# INFORMATION

# eFarmer – project "Bringing the datawarehouse for agriculture into practice"

J. Vaněk, J. Jarolímek, P. Šimek

Faculty of Economics and Management, Czech University of Life Sciences, Prague, Czech Republic

#### **ABSTRACT**

The key focus of the project "Bringing the datawarehouse for agriculture into practice" is to develop an IT solution for the implementation of data integration in agriculture. This solution is elaborated in cooperation with a whole range of bodies in the Czech Republic; the lead partner is the Pardubice region where the pilot phase of datawarehouse implementation is being performed. Apart from Czech legal subjects, partners from Finland, Estonia and Poland are also participating in the eFarmer operation. The datawarehouse project is one of the biggest subprojects of the eFarmer operation and is directly connected to the subproject of "Analysis of data and information flow and storage in rural regions", which was carried out within the eFarmer operation in the last period.

Keywords: datawarehouse; ICT; Internet; information; eFarmer

Agriculture is a field characterized by a very large and diversified database. Transfer, storage and processing of data gathered from individual activities of agricultural production are currently supported by information and communication technologies, at the same extent as in other sectors of industrial production or tertiary sector of services. The datawarehouse system aims at providing farmers with access to integrated data on their production practice and thus ensure relevant background materials for management decision-making, make such information available to agricultural advisors, offer statistical evaluation of farmers' groups, including their mutual comparison and on the whole support the competitiveness of agricultural businesses within the integrated Europe.

#### **MATERIAL AND METHODS**

eFarmer project generally aims at improving the quality and sales of agricultural production through electronic support of the latest management methods; proper datawarehouse implementation is to guarantee farmers and eventually other legal subjects an access to integrated data, thus maximally supporting their competitiveness.

By creating the datawarehouse solution for the field of agriculture, farmers gain two main benefits. The first one derives from the possibility to interconnect data from different fields of agricultural production and therefore acquire a complex overview throughout all the production parts. The second benefit basically comes from the first one; in other words, a user of the data-

Supported by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. 6046070906.

warehouse system can obtain information, news and reports from integrated data from particular agricultural applications. Besides these benefits, which essentially stem from the datawarehouse principles, such a solution may offer a remote access to data by taking advantage of Internet technologies. Consequently, not only a farmer has an opportunity to acquire his/her data in complex overviews but his/her contract advisor may also access the data without having to visit the farm in person. Through this, the solution can bring more effective agricultural advisory service as well as a greater support to management decision-making, owing to which good functioning of agricultural businesses is supported and the situation in rural areas bound to agriculture for centuries, may be improved.

The datawarehouse system currently involves three basic areas – plant production, animal production and feed production. Data from these areas are furthermore supported by supplementing data from price records, market news and information news.

Animal production – the proposal of the datawarehouse system in this field of agricultural practice consists of data on livestock records and dairy cattle utility. These data are acquired from systems such as the Central register of livestock and Utility control.

Feed production – the draft of the system assumes availability of data on records of used feeding rations, nutrients standards and values of quantitative indicators of feeding quality.

Plant production – the proposal relies on availability of data on plant production plans and current practice in plant production.

Information service – data from some agricultural information portals are available. It mainly involves news service, price records, weather forecast and exchange information on listed prices of agricultural commodities.

### Datawarehouse architecture

With respect to fragmentation of entry data and lack of defined relations the architecture of the datawarehouse system is proposed in order to make the datawarehouse activities as effective as possible taking into account its further development. Therefore, a special attention is paid to the preparation of entry data, their processing and modification. To divide the whole system into the

following, relatively independent parts seems to be the most effective way:

- (i) integrated database,
- (ii) datawarehouse.

Integrated database. Regarding the fact that the system's entry point consists of partial applications gathering data from agricultural practice, the Integrated database may be considered as a datawarehouse system entry point. Its key significance lies in collecting entry data from independent partial source systems, their integration and preparation for transfer into the datawarehouse itself.

The integrated database is internally divided into three components:

- (1) entry structures structures into which source data are imported in order to gain supplementary data required for creating relations between entry data and enlargement by supplementary data;
- (2) basic structures they contain integrated data extended by supplementary figures before modifying the structures for the datawarehouse itself;
- (3) output structures they contain data modified for the transfer into the datawarehouse itself.

