen and fill of annular space),
- definition of observed quantities (units, chemical formulae etc.),
- standards (action levels, regulatory limits),
- definition of vertical intervals (well screen or sample interval in boring log),
- measurements tied to vertical intervals (e.g. chemical analyses or head measurements),
- measurements tied to specific depth (e.g. geophysical logging),
- samples (metadata about measurements and sets of measurements – sampling methods, conditions, etc.),
- anti-aliasing (e.g. a quantity has different names in different data sources),
- conversion of units (e.g. mg to g) and quantities (e.g. nitrate to nitrogen),
- time intervals, lookup tables etc.

We are using PostgreSQL (postgresql.org) database management system.

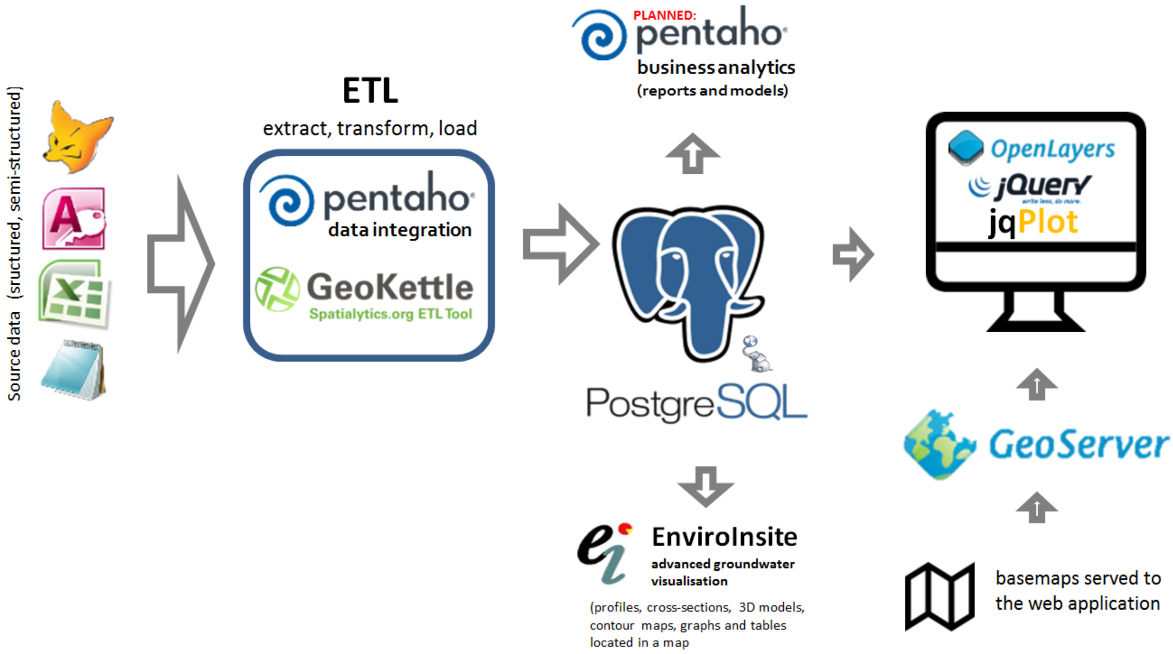

Figure 1: Schema of the presented system for management of groundwater data

2.2. ETL

The data is loaded to the database by ETL (extract, transform, load) tool Pentaho Data Integration, also known as Kettle [6]. It is an open-source software released under Apache Licence (community.pentaho.com). Data transformations in Kettle can be implemented without coding through an intuitive graphical user interface called Spoon (Figure 2), so they are also easy to maintain. Kettle transformations can be run in command-line interface (“Pan” and “Kitchen”) or on an ETL server (“Carte”). We implemented the loading of following data:

- analyses from laboratory information management system Labsystém (provided as 2 xBase files),
- boreholes and groundwater chemism from Czech Geological Survey (provided as MS Access files),
- exploratory boreholes (provided as MS Word documents created by a Geobanka software from Data-PC Sokolov),
- text files with precipitation and temperature served via FTP from a watershed authority (Povodí Ohře s. p.)
- original database format of EnviroInsite (MS Access, MS Excel) and
- “handmade” MS Excel spreadsheets.

