Data and Data Warehousing for Management of Small Region – from the Czech Republic Point of View

STANISLAVA SIMONOVA, JITKA KOMARKOVA
Institute of System Engineering and Informatics
Faculty of Economics and Administration
University of Pardubice
Studentská 84, 532 10 Pardubice,
CZECH REPUBLIC

Abstract: The term Public Administration is understood as administration of public affairs demonstrating the executive authorities of government and self-government that are applied within the framework of the valid legislation. Recently, the administration of region has become significantly connected with the problem of creation of efficient information environment supporting this administration, whether the region is meant in the dimensions of a district or other territory. The acquisition of outputs for support of decision-making requires either specialists in the field of information and communication technology (ICT) or a suitably prepared information environment. However, in a current municipality the information environment is oriented to expeditious data, i.e. it involves transactional information systems based on database and/or geographical systems. Recruitment of ICT specialists represents a difficult problem for most municipalities because most of them are rather small in the Czech Republic – 59% of Czech municipalities have up to 499 inhabitants, and 80% of municipalities fall within the category of municipalities having up to 999 inhabitants. In addition to it, the existing database is scattered because there is no coordination among institutions of public administration. This problem can be dealt with by applying the technology of data warehousing. Regional data warehouse can be realized on different levels of elaboration and efficiency. The selection of a suitable variant of data warehouse architecture must reflect current needs and possibilities of regional institutions.

Key-words: data sources, regional data warehouses, geographic information systems enquiry, regional management

1. Introduction

In the Czech Republic a reform of public administration is going on whose impacts are directed to strengthening of democracy in CR and also, of course, to harmonisation of our legislation environment with that of other countries of EU [1]. These reform trends require, inter alia, implementation or more efficient application of information and communication technology in the field of public administration in such a way as to fulfil the requirements such as increasing efficiency of activities of public administration organs, providing a high-quality information platform for well-informed decision-making in the area of economical use of municipal/regional/state property, simplification of contacts between the public and public administration at the municipal/regional/state level etc.

2. Administration of Municipalities and Micro-regions

The reform of public administration involved restructuring of territorial zoning with the aim of its harmonisation with the NUTS system (La nomenclature des unités territoriales statistiques). The individual levels of NUTS represent certain magnitude groups of territorial units (depending on the number of inhabitants and surface area of the territorial unit). From the point of view of regional administration important NUTS levels are NUTS 5, representing territoria of individual municipalities irrespective of their magnitude, and NUTS 4 – microregions, which are especially important from the standpoint of regional cooperation of municipalities. However, the changes in competencies and

authority of territorial units leads to a sertain ambiguity in the sense of data sources.

The database in terms of basic registers and other sets of public administration is formed by the data for basic registration (civil registration, economic and subject data files etc.) and geogrpgical data (map layers of territoria, space identification of territorial elements), which are composed of graphical data, non-graphical (alphanumerical) attributes, and relation data. Hence, this means an interconnection of the "basic" information system and the geographical information system (Fig. 1).
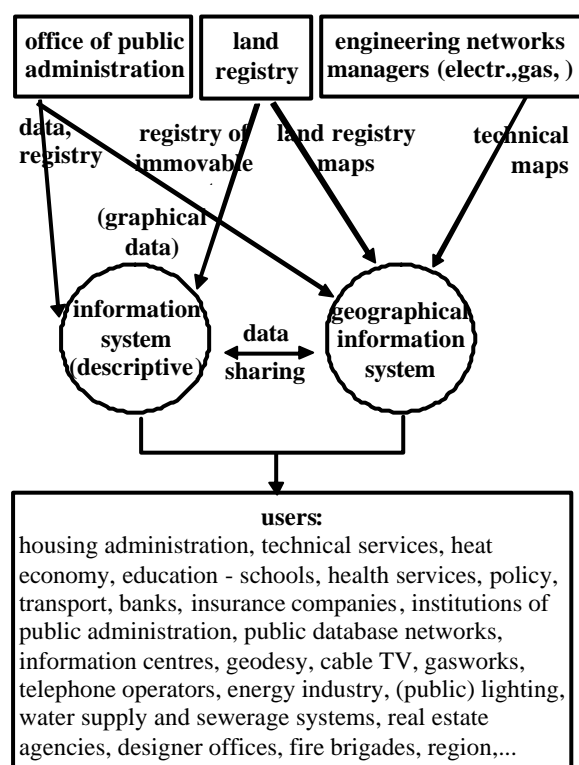

Fig.1: Scheme of information system in the area of public administration

The data about regions used to be collected by former District Authorities; however, these institutions were abolished in accordance with the reform of public administration. Other institutions collecting data about regions include representations of Statistical Office, Tax Office, Employment Agency and individual magistrates or local authorities. However, the data assembled within the above-mentioned institutions are not coordinated between the individual institutions (e.g. between Statistical Office and Employment Agency) and, what is even serious, they are not coordinated within a single institution (e.g.

individual Employment Agencies collect data that contain various amounts of details).