**Datawarehouse.** The datawarehouse itself is composed of several databases containing data prepared according to the defined method for output system configuration. Thus, data structure consistence within the datawarehouse and system outputs are ensured. The datawarehouse plays a pivotal role in the whole system; however, the datawarehouse system itself only stores data and keeps audit records. In short, it accepts and offers data but it does not perform any other operations.

Integration is practically a service functionality of the datawarehouse system. It guarantees data provision from a partial system, their verification and connection to data from other partial systems. Integrated data are exported into the datawarehouse itself in regular intervals.

Possible scenarios for the function of "Data integration" as follows:

Data import – based on the requirement of the data import into the datawarehouse system several activities are carried out. First, identification of the source and data content, in other words, recognition from whom the data are (business, system, etc.) and what their content is (animal production, plant production, etc.). The data content does not always have to be clearly defined by the

source system. As soon as the data are identified, the system verifies them. The first verification is syntactic correctness, i.e. whether data correspond to actual schemes of the data structure. If they do, data are accepted and their semasiological correctness is assessed, in other words, it is checked whether they do not contain nonsensical figures. This check is carried out in an interval defined from figures already included in the database or defined by the datawarehouse system administrator. If the data do not pass through both checks successfully they return to the user with an appropriate report.

Data integration – from entry database structures where data are imported after their arrival and check, they are transferred and inserted into structures of the integrated database where they are accompanied by data from other partial systems. If there is a collision of unique identificators or relation data this import is cancelled. At the same time, the system administrator is informed. This is not a part of "Data import" because it can be more time-consuming and user, or source system, can fail to remove the error.

Data export – data are exported from the integrated database into the datawarehouse itself in regular intervals. This export is performed automatically in a predefined interval or upon request of the system administrator. During the export data are modified according to the set rules to a datawarehouse form – aggregation, change in relations – modification of figures according to the method saved in the metainformation system. Such modified data are saved in output structures of the integrated database, from where they are transferred into the datawarehouse itself.

The metainformation system contains a description of every single data in the database with its history. At the same time, it includes a methodology of data collection, their relations and interpretation. Thus, it is possible to form individual summary reports.

Possible scenarios for the function of "Metainformation system" are as follows:

Metainformation administration – collection, storage and change of information on every single entry stored in the datawarehouse itself. The administration of the methodology of data collection and their interpretation. These activities are only provided by the datawarehouse system administrator. All the data have their history. For this reason, when changing the methodology of data collection (e.g. change of units) the update

does not concern all the entries in the database but merely the change of metainformation on this entry and its methodology. Database consistence is thus easily guaranteed.

Creation of summary reports – system administrator can create new summary reports for individual system users based on the metainformation on entries in the database. By this relevant datawarehouse system outputs are provided without having undergone a change.

#### System users

The original plan counts with three interest groups of datawarehouse users:

- (1) state administration and public administration bodies or other institutions,
- (2) agricultural advisors,
- (3) farmers.

These interest groups cover the whole sector for which the datawarehouse solution is planned. Nevertheless, it is possible to extend these interest groups about at least one more significant group of potential users of the datawarehouse system – agricultural schools and universities.

Each of these groups comprises a large number of system users. Apart from people it also concerns computer applications and systems, which communicate with the datawarehouse.

#### **RESULTS AND DISCUSSION**

Maintenance of the server and all the endpoint computers is fully provided by the non-profit making company AgroVenkov, o.p.s., integrated SW applications are maintained by individual producers, and the agrarian WWW portal Agris, commodity exchange electronic system and eLearning system by the Czech University of Life Sciences in Prague (see the solution scheme in the Figure 1).