Subsequent transformations provide data cleaning, anti-aliasing, validation and loading to the database. Coordinate conversion is performed by a spatial fork of Kettle called GeoKettle [7]. Future versions of GeoKettle should be an extension of Kettle and not a stand-alone application.

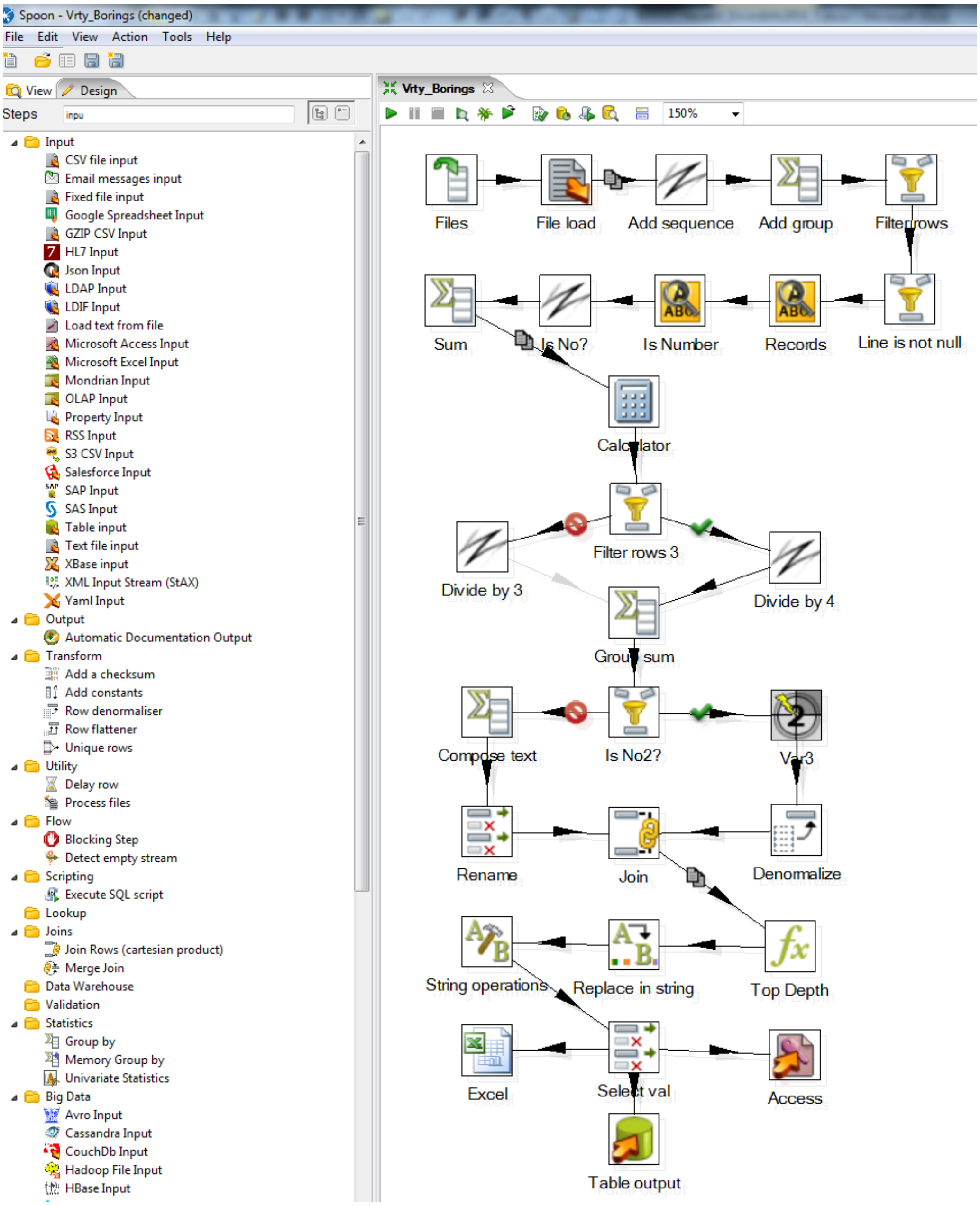

Figure 2: The data transformation implemented in Spoon (development environment of Pentaho Data Integration aka. Kettle)

