A Software Architecture for the Cooperation Project HADU:  
(Hamburg’s Dynamical Geological Underground)

Jochen Wittmann

Abstract

The cooperation project on Hamburg’s dynamical geological underground (HADU) intends an integrative view on recent geological subsurface processes in the region of the City of Hamburg. Measurement data resulting from geological georadar methods, geophysical ambient vibrations techniques, and conventional well databases shall be combined to a 3D-underground model. From the geological point of view, such a model will help to understand the georisks caused by the salt diapirs in the underground. The computer science part of the project concentrates on a common data handling, the visualization of the geological data, and the integration of Fuzzy-methods to handle incomplete and uncertain data. In this context, this paper presents the concept for the over-all architecture of the software system developed for the use in the HADU project. Main task for the developers was to bridge the gap between the demands of the projects partners concerning the integration of already existing, highly proprietary software modules on the one hand, and to offer an integrated data handling and analysis in regard to data consistency and reproducibility of the results on the other hand. Focused on this aspect, the paper defines the HADU software architecture and discusses its adequacy and practicability within the project contexts.

1. The HADU Project

As already described in the overview paper (Geotechnologien, 2005) and a first computer science state of the art analysis in Wittmann (2006), the cooperation project HADU (Hamburg’s dynamical underground) deals with the modelling of recent geological structures and their dynamics in the underground under the city of Hamburg. In this context, different measurement methods are applied: georadar, gravimetry, and ambient vibrations techniques.

The objectives for the computer science group in the project lie in a consistent data handling with distributed access for the project partners, and to guarantee consistency and reproducability of the results of the data analysis process. There are two contradicting aspects that will influence the software architecture intended: on the one hand, the software shall enable an interdisciplinary analysis by an unified view on the data, on the other hand, the data formats and analysing methods used by the partners are very complex, distinguished, and proprietary. Therefore, it does not seem to be adequate to develop a data model in the sense of a conventional relational database on the level of measurement entries and to confront the user with just another common user interface with integrated analysis functionality that does nothing else than to duplicate the interfaces of the analysis tools already in use. This conflict between unification and integration of existing tools is the main topic of this paper. Sections 2 and 3 will provide a detailed view on the software tools already in use and on the data formats generated by this software. Afterwards, section 4 proposes a pragmatic solution for the HADU software architecture that provides integration, data consistency, and usability.

---

1 University of Hamburg, MIN Faculty, Dept. Informatik, Arbeitsbereich Technische Informatiksysteme, Vogt-Kölln-Straße, 30, 22527 Hamburg, Germany; e-mail: wittmann@informatik.uni-hamburg.de

EnviroInfo 2007 (Warschau)  
Environmental Informatics and Systems Research  
Copyright © Shaker Verlag, Aachen 2007. ISBN: 978-3-8322-6397-3

159
2. Problem Analysis: Data and Workflows

This section will analyse the workflow on the side of project partners, a task that splits into two sub-tasks: Firstly the analysis of the format and the amount of the data generated, secondly the analysis workflow starting with the data acquisition by the measurements and ending with data analysis and presentation of results.

2.1 Georadar-Measurements

The investigations by georadar are executed by the institute of geophysics at the university of Hamburg (www.geophysics.zmaw.de). The measurement device is pulled over the surface according to a predefined trajectory. Thereby a high frequent signal is emitted, reflected by the ground layers, and re-collected by the georadar device. Dispersion and diffraction of the reflected signal allow implications on the situation in the ground. Doing so, reflection profiles are generated that are orthogonal to the trajectory and reach depths up to 5 to 10 meters. These profiles are stored in a separate file each. By successively changing the surface trajectory, a 3-dimensional view of the underground situation is elaborated. Basically, this task is supported by data filtering algorithms on the profile data itself and by varying the perspective on the 3-dim situation by visualization methods to gain information on the spatial relations between the measurement data collected.

Under the aspect of software technology the analysis of the workflow is easy: the raw data coming from the measurement device are stored in a pre-defined folder and afterwards processed by the open source tool „KoGeo“ (www.kogeo.de). This tool provides the necessary mathematical methods as well as visualization functionality. All the processing steps executed interactively are protocolled in an Access database by KoGeo. Structural unit on database level is the so-called project, which includes references to all data treated together with the complete log of user interactions. Each project is stored as separate *.db-file in Access format.