With compliance to the project schedule the eFarmer operation is currently (at the end of the first half of 2007) focused on approximately 60 end-users from the agrarian sector in Pardubice region (Figures 2 and 3). These users are relatively equally distributed within the whole region (always 15 users in each district – Pardubice, Svitavy, Chrudim and Ústí nad Orlicí). End-users who contribute with data to the datawarehouse system via a number of different desktop applications focused on plant production, animal production

## eFarmer (DW) structure

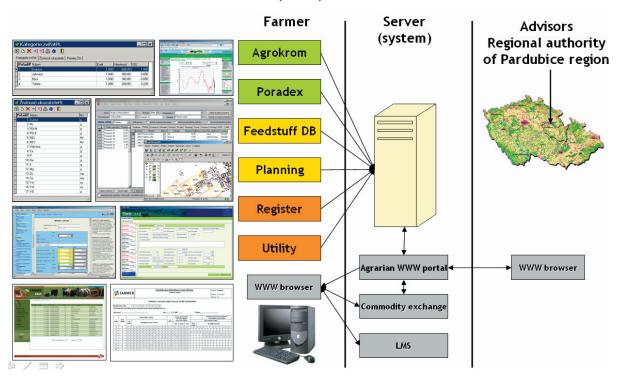


Figure 1. Datawarehouse structure

and production of feeding diets (namely Poradex, Agrokrom, Animal register, Database of feeding diets and rules, other utilities) are now managers of agricultural and food businesses and private farmers. Each of them disposes of a personal computer equipped with the above-mentioned

applications and connected to the datawarehouse system through the Internet. Gathered data, divided into publicly accessible and closed (according to the legislation of the Czech Republic), are accessible mainly to their owners, possibly to other interest groups.

# Pardubice region

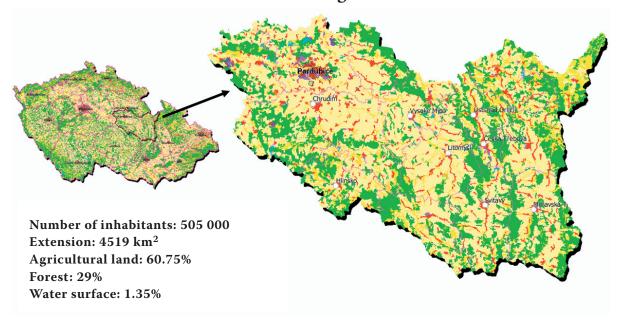


Figure 2. Pardubice region – solution to the pilot project

# 60 end-users in the pilot project

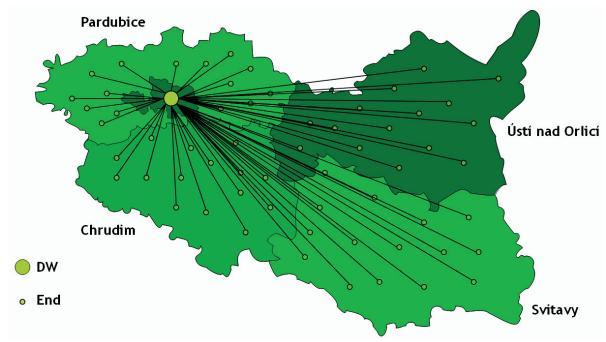


Figure 3. Pilot project – Pardubice region

System output data entirely follow the dataware-house system methodology, which was created by professionals in particular fields of agricultural production. When forming outputs in the publicly accessible part of the system the anonymity of data sources is naturally guaranteed as well as relevance of the provided output is guaranteed in the case of statistical outputs or when choosing a group of businesses.

New end-users also have, under certain conditions, an open access to the system implemented within eFarmer operation. It concerns purchase of a commercial SW application, which has an access to the central datawarehouse. The whole solution is thoroughly being developed as modular and transferable (extendable), mainly into other

regions in the Czech Republic; nevertheless, it may be used abroad as well.

#### REFERENCES

Mareth F., Jarolímek J., Vaněk J., Kubata K. (2006): Data integration in agriculture. In: Information Systems in Agriculture and Forestry 2006. PEF ČZU, Praha.

Šimek P., Jarolímek J., Vaněk J. (2007): Subproject DW in the eFarmer operation – Encouraging entrepreneurial competitiveness and regional development in rural areas. In: Environmental and Rural Sustainability through ICT – EFITA/WCCA 2007. Glasgow Caledonian University, Glasgow.

Received on July 19, 2007

#### Corresponding author:

Ing. Jiří Vaněk, Ph.D., Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta, 165 21 Praha 6-Suchdol, Česká republika

phone: + 420 224 382 279, e-mail: vanek@pef.czu.cz