GEORD – An Open Data Warehouse Architecture for Environmental Data

Rainer Haselberger¹, Norbert Ondra²

Abstract
For the sake of quick implementation, information systems are often tailored towards specific tools for editing, processing and querying of data. They normally have a high structural dependency on the tools used to build them, which usually makes changes to other tools and technologies both difficult and costly.

As product cycles in software industry are constantly getting shorter, we have to face the rapid aging of existing tools, while the data involved still show growing longevity. Thus, a stable basement in terms of data, which supports the deployment of various different technologies at the interface level while preserving the internal structures, is of utmost importance.

This is even more true in the context of geographical information systems (GIS), which are used for representing spatial contexts of environmental data, such as the sampling of pollution or noise and the water quality of wells, lakes, and rivers.

In order to meet the requirements to gather, manage, and exploit various environmental data with complex structures, huge volumes, and geo-referencing, the GEORD architecture was developed in 1992 to suit the purposes of the Vienna City Administration and has been continuously enhanced since then.

1. Overview

The concept of GEORD is a geographically enhanced, object-relational, distributed data warehouse, designed as a framework to tie together different application data from database servers distributed in a network.

The main goals to be achieved are:

- enabling use of data without requiring detailed knowledge of the data model

1 Vienna City Administration, Municipal Department 14 – EDP, Division MV1, Rathausstraße 1, A-1082 Wien, Phone: +43(1)4000-91981, FAX: +43(1)4000-99-91981, Email: har@adv.magwien.gv.at

2 Vienna City Administration, Municipal Department 14 – EDP, Division MV1, Rathausstraße 1, A-1082 Wien, Phone: +43(1)4000-91213, FAX: +43(1)4000-99-91213, Email: ond@adv.magwien.gv.at
facilitating the development of applications by using uniform component libraries
minimizing data redundancies
combining unrelated application data by using spatial context

Main benefits of deploying the GEORD framework are:

- establishing a stable and uniform data basement for the future
- usage of widely different base data across organisational units
- flexibility in choosing a suitable application architecture built on XML interfaces
- efficient realisation of future user needs by re-using basic components
- scalability – stepwise migration of external data collections into an object-relational database with spatial context
- value added to external data by relating to central base data stored in the data warehouse
- an integrated uniform address service helps to establish spatial references among data from different applications

2. Application model

The core of GEORD primarily consists of:

- meta data descriptions as an integral part of the application data itself
- application data (both conventional and GIS data) which are subject to a uniform way of data access
- tools for managing the meta data descriptions
- applications built on top of base functions interpret the meta data enabling the use of a uniform data access method

The following diagram shows the logical structure of GEORD:
Addresses are basic data used in almost every application programme of the City Administration and have a long-lasting nature. Therefore, the “addresses” (of Vienna) constitute an integral part of the GEORD basic data, accessible via the same methods as other application data in GEORD.

One reason for their integral role is that addresses are often necessary to establish the spatial context of application data. In that respect, addresses support the combination of data from different application domains by means of the spatial context.

2.1 Data structures

All data is stored in standard relational databases (currently Oracle) making up the GEORD data warehouse. There are two different kinds of data, which are both kept in ordinary tables of the same database(s):

- meta data structures describing the actual structure of the application data
- the application data itself

Each kind of object (= object type) belonging to the subject matter of the environmental information system (EIS) is represented by a “prime table” storing the core data of that kind of object. Additional data associated with the object, such as multiple attributes and relationships to other object types, is stored in related tables. The management of historical data is supported as well.

Commonalities among objects types can be represented in inheritance hierarchies. In the end, any table, attribute, key, or index needed by environmental data is described by means of meta structures in GEORD.
In addition, data with spatial context have descriptions for the graphical context used. Each such kind of object may have a particular kind of spatial reference, a link to a theme of a coverage or shape file from the ESRI tool "ArcView".

2.2 Meta data maintenance

GEORD provides a special tool with GUI to administer meta data structures. In order to introduce a new object type into the data warehouse, you just have to define the necessary entries in the meta data structures. As a consequence, all existing software components, queries, reports, programmes, etc. can be used to work on the newly defined data.

This is a very powerful and yet simple way of making existing software available to new data.

2.3 Editing and querying application data

GEORD provides a generic way of designing user dialogues based on the meta description of user interface forms.

The meta descriptions are then interpreted by a generic client programme (implemented in Visual Basic) which brings the proper user interface forms, GIS maps as graphical layers, objects connected to graphical layers to the users’ screen, performing all data operations as defined.

This generic client supports all types of relational data: prime tables used for object types, multiple attributes, sub-type objects (inheritance) and relations between object types.

2.4 Graphical presentation of data with spatial context

All spatial information related to concrete objects is rendered by means of modules based on the ESRI tool "Map Objects". The geographical context of a georeferenced object, as defined in the meta data structures, gets interpreted in order to place the proper symbols in the right positions on the thematic map which is used as graphical background layer.

3. Applications

The concept GEORD is realised in several applications with environmental data which all make use of the structures and methods described above.
3.1 WWDBS – Database system for water management and water resources in Vienna

The WWDBS application is implemented as a Visual Basic client programme with database access via ODBC. All data access routines for an object type or relationships with other object types first interpret the meta data definitions to find out which tables, attributes, indexes, etc. are involved in performing the actual data access. The main functions of WWDBS are:

1. editing and maintaining the water resource data of Vienna ("Wasserbuch")
2. defining and performing all relevant legal transactions of the Federal Province of Vienna based on the Austrian water management act
3. which in turn are the base data for legal decisions based on that law.