2.3. Visualization

Due to the compatibility of our data model with that of EnviroInsite, data can be easily exported to MS Access or MS Excel file and visualized in EnviroInsite. This is suitable for professional hydrogeologists. Stakeholders and other nonspecialists can view data in a web application we have developed (Figure 3). The application combines a table, a graph and a map on a single screen. Upon choosing a borehole on a map and a quantity from a drop down menu, table (in left vertical strip) and graph (in upper horizontal strip) are displayed. Attributes of an observation object are displayed in a pop-up window. Different map layers provided as WMS (Web Mapping Service) and WMTS (Web Map Tile Service) can be switched on and off as basemaps. The application is implemented mostly in JavaScript (jQuery, jqPlot). Maps are integrated with use of widely used JavaScript library OpenLayers (openlayers.org).

Interpretations and non-point data (arcs, polygons etc.) are stored in the PostgreSQL database due to PostGIS. PostGIS (postgis.net) is a spatial extension of PostgreSQL. Those data and georeferenced images are served via GeoServer (geoserver.org) as WMS service. WMS can be loaded as a basemap to our web application or to any geographic information system (GIS).

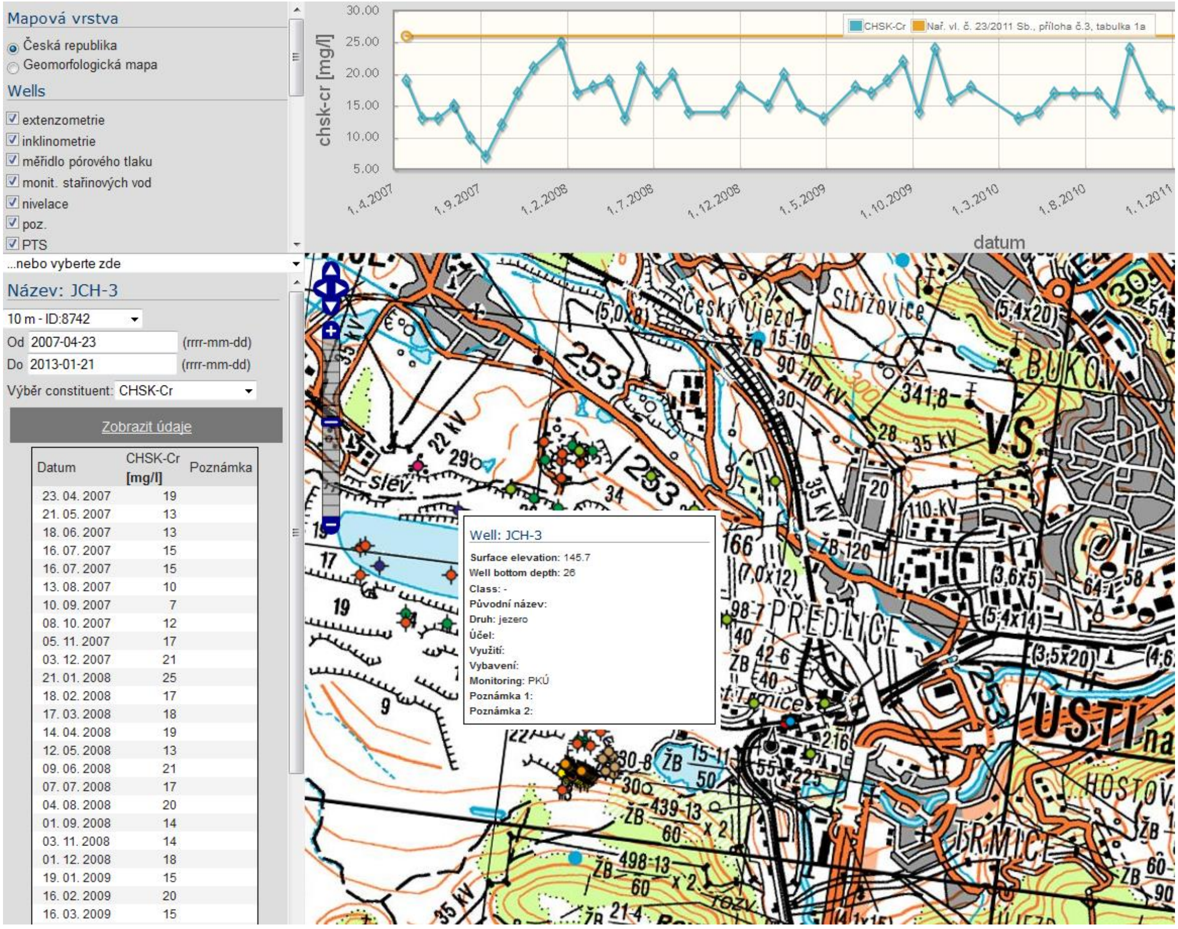

Figure 3: The web application depicting time series in the table and in the graph

3. Further plans

Common hydrogeologist's data management tasks are to automatically load data to database, create reports and alerts and export data to a third party software for analysis and modelling. Those tasks are not unique for hydrogeology but can be provided by existing open source business intelligence software. We made a very good experience with the ETL tool Kettle from Pentaho Corporation. Pentaho develops a whole stack of open source business intelligence software such as Pentaho Report Designer, Pentaho Business Analytics Platform (server) and others. By the time of the conference the reporting platform should be deployed. The exporting to third party software for analysis and modelling shall be implemented in Kettle. An alternative to Pentaho BI suite is SpagoBI (spagobi.org) with powerful visualisation and analysis of geodata including non-point data and WMS services.

4. Conclusions

