Data integration in agriculture

František MARETH[1], Jan JAROLÍMEK[2], Jiří VANĚK[3], Karel KUBATA[4,]

Czech University of Agriculture Prague
Faculty of Economics and Management
Kamýcká 129, Prague 6, Czech Republic
e-mail[1]: xmarf100@studenti.czu.cz
e-mail[2]: jarolimek@pef.czu.cz
e-mail[3]: vanek@pef.czu.cz
e-mail[4]: kubata@pef.czu.cz

Abstract

Agriculture is specific sector also regarding data complexity. One subject may have data of its plant production results, livestock production or farm food production. There is an essential need for data integration. The farmer could than at the first sight combine results from all fields of interest. Second aspect is timing. This aspect could be realized by data warehousing.

Key words

Data, Integration, Data, Warehouse, Data warehouse, Agriculture, Plant, Livestock, Farm food, XML, SOAP, Application, AgroXML, ISOBUS

Introduction

Czech university of agriculture become participant of European project eFarmer. This project is concerned in stimulating enterpreneurship competitiveness and regional development in rural areas. One of the subprojects is aimed on data warehousing and data integration in agricultural area.

In agriculture there are several standalone applications specialized on specific area of farming like plant production, livestock production or farm management. The farmer must hence use multiple applications to fulfill his needs. On larger farms is the situation a bit different. There are several specialized groups of workers, when every single group usually works with one type of application.

Due to this circumstances is the data integration an essential need for agriculture. Data from each application might be stored in common database and used for comprehensive data analysis. The farmer would have possibility to see what was foddered from his own plant production or vise versa where was the dung used on his soil.

Data integration is only the first step. If the time factor is mentioned, then data integrated from several areas are stored in longer period and represent huge amount of information. Hereafter are the data warehousing techniques quite good solution. Unfortunately is this solution more expensive and might be used only by larger companies. By the help of European union might be realized common data warehouse for agricultural sector, that may be used by small farmers as well.

The eFarmer project is in the beginning at the present time. Efforts are mostly aimed on description of needs and realizing the conceptual model of data integration. The needs are described in two levels actually. The first level is represented by existing applications functionality and because these applications are used in agriculture for several years, that

should fulfill most of farmers needs. The second level is based on description of processes in farming and is extending the first level needs.

Data integration

In the early beginnings of subproject was the idea of data integration based on data warehousing. With further development was realized, that data warehouse is not the best solution. Data warehouse is not normalized database with huge amount data stored inside. Its query might take longer time for answering so its useless for operational purposes.

Next factor is, that this data warehouse should contain data from several applications and systems. This data should be not only stored but also provided to other applications in real time. The first idea was to create some kind of data market as a pre-storage of data warehouse. But due to fact, that data from application could be exported irregularly, was the standalone database chosen.

The standalone application will be connected by the set of interfaces. Content of this database might be regularly exported into data warehouse for more complex analysis. But the database will stay operative to fulfill request from connected applications.

As a solution for interface was the common interface SOAP[1] based on XML[2] chosen. The other possibilities like native connectors of performed database were rejected. The rejection was based on fact, that the interface has more functions then provide data. The problem is in usage of different code lists and data structures in each single application. Therefore should the interface fulfill also the functionality of converter from application data structure into integrated database structures and vise versa. Data structures needs to be described by XML schemes[3] or by similar ways.

