Data processing and computer developments of milk recording in Estonia

K. Pedastsaar & T. Murulo

Estonian Animal Recording Centre
Tartu, Estonia

Summary

1. The transmission from the mainframe data processing system to the use of Relational Database Management System (RDBMS) is a complicated task:

    *   it is a long-term process during which the data have to be synchronised;
    *   the quality of the data used in the mainframe system was deficient for the implementation of the RDBMS. For the data transmission, we had to improve the quality of the data saved in the old system;
    *   the designers, programmers and users of the system need some time to adapt to the ideology of RDBMS.

2. Soft- and hardware experience:

    *   there should be widespread use of the RDBMS version. It gives us the tools for the development, flexibility and support for the future;
    *   at the time of development of the system, a powerful hardware and expensive operation system were not needed. For example in the case of Oracle the changing of the database server and the operational system was not a problem. Besides, the development of soft- and hardware is fast;
    *   in the case of a relatively small population of animals, the system can successfully be exploited at the cheap INTEL/Windows NT platform.

3. Development experience:

    *   the advice of the foreign experts was important to us;
    *   people need extensive training;
    *   the data model has to be planned in the early stage of the development of the system;

* for the unique identification of animals (objects) in the database, it is very useful to apply the identifications generated by the system (in the case of Oracle by Sequence). The identifications of animals used by people, register numbers, inventory numbers, names, etc., are not suitable as a primary key;
* as few calculated values as possible should be saved into the database. For instance the calculation of the current lactation production and the control-year production "in-the-fly" is justified;
* it is practical to save the pedigree of animals recursively into the database, the data of the animal sire and dam is not saved in the animal's own record. Only the pointers to sire and dam records from the same table are saved in the records.

Introduction

The Animal Recording Centre (ARC), a service organisation for milk recording, occupies a central place in Estonian animal breeding providing services for dairy and pig farmers, breeding organisations as well as veterinary services, dairy processing industry, advisers and research institutions.

This report gives the historical overview of the development of data processing for the milk recording (MR) and cattle breeding in Estonia. Above all, we would like to share our efforts and experience with the so-called "rightsizing" project (1994-1998): the transmission of the data from the mainframe data processing system to the PC system of RDBMS on Intel/Win NT. The whole technology was completely changed and all the operations reprogrammed.

There were two main reasons to accept the challenge:

- the service was not satisfactory for the farmers and breeding organisations therefore there was danger that they would not be motivated to use (and pay) for the service;
- the mainframe system was morally and physically old and maintenance and development of the system would be too costly.

Getting started with computers

The introduction of computers and electronic data processing of Estonian MR can be divided into four stages:

1. The first attempts at the computerised data processing in Estonian milk recording were made in 1964. At this time the computer only calculated the herd yearly productivity average. The estimation of the breeding value for the bulls (contemporary comparison) was also started.
2. During the second period, the elaboration and implementation of computerised data processing of milk recording got under way: regular monthly reporting, herd year summaries, etc. In 1969 the application

Table 1. Abstract of process of development.