We discovered real needs of groundwater professionals, available technologies and tools. We developed a flexible information system focused on needs of practitioners. We used technologies that enable long term maintenance, customization and extensibility. The architecture of the system corresponds to a spatial business intelligence solution (GeoBI) – combination of business intelligence (BI) and spatial data infrastructure (SDI). Therefore it could be used for geographical analyses and management of big data sets. Groundwater practitioners are used to work with GIS software for decades but not with BI tools. Our effort is to introduce BI to groundwater community. The system is available commercially, upon request (contact the corresponding author).

5. Acknowledgement

Work of Daniel Vrbík and students David Krejbich (transformations in Kettle) and Tomáš Jodas (web application) is greatly acknowledged. Christopher Jon Pierson proofread an early version of the manuscript.

The contribution was prepared with support of the Technology Agency of the Czech Republic via the project Nr. TA02020177 (Šembera) and by the Ministry of Education of the Czech Republic within the SGS project no. 21066/115 on the Technical University of Liberec (Nešetřil).

References

[1] Rich, D.W., *Relational Management and Display of Site Environmental Data*, Boca Raton, Lewis Publishers, 2002.

[2] E. Boisvert and B. Brodaric, "GroundWater Markup Language (GWML) - Enabling groundwater data interoperability in spatial data infrastructures," *Journal of Hydroinformatics*, vol. 14, no. 1, pp. 93–107, 2012.

[3] P. Wojda and S. Brouyère, "An object-oriented hydrogeological data model for groundwater projects," *Environmental Modelling & Software*, vol. 43, pp. 109–123, May 2013.

[4] R. Chesnaux, M. Lambert, J. Walter, U. Fillastre, M. Hay, A. Rouleau, R. Daigneault, A. Moisan, and D. Germaneau, "Building a geodatabase for mapping hydrogeological features and 3D modeling of groundwater systems: Application to the Saguenay–Lac-St.-Jean region, Canada," *Computers & Geosciences*, vol. 37, no. 11, pp. 1870–1882, Nov. 2011.

[5] Australian Government: Bureau of Meteorology, (2014, Jul. 10). *National groundwater information system* [Online]. Available: www.bom.gov.au/water/groundwater/ngis, 2013

[6] Casters, M.R., Bouman, R., and Dongen, J., *Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration*. Wiley, 2010.

[7] Spatialytics, (2013, Apr. 2). GeoKettle. Version 2.5 [Online]. Available: http://www.spatialytics.org/projects/geokettle/, Spatialytics solutions inc., 2013.