Unfortunately interface does not solve the problem with code lists. For example one company must be uniquely identified across all the applications. But it's not the problem of company id only, but also the problem of identification of species, livestock, etc. Some kind of standard has to be defined. Due to limited resources, some existing standard should be used. The possibilities are AgroXML, ISOBUS or standard used in one of existing application. The analysis of the possibilities is done at this time.
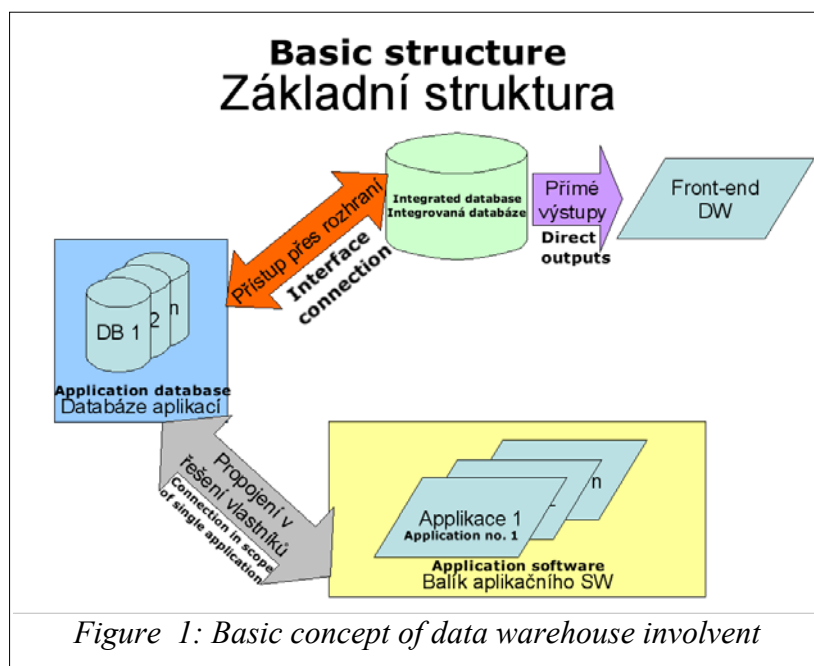
Data warehouse

The concept of data warehouse is still a little bit misty. The main idea is in regular export of data from integrated database into data warehouse, where the data may be used for comprehensive analysis. Its output should be used not only by farmers, but also by farm advisers or official authorities. In regard with this we can consider three groups of users – farmers, advisers and authorities. This extends the needs analysis by two more groups.

One of the main problem in this area is the copyright. If the data will be available for other users, not only to the owner, then might be the data misapplied. Due to this fact, there has to be defined detailed security policy which must be strictly implemented.

---

1 SOAP - Single Object Access Protocol (more on URL <http://www.w3c.org/2000/xp/Group/>)
2 XML – eXtensive Markup Language (more on URL <http://www.w3c.org/XML/>)
3 XMLSchema – A possibility how to describe XML data structure based on XML (more on URL <http://www.w3.org/XML/Schema/>)

Figure 1: Basic concept of data warehouse involvent

Conclusion

Integrated database for storing farm data from different areas of agriculture is one of the possibilities how to improve position of farmers in the beginning of 21th century. Building this solution on widely spread technologies as XML will help the project to succeed. Realizing solution opened for usage may improve not only farms, but also software companies to develop more variant application for agriculture. And this may in fact again improve the farmers and help to solve one not mentioned problem. The problem is low level of penetration of IT in agriculture especially among small farms and holders.

Realizing system based on common data structures will also help to improve the process of standardization in agricultural, where due to specific circumstances is this process much slower and complicated then in other areas of business or industry.

As a by-product of this system is a conceptual model of farm processes. This model might be used to further analysis of farm organization and the management might extend the usage of management tools like internal audit or controlling to improve the competitiveness.

References

[1] Pokorný, J.: *Konstrukce databázových systémů*. Praha, ČVUT 2001.
ISBN 80-01-01935-7.
[2] Parr Rud, O.: *Data mining*. Praha, Computer press 2001. ISBN 80-7226-577-6.
[3] World Wide Web Consorcium: *SOAP*. [ONLINE]
Available on URL <http://www.w3.org/2000/xp/Group/>
[4] World Wide Web Consorcium: *XML*. [ONLINE]
Available on URL <http://www.w3.org/XML/>
[5] World Wide Web Consorcium: *XML Schema*. [ONLINE]
Available on URL <http://www.w3.org/XML/Schema/>
[6] World Wide Web Consorcium: *XSL, XSLT*. [ONLINE]
Available on URL <http://www.w3.org/Style/XSL/>
[7] *European eFarmer project*. [ONLINE] Availible on URL <http://www.efarmer.info/>