eFarmer - Project of Data Integration

František MARETH[1], Karel KUBATA[2], Jiří VANĚK[3], Jan JAROLÍMEK[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]: kubata@pef.czu.cz
e-mail[3]: vanek@pef.czu.cz
e-mail[4]: jarolimek@pef.czu.cz

Abstract

The problem of data complexity is often discussed as a part of management support. Decision-making is based on available information. The information is usually provided from heterogeneous sources without any formal relation between them. Agriculture is not an exception otherwise in agriculture the problem is more important then in other areas. Linkage between plant production and livestock is particular, but usually not in scope of information systems. Described project shall provide data integration solution.

Key words

Agriculture, Data integration, Project management, Information system, Design, Implementation, Analysis, Data standard, Data warehouse, Consulting, Farm

Abstrakt

V souvislosti s podporou řízení podniků je často zmiňován problém komplexnosti údajů. Rozhodování řídících pracovníků je založeno na dostupných informacích, které jsou často z heterogenních zdrojů bez formálních vztahů mezi nimi. Zemědělství v tomto není výjimkou, naopak je zde tento problém zřetelnější než kde jinde. Provázanost rostlinné a živočišné produkce je zřejmá, ale ne obvykle v oblasti informačních systémů. Popisovaný projekt má za cíl vytvořit řešení problému datové integrace.

Klíčová slova

Zamědělství, Datová integrace, Projektové řízení, Informační systém, Návrh, Implementace, Analýza, Datový standard, Datové skladiště, Konzultace, Farma

Introduction

Czech university of agriculture is a lead participant in the project of data integration and usability of data warehousing in agricultural and rural area. In this project are also involved most important companies realizing applications for agriculture in Czech republic. The project considers providing independent platform and described data standard for data integration and exchange among these applications. However it's connected with commercial enterprises the project should provide open platform for any application in agricultural area.

Right now we are at the beginning of the project. We've several applications from different areas of agriculture like livestock production, plant production, farm food evidence, business planning, e-learning and market information such as market prices. The mission of the project is to provide open platform for data integration and exchange to help the competitiveness of farmers – to support their decision-making. This project is divided into three main parts. The

first (the one we are actually working on) part is the prior analysis of project feasibility – the initial analysis. This part is focused on analysis of available data sources, application usability and users' requirements. The second part is based on detailed analysis of data structures and relations between data, realizing data standard structure, database structure and data connectors to other systems. The last third part is focused on realizing the platform and pilot run if the system. There will participate about 60 farmers from Czech republic in the last part. These farmers will receive several applications to realize initial data upload and first live database tests.

This report is concerned in the first part of the subproject, the part we are actually solving.

Initial analysis

According to general project plan we have to find out the project framework. This part is called the initial analysis in this report. The output of this part shall be a document that might be called initial study – study of project feasibility, specifying project framework with available information sources.

First step was to specify the project strategic focus in more formal way – to specify the problem. The strategic mission was mentioned in previous text. The problem was specified as a problem of integrating data from plant production, animal production (livestock, sheep, coats), farm food evidence, farm economic (financial) data and market data that will be used for reporting purposes, remote farm advisory services, as a source for external web services and of course for improving the farming. The e-learning application mentioned above will be used as some kind of user guide for the whole platform; no data from this application will be integrated. The goal is to provide the platform opened for each application in agricultural area to integrate applications' data in specified standard.

As was the problem specified, it was possible to specify the project framework. The specification is based on sets of user roles and role requirements. This led to the matrix of roles and requirements. This matrix was not filled yet. According to this matrix the next goal was to find out necessary data for requirements fulfillment. However the requirement will be the same such as "Evidence of livestock" the data will differ for each role. The farm CFO will rather have data in financial units otherwise the breeder need the real amount of cattle in the stable. This is the reason why the matrix of roles and requirements was found useful.

If the data were specified (in general level), it's possible to specify the technical background of the platform – to specify the number of user requests, amount of stored data and to figure out the number of relationships between data. The reason is to specify the amount of investments needed to build the platform background. Available solutions are the Oracle database server running on certified Red Hat Linux distribution, the Microsoft SQL server running on Windows Server. The solutions based on open source database machines such as MySQL or any other databases like Interbase or GemStone were found as unsuitable due to leak of data mining or data warehousing tools. Both Oracle and MSSQL database are widely used for this kind of solutions – data integration or data warehousing. The amount of data was not specified at the moment, but the first approximation is that there will be thousands of data elements and hundreds of relations.