Most relevant object types, such as stagnant or flowing waters, wells, etc. are georeferenced data connected to a graphical representation in a GIS layer.

The VB-client is designed in an extremely generic way so that it is apt to be used in other application domains with similar requirements and comparable complexity.

Adapting the meta data descriptions in GEORD to the new application domain by using the GEORD tools is a necessary prerequisite.

3.2 WUIS2000 – Environmental data catalogue of Vienna

The WUIS2000 application, again a Visual Basic programme, interprets the meta data descriptions of the environmental data sources of Vienna – all stored in the GEORD data warehouse – and generates static HTML pages to be published at "wien.at", the City of Vienna’s website.

The information shown contains only the descriptions of environmental data, furthermore hyperlinks to other environmental information or real data as published in "wien.at" or in other WWW domains. In addition, cartographic representations help to visualise the spatial context of environmental data sources.

For data editing and maintenance the generic client programme described under point 4.1 is used.

3.3 FIS – Environmental information system

FIS is a prototypical server application which provides an easy way of publishing complex information - consisting of textual data, GIS data, and table data – in the Web. Its current usage is confined to the Intranet.
The HTML pages to be presented to the web browser are derived from server pages with embedded XML tags referring to templates containing format information and references to GIS maps and tabular data.

The following tools are made available by the FIS server for usage in HTML pages:

1. **TreeView**
   - graphical visualisation of a directory tree to structure the information base.

2. **TableView**
   - visualisation of database tables or GIS attribute tables. The data to be shown on the HTML page can be restricted by applying filters and the choice of attributes, data can also be sorted.

3. **MapView**
   - cartographic representation of geographical information based on geographical data sources in the formats of ESRI (coverages, shapefiles, ...). The user can modify the area of interest by panning and zooming and via addresses. Graphical layers can also be activated and deactivated as needed.

Currently, this server application is still a prototype which is, however, used by several Municipal Departments for production work. To provide for quick implementation, FIS was built on MS IE 5+ as browser technology, which is the internal standard.

The next version will aim at wide-spread use among the departments.

### 3.4 Future plans

The Map View component will be callable separately from FIS to be used with other applications, hence another GEORD component.

There will also, be end-user tools to design HTML pages to be used in a simple way with the 3 existing tools mentioned above.

An open interface based on XML to access application data in GEORD is under development. GEORD server data can then be accessed through either conventional “fat clients” (such as Visual basic application) or through “thin clients” (such as Web forms). The software developer then has the opportunity to choose the most suitable design for the requirements of a given application.

Actually, both technologies will use the same methods at the data access layer and business layer, thus having a uniform way of accessing the application data. In particular, changing the client chosen should not require any changes to the data access layer.
4. **Data migration**

A major strength of GEORD is to provide an easy way of migrating external data into the data warehouse. Several degrees of integration are possible depending on how tightly data is connected to the data warehouse.

The following 4 steps are necessary to obtain full data integration:

1. **incorporating the application data into the server database:**
   Application front-ends such as MS Access will continue to work by means of linked tables and can therefore still be used for the operational part of data editing and maintenance.

2. **adding meta data to describe the application data:**
   Tools with graphical user interfaces are available to support that task. The application data are now accessible to existing querying and reporting tools or other applications based on the meta data descriptions. Data usage is read-only, all operational tasks, such as editing and maintenance continue to be carried out with the application front-ends.

3. **tying together application data with spatial context and existing GIS data:**
   The connection between geo-referenced application data and GIS data is established via coordinates as part of the application data. Existing GEORD front-end tools, such as the generic client for WWDBS, can now be deployed for operational tasks on the application data. Likewise, applications with specific requirements may be developed based on the GEORD architecture and interfaces. Existing front-end applications can still be used for data maintenance.
4. **abandoning old applications:**

Existing GEORD tools and newly developed applications fully replace old application programmes. The migration is now complete.

The 4 steps are loosely coupled, i.e. you can stop the migration process after every completed stage. In order to obtain full benefit, however, you have to pass beyond stage 2.

5. **Practical experience**

The GEORD concept offers substantial synergies among applications with very similar requirements with respect to

- data editing and maintenance,
- combining data with other geo-referenced data
- visualising geo-referenced application data
- and performing evaluations of such data.

In fact, there is a huge potential in every city administration dealing with geographically related data.

It is a pity that those synergies are rarely exploited across organisational boundaries. Instead, special applications are developed tailored to the requirements and the budget of a single organisational unit. The potential to bring together data from different application domains thus remains unused.

This is also a major reason why there is no external product GEORD yet. Its current usage is confined to the City Administration. The goal is to promote the wide-spread “spatial” deployment of GEORD, within the Vienna City Administration, with no commercial interests in the first place.

6. **Summary**

GEORD is an architecture which facilitates the integration of external data sources and enables the combination with geo-referenced data and other central basic data.

Flexible open interfaces provide for an easy replaceability of tools for data access, while preserving the longevity and continuity of the business data itself, representing the most valuable piece of goods of any organisation in the long run.