2.1 Characteristics of Information Milieu of Public Administration from the Czech Republic Point of View

The information milieu of public administration in CR exhibits certain specifics. These involve the magnitude of the regions administered from the standpoint of number of inhabitants; moreover, it is necessary to take into account the typical user of data sources of public administration. The structure of Czech municipalities according to numbers of inhabitants shows that 27% of municipalities have up to 199 inhabitants and 33% of municipalities have populations between 200 and 499 (Table 1) [2].

Table 1: Structure of Czech municipalities according to numbers of inhabitants

Number of inhabitants	No.of municipalities	Municipalities, %
up to 199	1 659	26.53%
from 200 up to 499	2 037	32.58%
from 500 up to 999	1 275	20.39%
from 1 000 up to 1 999	657	10.51%
from 2 000 up to 4 999	365	5.84%
from 5 000 up to 9 999	130	2.08%
from 10 000 up to 19 999	68	1.09%
from 20 000 up to 49 999	41	0.66%
from 50 000 up to 99 999	17	0.27%
over 100 000	4	0.06%
Total	6 253	100%

Therefore, many municipalities do not fulfil the conditions for employing their own ICT specialists, due to both their low budgets and the absence of the said specialists from the given (small) region.

Another problem consists in the structure of users of information milieu of public administration (Fig. 2)

Users of information systems are represented by public administration personnel, staff of other organizations and

especially citizens, i.e. enquiries to data sources are put with different accuracy, different knowledge in information technologies and different knowledge in public administration problems [3] [4].
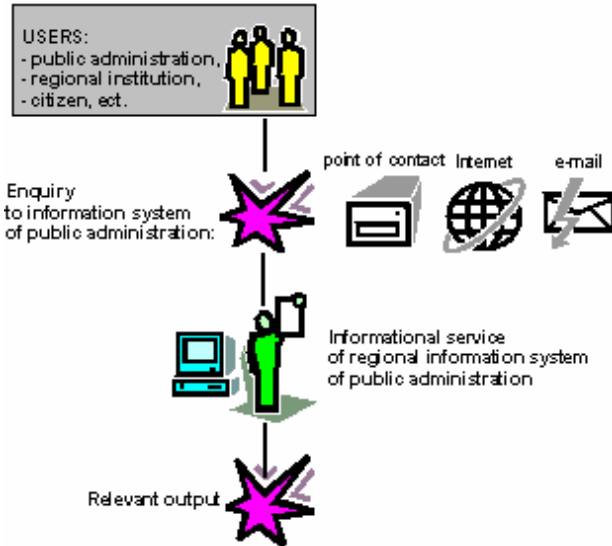

Fig.2: Specific group of users of information systems in public administration.

Another big problem is the above-mentioned non-coordination with regard to the database across the public administration (structure of the indexes registered, degree of comprehensiveness etc.).

3. Technology of Data Warehousing for Public Administration

Information systems within public administration are most frequently realized by database software. Above all, these are so-called transaction database systems, which are designed for work with operational data of the organization [5]. Transaction systems work with actual operational data; they are, however, less suitable for analyses in time relations, more complex enquiries etc. Another approach to data sources is brought in by data warehouse technologies [6], when data are drawn from heterogeneous sources of transaction applications and are stored for a certain period of time, so that they can be used for comparisons, analyses and predictions.

3.1 Modelling
The modelling approaches of database technology are based on principles of entity-relation modelling or multi-dimensional

modelling. The relation concept results in designing a denormalised database with a high degree of aggregated data.
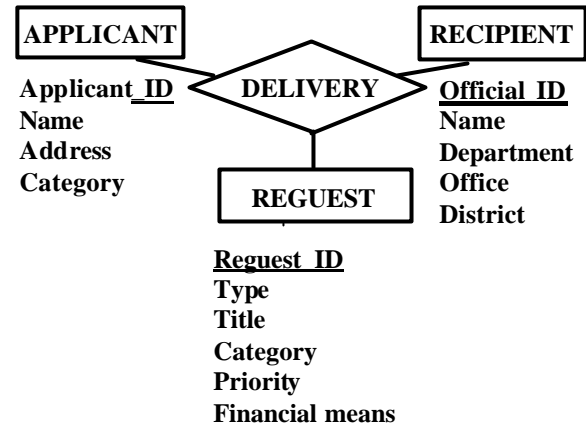

