

Building Common Infrastructure for Dissemination of Real Estate Data in Slovenia

Jože SENEGAČNIK

(permanent associate of S&T Hermes-Plus d.d., Ljubljana, Slovenia, email: joze.senegacnik@snt.si)

1 ABSTRACT

S&T Hermes-Plus d.d., Slovenia, together with other partners is currently working on several ongoing projects concerning data dissemination. All projects have similar objectives and are encountering similar problems. This paper describes a logical model of the web based system for data dissemination used to support the dissemination process in the Slovenian government. This logical model was prepared in a preceding theoretical project and is a good entry point in preparation of physical implementation of data dissemination process in the surveying and mapping authority of the republic of Slovenia.

2 LOGICAL MODEL OF THE DISSEMINATION PROCESS

The dissemination model is based on web services accessed through a unified portal. The logical concept contains one three levels:

The first level contains the entry point to all services. The second layer consists of one or several dissemination databases that provide all necessary data. The third level contains background system services which support security and system administration and support (like data transformation, recording of all data orders, payment system, etc....)

The end users perceive the dissemination system as one entry point – as a Spatial Portal. This portal contains links to different services of the portal (ordering, data browsing, data acquiring, general and special information, etc...). The same portal is used also for the internal governmental users. Figure 1 displays the logical structure of the spatial portal.

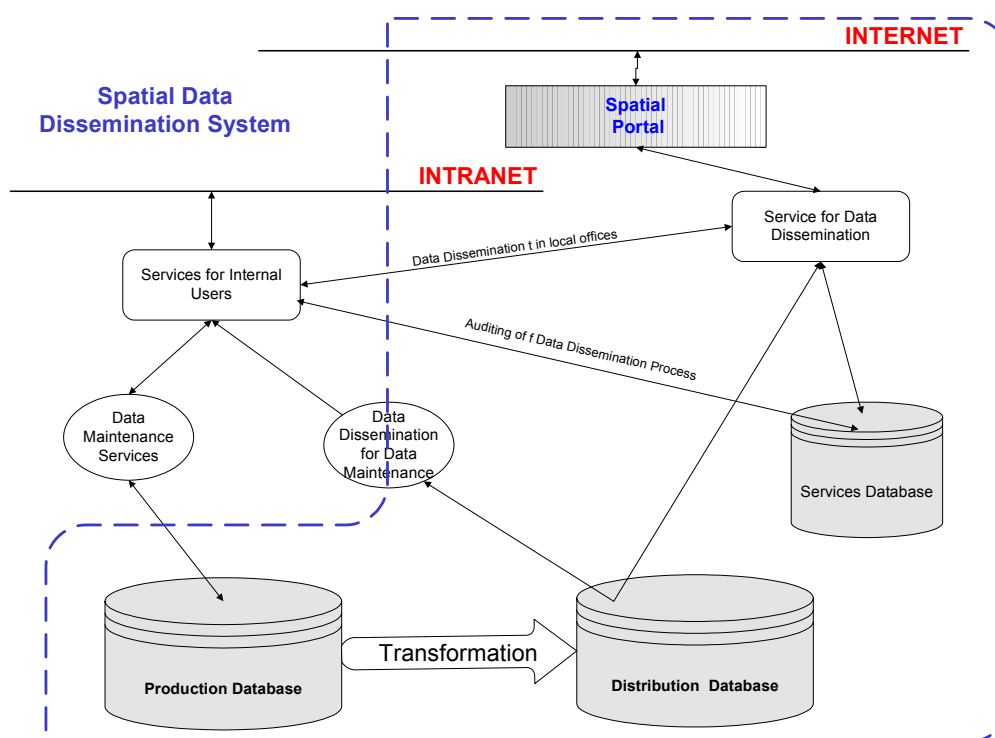


Figure 7: Logical Model for Dissemination of Spatial Data

3 THE KEY PROCESSES AND THEIR RELATIONS

From the above figure we can see the logical model of our system with the main data flows and processes. On the highest level the logical model is presented with the data flow diagram of the Spatial Portal that consists of:

- global system processes
- data warehouses
- data flows among data warehouses and processes

Data sources in the dissemination system are the following:

1. *Dissemination database* – is the main data source of the dissemination system
2. *Production databases* – these databases are not a part of the dissemination system but are the main source of the data flow for new or updated data that are through transformation process entered in the dissemination system.
3. *The metadata database* - in this database we keep the metadata about our data records.

4. *The overview layers database* – in this database we keep additional metadata attributes at the level of data maintenance, all planned data maintenance or data acquisition activities and also the constraints about the minimum amount of data that can be acquired.
5. *The Statistics or Services database* – keeps all necessary data for overview and managing the dissemination process.
6. *The system database* – contains the system setup data and all other data used for controlling the dissemination process.
7. *The users database* – contains the data about the registered users, their attributes and rights that influence the data dissemination process and also the price of the service.
8. *The price list database* – keeps the records for different services and data types.