2.2 Gravimetry-Measurements

There is a similar situation for the execution and analysis of the gravimetry measurements (by the geological institute of the University of Hamburg). The measurement takes place at a certain point in the region of interest and provides a frequency answer for exactly this point with variable depths. This frequency spectrum is analysed manually with the intention to find a significant peak-value. At the corresponding depth to the peak a significant change in stratigraphic sequence is supposed. From methodological point of view, the precise calibration of the measurement device in regard to the exact geographical position (space coordinates and height) is of importance. This is made by methods sequentially calculated correction that allow a more precise determination of the depth of the peak value the investigator is interested in.

Also this case, the workflow stays sequential: it starts with the spectrum of frequencies measured at a certain point in space. This spectrum is stored in a distinguished file, again. The peak value is determined manually followed by a sequence of manually started methods for calibration and correction of the measurement. All these methods are realized as independent main programs and provide a single real value as result. For the complete region of interest, the investigators work with a single, comprehensive ASCII-file that contains in its rows: the exact coordinates of the measurement, the peak value, the corresponding depths, and the corrected values for depth as result of the sequentially executed correction methods. Additionally, there is some information as date of the measurement, ID of the measurement device, remarks, etc. For visualization of this result-file the Generic Mapping Tool GMT (gtm.soest.hawaii.edu) is used. As standard visualization for each measurement position a circle is plotted on a map of the region of inte-
rest. The colour of the circle codes the depths of the peak and thereby the depths of stratigraphic change at this position.

2.3 Ambient-Vibrations-Measurements

More complex is the situation for the processing of the ambient vibrations measurements of the institute of geo-sciences at the University of Potsdam (www.geo.uni-potsdam.de). It is again the intention to determine the depth of a change in stratigraphy for a certain point of measurement. This is done by measuring resonance spectra for a set of measurement devices and calculating correlations between the gathered data afterwards. Finally, there is a set of so-called velocity-depth-profiles that allow an estimation of the change in stratigraphy. Figure 1 shows the complete course of the analysis more in detail. All calculation methods needed are implemented in the freeware package “geopsy” (www.geopsy.org). However, geopsy has no data administration component with database-like functionality, but allows data export of the results in separate files with detailed header and meta-information for each step in the workflow. In consequence, the user himself is responsible for a complete and consistent storage of the results of the analysis. In the context of the HADU-project, there are conventions for file-names and a standardized structure of the file-tree.

3. Requirements Analysis and Design Decisions for the HADU-System

3.1 Analysis of the Current Situation

After the description of the processing of the raw measurement data within the subprojects an analysis of the current situation concerning methodological approach, data handling and storage, workflow, and interpretation of results shall be undertaken. A view on the data measured and the data processed leads to three main aspects: Firstly, there are huge amounts of data to be stored in highly proprietary format. Se-
condly, there is no fast and direct access to the single data entries necessary. For analysis purposes the access is restricted to always requiring the complete set of data corresponding to one measurement. Thirdly, the measurement data are extended by additional information such as data of measurement, name of the investigator, type and identification of the measurement device, etc. These data are for documentation purposes only and do not have any significant relevance for the data analysis itself.

Concerning the workflow, georadar- and gravimetry investigations are purely sequential. The evaluation workflow for ambient-vibrations-measurements may split, but there are neither cycles nor rejoins of workflow branches after a split. All the steps in the workflow can be interpreted as separately treatable method calls that get their input from one or many input-files and put their output into an output-file with well defined format. The analysis workflows for the different measurement methods are independent to each other and are linked only by the common position of the coordinates of the measurement. Only at the very end the different workflows refer to each other: They all provide an estimation for the depth of the stratigraphic change (in the context of the HADU project the estimated position of the salt diaper) at a given geographical position on surface. Besides this geographical link there is no methodological interrelation between the data during analysis. In contrary, this analysis seems to be highly proprietary for each measurement method and therefore prohibits any mutual integration on the level of intermediate results. But even independently of the possibility of such interdependencies, an interpretation on this level is not intended by the project objectives and hence can be cut for the design of the HADU-architecture.