Next thing that needs to be specified in this part of project – initial analysis – is to choose the methodic for further detailed analysis. It was agreed that the outputs had to be in UML, but the methodic was not specified yet.

Once we have finished this initial analysis we shall know what exactly we want to solve, what is the framework of the subproject, approximate financial and time costs. This information will be contained in the initial study – study of feasibility – that is a good starting point for further detailed analysis planned for next part of project.

Consultant services will be needed to fulfill the goals of initial analysis. Not only from the field of information technology but also from the field of agriculture. The experts from agriculture will be provided by other project participant – the software companies, the experts from IT are still unknown, but the negotiations are led with Microsoft consultancy services, the Hewlett-Packard Company and several middle-sized companies from Czech republic.

Detailed analysis

Because the project is still in the part of initial analysis, the detailed analysis will be mention only shortly. The output of this analysis shall be the description of all data elements, relations between data elements. The data standard will be specified based on the data elements and data relations. The connectors to the database will be also specified.

The data standard will be used for internal communication and also for communication with external systems. The database will of course use native data format, but we have to also manage the communication between possible modules of the platform. The problem is that there is no available commonly used data standard for agriculture. It's also not the goal of the project. As a part of detailed analysis all the needed code lists shall be specified.

The problem of code lists is one of the biggest problems in data integration. Unfortunately there is no widely accepted set of code lists for agriculture. The analysis shall solve the transformation between different code lists if it is possible.

The database connectors will be based on XML/SOAP/WSDL and appropriate data standard. Nevertheless the connectors are based on widely spread technologies there could appear significant problems with realizing communication with external systems.

The documentation of the system will be created continuously during the second part of the project according to the accepted partial results of analysis.

There will be a strong need for consultancy services in this part of project. The consultancy services will not be only needed with IT experts but more then ever with experts from agriculture to find out the significant relations between data elements. This part of the project is the most difficult and its' result has direct impact on system realization.

Realization and Pilot run

The last part is the system realization. It's really not possible to describe the realization of the system if the analysis didn't finished already. It's planned that on the system design and system implementation will participate the Czech university of agriculture and also some professional commercial IT company that was not specified yet.

Before the pilot run all the users will be trainer in using the system (applications communicating with the system). This training will be based on documentation that is the output of the second part of the project. There will be used the e-learning application provided by the University as a tool for training. This training will take place during the implementation of the system to approve the user knowledge before the pilot run. The applications communicating with the system will be provided to the users to improve the results of training (the applications will not communicate with the systems of course, but other functionality of the applications will be available).

There will participate about 60 farmers in the pilot run. All applications will be provided for free. The only requirement is the Internet connection that will not be part of the pilot run.

Time plan

The first time plan was specified as follows. The outputs of the first part will be available till the end of June, the second part (incl. Implementation) will finish at the end of December and the pilot run will take part from January till June 2007. Due to change in project plan – the implementation was separated from detailed analysis and added to the realization and pilot run – the time plan has changed as follows. The first part will be ready till the end of August. The second part will finish at the end of February and the last part will be from March till the end of June 2007. The pilot run shall be at least 2 month.

Conclusion

This project is concerned in comprehensive problem – realize data integration in agriculture. There is about one and half a year for the whole analysis and realization. It's not too much time, no time to loose. Precision project management might help to success realization. By dividing problem into the simple specific tasks might improve effectiveness of its solution.

The realization of the data integration in agriculture was divided into three single parts. The first is concerned in problem specification and initial analysis, second one in the detailed analysis including describing data elements and relations. The last part is concerned in realization – system design, implementation and pilot run. These parts follow the common project modality – analysis, design, implementation and run. Due to specific problems in agriculture was the analysis divided into two separate parts. Otherwise was the design and implementation included into one part in anticipation of outstanding results of analysis.

During the whole project run the consultancy services of external experts from the IT and agriculture will be needed. This shall improve the result and help to realize open professional solution with long life presumption.

References

[1]   Voříšek J.: Strategické řízení informačního systému a systémová integrace. Praha. Management Press, 1997. ISBN 80-85943-40-9

[2]   Polák J., Merunka V., Carda A.: Umění systémového návrhu - objektově orientovaná tvorba informačních systémů pomocí původní metody BORM. Praha. GRADA, 2003. ISBN 80-247-0424-2