Processes in the Spatial Portal are the following:

1. *Registration* - This process provides the authentication of portal users. All new users that want to access services with restricted access must first register and they are automatically redirected to the registration process.
2. *Data dissemination processes* - The main part of the system are the following processes for data dissemination:
 - data browsing services for browsing the spatial data, metadata, overview layers etc...
 - data ordering services,
 - special services
3. The non-registered users can access some of these services as well.
4. *Services for the internal users* - The Spatial Portal also provides the services for activities that are not directly related with the dissemination process but are part of the internal procedures of the data providers.
5. *Common e-governmental services* - The Spatial Portal is connected to the common e-governmental services in the process of:
 - User authentication
 - Electronic data interchange of ordered data with the digital signature
 - Billing and Payment system - This is a general service used for preparing the necessary documents for the billing process.
6. *Statistics* - The system keeps records about all necessary statistics about all accesses to the system and data requests.
7. *Auditing system* - The purpose of the auditing system is to keep the track off all accesses to the data especially to the data with defined attribute of privacy. According to the Slovenian legislation the personal data has a high degree of privacy and thus special procedures of providing privacy of this data are required.
8. *Data transformation* - Data transformation process is used for transformation of the production data into the dissemination data and properly updating all the overview layers and metadata database.
9. *Help system* - The purpose of the help system is in providing context and technical help to the end users.

Implementation of the data dissemination system

The whole dissemination system is implemented in web technology. The database system is ORACLE RDBMS that is a de-facto standard in the Slovenian governmental environment. There is a significant difference between the production and distribution data organization which can be expressed with the following:

The organization of *production database* supports:

- The normalized structure of all data types that is easy to maintain (update); each attribute appears only once; the de-normalized structure is an exception in the case of serious performance problems.
- All data changes are recorded in the history of changes. The audit record contains also the information by whom the change was performed.
- All business rules how to maintain data are implemented at the database level (triggers, database procedures) or are provided by the application code.
- In the application data browsing is supported for data maintenance function and not for the data dissemination process.

The organization of *dissemination database* supports:

- The data model is optimized for querying data regardless the point in time
- The number of business rules implemented in dissemination process is smaller than in production database.
- The data organization can be also optimized for supporting certain standard formats.
- For all personal data with certain degree of privacy special rules are in order. The user rights to access data are checked differently than in production systems.
- The process of de-normalization of production data is one the key success factors for successful dissemination model.

4 IMPLEMENTATION

The system was implemented in past two years in order to replicate the production data from Survey and Mapping Authority to the common distribution environment available for all governmental institutions and all other users. As already stated the distribution environment is optimized for data dissemination with the following goals:

- Good response time
- Data structure is optimized for data dissemination
- Two data sets: the current data (used mostly) and historical data

The data structures, which are used in the production environment, are de-normalized and optimized for the dissemination as we already stated. The history of changes on production database is de-normalized and kept differently in the distribution environment to enable simple and efficient querying of historical records.

Because the distribution system is de-normalized the regular procedures available for data replication are not suitable any more. Therefore a complete set of replication procedures was developed which in the first step replicate original data structures from production environment and then in the second step build de-normalized data for the current and historical state of the data. Currently there is only one refresh of the data performed per day during the nightly batch processing window. Although this is enough to fulfill almost all requirements a refresh on demand is also available and can be performed when required. Most of such urgent refreshing cases evolve in the time before the elections when the population register is intensively updated and the most recent data from Spatial Units Register are required to prepare the required material for local elections committees.

The security issues require that only the owner of the data has the right to update the distribution database. All procedures, which were implemented to support web services for data retrieval, are implemented in a separate schema, which has only read rights on the dissemination database.