3.2 Implications for the System Architecture

For the design of the system there are the following implications coming from this analysis:

1. A database system with focus on complex relation scheme, intensive usage of roles and rights and parallel transactions does not seem to be appropriate.
2. The architecture should focus on complete documentation and warranty for reproducibility of the analysis workflow in the subprojects (see details in Wittmann (2006)).
3. In addition, there should be the possibility to inform about the (intermediate) results of the partners by datafile exchange or by access to a comprehensive and integrative visualization of the current data in form of the current state of the model for the salt diaper.
4. The methods needed already exist in form of software packages. They are well adapted to the needs of the investigators. There is no need to integrate them under a common interface, especially because there is methodologically neither interrelation nor need for it.

4. The HADU System Architecture

Figure 2 shows the top level view on the HADU-architecture that will be explained by its two access paths in the next three subsections.

4.1 Access Paths to the HADU-System

Browser-based access for information and documentation purposes
Technically, the access is browser-based by a web-portal. Within the area with overall information on the project there is a non-public area accessible by personalized login with password. In this internal area the data analysis workflow is administrated by a seamless sequence of method
calls together with the necessary meta information such as input- and output- parameters and pre- and post- conditions. Details about the design of this method chain are explained in Wittmann (2006) again.

From the viewpoint of the geological researcher, this part of the HADU-system represents the database component. It is there, where he can deposit results of the analysis, regard the origins of the results, and have a look to data and results of the partners. The access is managed by defined roles and rights, that are initialized as read and write on data concerning to the own sub-project and read-only for others.

Access to analysis functionalities
Besides the classical web access for information purposes the HADU architecture supports the data analysis workflow. There are two possible ways to use the system: In the first case, the user executes his analysis as usual on his own computer with the software installed there and stores the results of the analysis on the central data base server by uploading the corresponding files according to the file tree he is already acquainted to. If the conventions for file format and file extensions are kept (which are already existing and already are accepted by the users) the data analysis can be executed as normal and the resulting data can be stored in the central database without any changes in the workflow of the partners. This concept works for any proprietary software package as well as for GIS projects that implies the storage of the corresponding shape files.

GIS-based access
In a second step of integration the different, more or less independent locations of the measurements have to be integrated for an overall view for the complete region of interest. For this task, the best choice is the view within GIS. For each of the measurement methods a separate layer with the exact position of measurements is created. By selection of one of these locations a certain analysis and visualization program is started as a separate application with its own window. Within this interface all non-writing operations can be executed, the control is assigned from the GIS to the application started. When the application is finished, the user is back in the GIS exactly in the state he left it. Condition for this loose coupling is the existence of a structure in the application activated, that allows combining data for one (or more) locations such as projects, layers, etc. Section 4.4 will go in some details by demonstrating the coupling between GIS and the KoGeo-tool. In contrary to the access to the project data with focus on the method chain this access path focuses on the access by geographical position. Reasonable visualizations and interpre-
ting combinations of the data collected are enabled by the use of the structuring units (projects or comparable semantical units) within the software tools plugged in.

As a subsumption, the HADU-architecture is characterized by the following features:

1. Autarchy for the sub-projects with their already existing software solutions.
2. Data storage on central server.
3. Data access by web-access on file level with rights management.
4. Export/import by metadata format to software tools already established.
5. Access to data by its position in the analysis workflow as well as by geographical position.

Two examples shall demonstrate the HADU-approach in some more detail:

### 4.2 Example 1: Workflow-based Access

Figure 1 shows the analysis workflow for ambient vibrations measurements; figure 3 the schematical structure of the method chain, consisting of method- and data-nodes. The workflow is mainly realized interactively within the tool geopsy. Intermediate data can be exported in file format together with the relating metadata. This mirrors exactly the preconditions necessary for building a close method chain. If the analysis is interrupted by the user, it can by continued at any position after an import of the related intermediate data file.

Therefore, the basic structuring element of the method chain consists of the method applied together with the file with the results and the metainfo for the method call. Such a basic element can be attached arbitrarily at any data node of an already existing method chain, thus defining a simple prolongation of the chain or a branching of the workflow. The resulting bipartite tree allows an easy access to its elements in full analogy to a filetree. A trivial user interface for access is described in Wittmann (2006).