1964	Herd yearly productivity averages. Sire evaluation: contemporary comparison.
1969	Start of computerised data processing of milk recording.
Up to 1982	Data processing in different units on computers of two generations.
1978	Start of designing new data processing system for the IBM/370 compatible computer.
1982	Estonian Research Institute of Animal Husbandry received a computer EC-1035 (1MB RAM, 3x29MB disk units, SVS).
1982-1984	Transition period to the new mainframe system. Since 1984 there has been centralised data processing in Tartu.
1986	Replacement of the computer and the operating system:
	EC-1055M with 4MB, 8x100MB disk units, CBM 3.0 (VM/SP) Trial of ADABAS. It was clear that the capacity of the computer was not sufficient for using the database system.
1986-1992	Quiet development.
1993	The Animal Recording Centre was formed. We made a decision to replace the mainframe background by the PC background for the future. We chose Oracle 7 under NetWare 3.12 for development.
1994	Start of putting local network and Oracle into use.
	Milk Analysing Laboratory got two lines of Combifoss 4300 and the so-called *Laboratory* programme was put into use on PCs. It was an essential improvement of the milk recording service.
	Data Processing Unit replaced the mainframe EC-1055M with mainframe HDS-8053.
1995	Enjoyment of the increased capacity of the mainframe and continued development of services on the mainframe.
	The rough draft of data model and technology was designed for the ORACLE System.
1996-1997	Stopping of the development for the mainframe system, however, mainframe was operational for all the services.
	Programming in Oracle.
	The first run of BLUP Animal Model, the first operation, application and output, utilising new cattle database in ORACLE. Synchronisation of data between the mainframe and the Oracle database every night. Replacement of the ORACLE database server and new operating system.
January, 1998	Start of production of MR services in Oracle. The so-called *Laboratory* programme is running in Oracle. That means that the data input of milking tests, AI, events and output: four different reports (milk recording reports, pre-print of barn lists, udder health report and invoice) is executed in Oracle.
May, 1998	Printing modified pre-printed barn lists for small size herds.
June, 1998	Full cycle of data processing (input, check and maintenance, output reports) for small herds in Oracle. Milk recording reports, error-list and as option action-list added.
	The batch technology of data processing was put into use.
	The whole technology will be applied to all the herds in September 1998.

programme was launched in one of the counties, by 1973 the system included as many as 105 000 cows which made up 43% of the total number of cows in collective farms. The work was disbursed amongst several computing centres. The use of computers of different type and different applications called for the adaptation of programmes.

3. In 1984, centralised data processing began. All the production, reproduction and breeding data were organised in one data bank in Tartu.

4. The application of the RDBMS Oracle was started in 1998.

A Survey of the So-called "Rightsizing" Project - Our experience

Something had to be changed – 1992 to 1993

The re-establishment of Estonian independence in 1991 brought about major changes in society as well as in agriculture. The creation of private farms was explosive: in 1992 there were 429 herds using the MR services of the ARC, whereas by the end of 1993 the number had gone up to 3767. The tendency was accompanied by a rapid decrease in the average herd size caused by the agricultural reform in Estonia. The data processing system of milk recording, designed at the beginning of the 1980s for a small number of large-scale farms and often with a view to meeting the requirements set by the Soviet Government in Moscow, was well out of date. The lack of computer capacity determined the primitive nature of the data processing technology. The service was no longer acceptable to the client. The State obviously could not provide 100% funding for the MR service any longer.

Taking the decision - 1993

It was decided to change the mainframe background, as its development and exploitation costs were too high. By that time some of us had seen the first PCs and had had some experience with the database (FoxPro). Within a very short time, it was decided in favour of a server of RDBMS architecture: a short list between RDBMS Progress, Gupta and Oracle was made. We discussed the choice of the operating system which in retrospect, did not make much sense.

We opted for Oracle 7 for NetWare 3.12 and Oracle Tools for DOS as the development tools. As for the database server as well as the local network server, Olivetti Suprema M6-880 with 60 MHz Pentium processor, 64MB, 3GB were bought. Thus, in the beginning we had to make a start with modest means.

It was in March 1994, when the local network and Oracle were installed and most of the programmers got their very first working experience with PCs. We were amazed by the obsoleteness of the Oracle Tools for DOS. We were not able to put it into actual use, but at the same time the service was in need of improvement.

The first acquaintance with Oracle and its developer tools - 1994

In October 1994 we laid a foundation for the further improvement of the service by fitting the Milk Analysing Laboratory with two Combifoss 4300. These devices allowed the real electronic recording of the milk analyses results. In cooperation with the Verden Computer Centre in Germany and in particular, with Dr Jürgen Claus, we had completed a FoxPro application for the merge of the milk analyses and barn lists results and for printing the test-day reports for the clients. In the same year, the lifetime numbers for the cattle were introduced, since there were two different types of ID numbers (intraherd and/or lifetime) in use. The collection and input of AI data, as well as type classification data, was reorganised.

The improvement of the service quality - 1994-1995

At the same time, the small capacity mainframe EC-1055 of the DPU was replaced by a real mainframe HDS-8053 which allowed the large-scale implementation of the database system Adabas and granted a good contact between the PCs and the mainframe.