Fig.3: Delivery of application for subsidy

The dimensional method of data modelling results in denormalised database projects with relatively high degree of data aggregation (a normalised / standardised database would not be suitable from the point of view of the complexity of SQL enquiries for analysis).
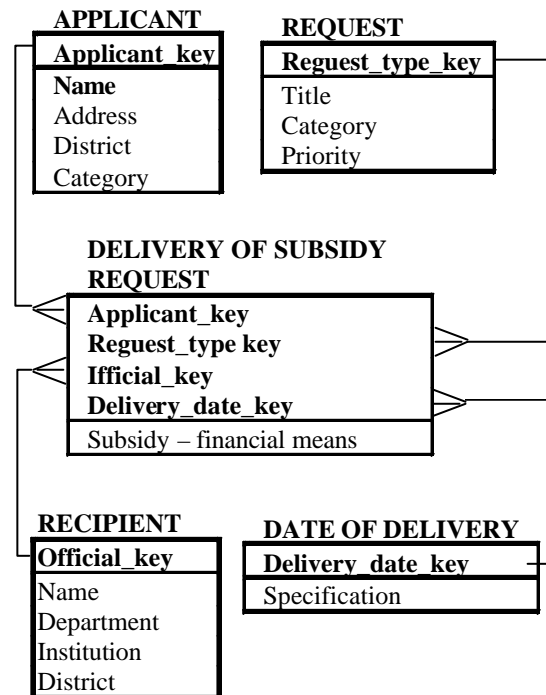

Fig.4: Delivery of application for subsidy: star-shaped scheme of tables

The approaches of dimensional modelling are grounded on constructions such as the monitored values (facts), dimensions, and attributes. The model obtained is usually

represented in the form of star-shaped scheme of tables or in the form of multi-dimensional cube.

Let us consider a case of registered application for dotation (Fig. 3). We can apply dimensional modelling represented by the star-shaped shceme of tables (Fig. 4).

A number of approaches can be adopted in the modelling, and it is a question which of them will be suitable for construction of data warehousing of a regional institution of public administration. The data in public administration are often of such character that they are manageable by the relation modelling approach rather than by the dimensional approach in the sense of, e.g., multi-dimensional cube (the values of dimensions are then unoccupied). The entity-relation model can seem more suitable also because of its variability for both the data marty and data warehouse. Therefore, the decision about the modelling principle must start from the data themselves which should be avalable within the regional data warehouse and/or from the type of enquiries that will be addressed to the regional data warehouse.

3.2 Architecture

An obvious premise of users' access to regional data warehouse should be the access via a network [7], most often through Internet (Fig. 5). This possibility should also be offered by the so-called contact points (kiosks) [1], which could differentiate between a current approach of every user and a registered approach.

Regarding the demands for data within public administration, we can recognise the following requirements that an architecture of regional data warehouse should meet: first - regional data warehouse should work with data extracted from operational applications of public administrations, i.e. without any direct access to the operational databases; second - the data must be transformed and purified, i.e. potential redundancies must be removed, the data must be consistent, aggregated data will be created etc.

The demands on regional data warehouse specified in the above way can be realised by constructing the data marts of by the architecture of so-called enterprise data warehouse.
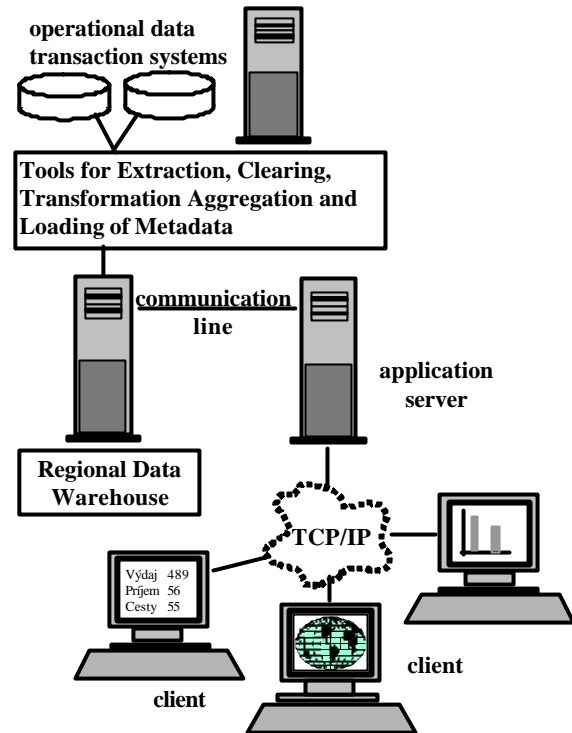