![Figure 3: the common structure of the method chain](image)
4.3 Example 2: Geographical-Based Access

The second access variant is depicted in figure 4. The overall structure is a map within GIS with the locations of measurements in distinguished layers for each measuring method. By selection of a position in the map, the user gets a list of existing projects for the selected measurement location. By organising the data according to a M: N mapping between location and project it becomes possible to realise the distinguished access to the designated data as well as a comparative presentation of different measurement locations within one project. After selection, the corresponding project is started by calling the related tool. The control is given from GIS to this tool, enabling the parallel call of different tools at the same time for comprehensive visualizations in separate windows. Closing those windows implies giving back control to the GIS.

With this approach both of the use cases necessary can be realised. Firstly, the access for information purposes and for presentation of results a the common GIS user interface extending GIS to 3D visualization by using loose coupled special tools. Secondly, a full access to the methods needed for data analysis offered by the tools accessible by the GIS entry points. The difference between these two alternative access methods lies in the rights given for the call. In the first case read-only rights are sufficient, the second case demands writing results of an analysis into the method chain.

4.4 Integration of Visualization

With the functionality described so far, the demands for unified data storage and widely autarkic analysis processes are satisfied. What is missing is an integration of the data to a joined 3D visualization of the region of interest. This nor can be done by GIS neither by one of the tools used. A look on the end nodes of the method chains of analysis shows for all the different approaches a quite similar final data element: There is always an estimation for the depth of the interesting changes between two ground-layers. These

Figure 4: access path to analysis methods
depth-values might be collected in a common matrix for underground of the region of interest. The matrix will be sparsely filled what forbids a direct visualization of its content. There will be special interpolation methods necessary to obtain a nearly continuous model on the one hand, but there should also be a measure for accuracy of the measurements, analysis and interpolation. How to solve this visualization task shall not be discussed here. However, in the context of this paper the integration into the analysis system is of interest: First step will be an additional method in the method chain that reads the estimations for the different depths from the different measurement approaches and transfers these values into the overall matrix. For each position in 3D-space a triple of measurement-method, depth-value, and accuracy are stored. Doing so, the over-all matrix merges the results from the different analysis workflows. Afterwards, an access to the matrix will be read-only, a fact that allows the exception to the rule for concatenation and branching of the method chain of section 4.3.

On base of this matrix any 3D-visualization can be coupled to the HADU-system in full analogy to the integration of the analysis tools. The call is from the known GIS selection, a separate window opens for free interaction with the 3D-visualization provided by the visualization tool. This has to interpolate the values of the matrix and to visualize the attributed accuracy by attributes of the curve (e.g. by increasing transparency for less accuracy). For the realization a coupling to MatLab is intended, to benefit from its broad spectrum of interpolation methods as well as its functionality in focusing, rotating, and printing the 3D-model.

5. State of the project and summary
The system analysis phase for the HADU-software system is completed, the concept is accepted by the partners. The coupling is implemented exemplarily for the tool KoGeo. Currently the analysis workflows are specified, the discussion concentrates on the metadata necessary. However, by the modularity of the concept, changes in metadata will cause only local changes in the database and do not imply serious changes of the data model.

The experiences so far approve that the system structure is strong enough to integrate all the processes of measuring and data analysis on the one hand providing on the other hand compatibility to the workflow and the tools the projects partners trust in and are acquainted to. Thus, the system gains acceptance as well as it provides the postulated data integration without too much overhead in data administration for the partners.

Bibliography
Geotechnologien (21005) Science Report No. 6, BMBF 2005, pp.124-140
http://www.geophysics.zmaw.de Homepage des Geologisch-Paläontologischen Institutes, Universität Hamburg, Arbeitsgruppe Prof. C.-D. Reuther
http://www.kogeo.de/ Homepage des Kogeo Seismic Toolkit
http://gmt.soest.hawaii.edu/ Homepage des Generic Mapping Tools (GMT)
http://www.geo.uni-potsdam.de Homepage des Instituts für Geowissenschaften, Universität Potsdam, Arbeitsgruppe Prof. F. Scherbaum
http://www.geopsy.org/ Homepage des Sesarray Projekt