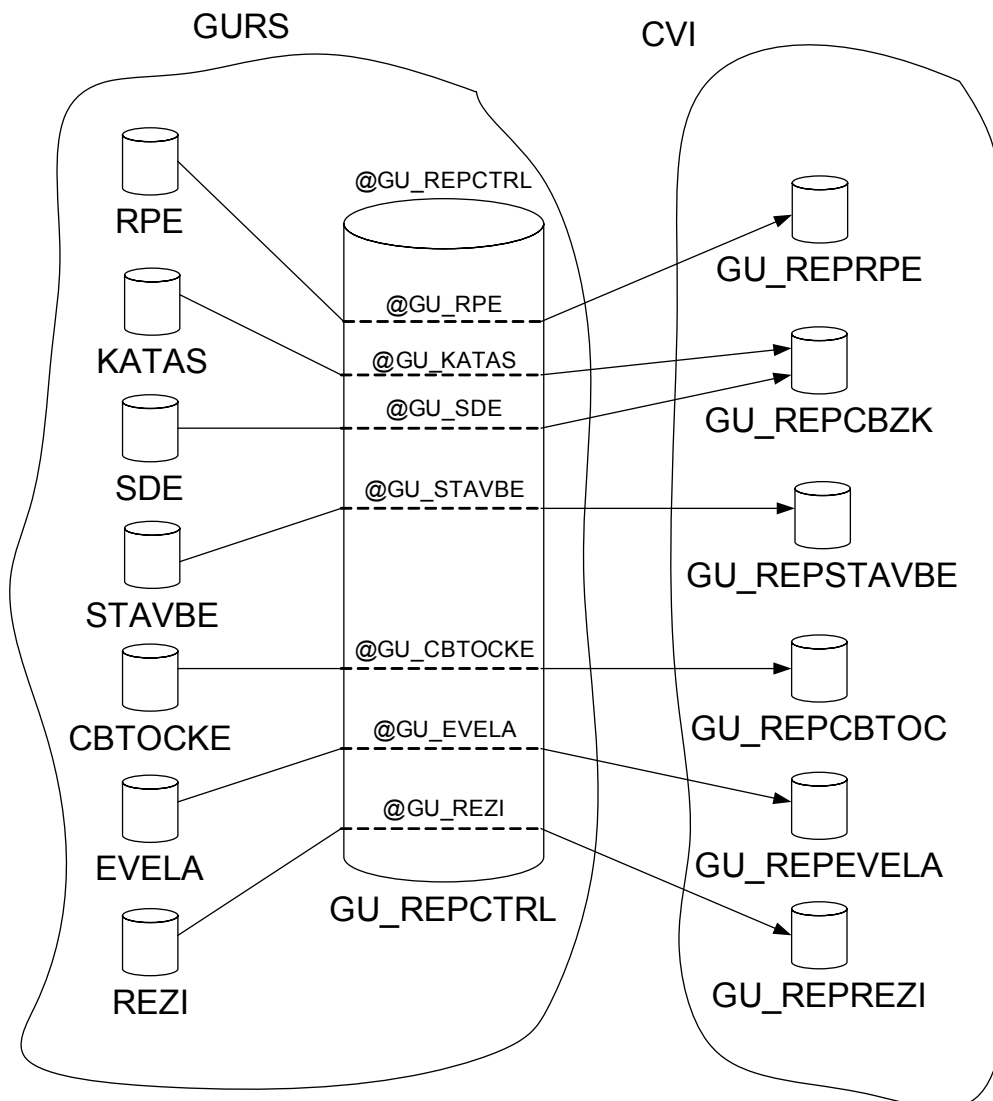


Figure 8: The logical model of replicated data - schemas

A special effort was made to fulfill the requirements to maintain and protect personal data according to the Slovenian legislation.

4.1 Replication of attribute data

All tables with the attribute data are replicated via Oracle replication mechanism using snapshots which enable incremental replication.

The graphical data in the production environment are stored in different formats due to historical reasons. Some of the systems were developed more than 10 years ago when there were no standards available yet and the developers used their proprietary data format.

During the planning phase for the common distribution environment a very important decision was made – to use Oracle spatial as a standard format for all graphical data. Therefore it was necessary to prepare the conversion routines to convert a variety of graphical data to Oracle spatial format.

Application System	Data Format
Registry of Spatial Units	Aster proprietary format
Registry of Geographical Names	Aster proprietary format
Central database of Geodetic Points	Aster proprietary format
Central database of Graphical Cadastre	ArcSDE
Central database of Buildings	ArcSDE

4.2 Existing graphical data formats

Aster d.o.o. in years 1993 and 1994 developed the Registry of Spatial Units which for the first time stored all graphical data in a relational database (Oracle). At that time there was no Oracle Spatial available and therefore a proprietary data format was developed that was capable to store point, line and polygon objects in classical relational tables. The graphical data model was highly optimized to minimize the required space for keeping the track of changes on different graphical objects. The same data structure was used later in several projects in Survey and Mapping Authority of the Republic of Slovenia.

The other data format used in some other application systems was ESRI’s Arc SDE format.

The implementation of new distribution environment required a full implementation of conversion modules for the above data.

5 CONCLUSION

During last years Slovenia made a lot of efforts to provide more and more e-governmental services to the citizens and other subjects like companies. One of the fields where Slovenia was quite successful was also the distribution of spatial data.

The dissemination database is operational for last two years and is used for the distribution of the graphical and attribute data in the governmental environment. During the last months the whole hardware infrastructure both in production and distribution environment were upgraded. At the same time the database software was upgraded to Oracle10g as well.

As new GIS data will be collected and maintained at the state level more and more data will be available to all kind of users.

6 BIBLIOGRAPHY:

10. The Model and Implementation Plan of the Dissemination of the Geodetic Data - University of Maribor - Faculty for the electronics, Igea d.o.o. and Aster d.o.o, Ljubljana, 1991
11. Senegacnik, Joze: LOGICAL DATA MODEL FOR DISSEMINATION OF SPATIAL DATA, UN-ECE-WPLA-Workshop Vienna 2002