Figure 5: Regional data warehouse accessible via network

The architecture of independent data marts works with various levels of both detailed data and aggregated data formed on the basis of extraction, purification, and transformation of operational data, which is positively reflected in their consistency (Fig. 6).
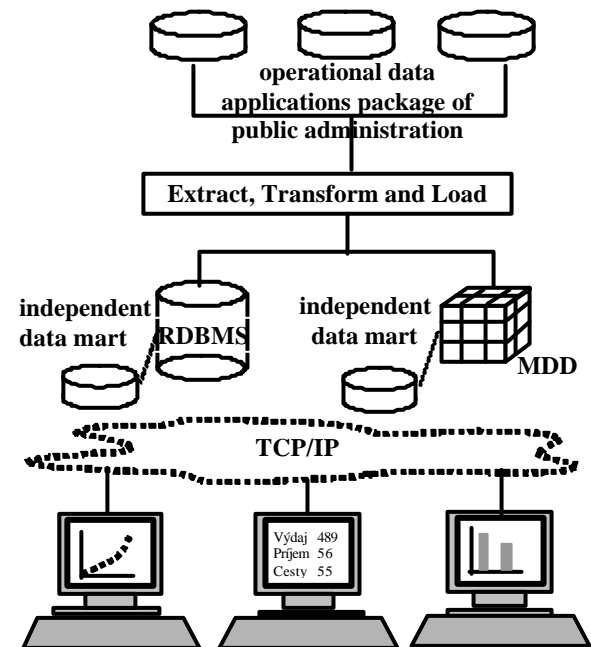

Figure 6: Architecture of Regional Data Warehouse

Data marts react to actual information needs, quick responses being obtainable within a framework of a defined set of enquiries

thanks to the dimensional approach to data modelling.

The ideal, most complex and most sophisticated variant of construction of regional data warehouse would consist in the architecture of so-called enterprise data warehouse. However, this variant is exacting technologically, organisationally as well as economically, i.e. a variant requiring considerable experience in the field of data warehouse technology.

However, the regional "informational" situation is not mature enough for an enterprise data warehouse (demanding sophisticated construction and technology, and high financial investments) to be constructed; nevertheless, this is not inevitably necessary for the above-specified obtaining of high-quality outputs. An acceptable architecture of regional data warehouse can be seen in a certain combination of both ways of solution, i.e. the architecture of independent data marts without constructing the database of central metadata but with application of coordination across the data marts within public administration inclusive of unification of professional terminologies of the individual areas of interest.

4. Conclusion

The chosen architecture of regional data warehouse is that of independent data marts without constructing the database of central metadata. Each of the independent data marts has its own database structure, the implementation of a data mart being easy from organisational standpoint. The data marts are owned and run by certain organisation units. The benefits from the investments into a data mart of regional institution can be seen the same as those from implementation / innovation of ICT within information system of any organisation, i.e. decrease in labouriousness, saving of overhead expenses, more efficient utilisation of database etc. However, when analysing the results of investments, one should also take into account the specific position of institutions of public administration with regard to citizens, i.e. increased support for the aims of the institution of public administration, support for more efficient administration of the respective territory, more efficient services for citizens by improving awarenes of the public etc.

References:

[1] Ministerstvo informatiky CR. URL: <http:// www.mcr.cz>. (Ministry of informatics Czech republic)

[2] Ceský statistický úrad. URL: <http:// http://www.czso.cz/>. (Czech Statistical Officce)

[3] Šimonová S., Komárková J. *Conception of the regional development of the microregion Pardubicko.(original in Czech)* Pardubice: Univerzita Pardubice, 2001. 98 s. AV-Upa-FES-ÚSII-08.01/02. ISBN 80-7194-373-8.

[4] Blazewicz, J., Kubiak, W., Morzy, T., Rusinkiewicz, M. *Handbook on Data Management in Information Systems.* Springer, Germany, 2003

[5] Whitehorn, M., Marklyn, B. *Inside Relational Databases.* Springer, Great Britain, 2003

[6] Humphries, M., Hawkins, M, W.,: *Data warehousing – Architecture and implementation (Czech translation)* Copmputer Press Praha 2002

[7] Abramowicz, W., Kalcynsksi, P., Wecel, K. *Filtering the Web to Feed Data Warehouses.* Springer, Poznan, 2002