The Oracle development environment Oracle Developer/2000 was launched and it was found that it was possible to programme in Oracle. The visit of Dr Eildert Groeneveld (from Germany) to Tartu in November 1995 was of great relevance and the results were stated in the article "Konzept zur Organisation der Datenbank und Datenverarbeitung am Jõudluskontrolli Keskus in Tartu". The implemented ideas about the automatic batch processing are derived from the very same paper. As for the method of production calculation, we opted for the interpolation method, suggested by ICAR and as the processing period we chose the interval between one test milking (day excluded) and the next one (day included). Intraherd and interherd comparison is based on a calendar month.

The elaboration of the data model and the conception of data processing for the cattle database in Oracle-1995

After the establishment and testing of the data model structure, the "rightsizing" project advanced at full speed. There were four DPU programmers engaged in the programming. The completion of the project took approximately 1000 man-days.

Programming. The data synchronisation between the mainframe and the PCs – 1996 to 1997

At the same time the saving of the current and historical data of the mainframe was started in Oracle. The synchronisation of the data between the mainframe and the Oracle database was now our every-day work. The loading of the data model and the synchronisation of the data, form one of the most complicated parts of the project. It was mainly due to the

deficient quality of the data saved in the mainframe. A lot of the data quality checks were done during the preparation for the implementation of Animal Model (1996) for the estimation of the breeding values and joining INTERBULL (1997) services later on.

Status quo - July 1998

Hardware and software of the "rightsizing" project

Database server, DELL POWEREDGE 6100, installed in July 1997 with:
    2 PentiumPro 200MHZ, 512KB Cache processors
    256 MB ECC EDO RAM
    6 x 4 GB SCSI-2 Ultra/Wide 7200 rpm HDD
    DLT 15/30 GB backup unit
    Windows NT operating system
    Oracle Server Enterprise Edition 8.03 for Windows NT.

Local network server, DIGITAL Prioris MX 6266, installed in January 1998 with:
    Pentium II 266MHZ processor
    128 MB RAM
    3 x 4GB HDD
    Novell NetWare 4.11 operating system for 50 user licenses and Connect 2.0 of NetWare with 8 licenses.

Network printer, 24-ppm HP LaserJet 5SI/MX with duplex unit, multi-bin mailbox and 2000–sheet input tray.

Developer tools

    Oracle Developer/2000 Release 1.5.1
    Microsoft FrontPage98
    Micrografx FlowCharter7.

Data processing services

Altogether there are 3033 herds with 123 000 cows (which make up 73% of the total number of Estonian cows) in milk recording. Since January 1998, the entry of data and the processing of test milking have been executed wholly in Oracle, i.e. all the farmers get the following reports:

*   milk recording reports (test-day results of individual cows and herd and/or milking group average: milk, fat, protein, urea, SCC);
*   pre-print of barn lists for test milking;
*   udder health reports (extra service);
*   invoices.

Printed Estimated Breeding Value (EBV) reports for cows and young-stock and cow sheets (pedigree + production + reproduction) are provided as extra services. New services for the farmers: reproduction advice, feeding advice, accordance with milk urea content, sire mating advice, etc., are designed.

The farmers and the breeding organisations have access to the EBVs for type, production and SCC. There are several options for all the clients, especially while the small holders' needs differ from those in large scale farms.

In addition we offer the on-line service. If the client is in possession of a PC and a modem, they have access to their data in the central database. They can make inquiries and download the data for local use.

The current calculation of milk production and monthly report printing is at present executed in two environments, namely:

* small size herds (up to 30 cows), i.e. 2500 herds with 21 000 cows i.e. 82% of the total number of herds and 17% of the total number of cows have been processed wholly in Oracle since 1 June. Besides the above-mentioned reports these clients receive the main report of the MR period and in the case of errors, an error report. The results are processed on the day of laboratory tests or the day after and the reports are posted via regular mail;
* in the case of large size herds, the milk recording data processing and the main print of production period (530 herds), are executed by old technology (up to September 1998) on the mainframe and the clients get the reports twice per the production period (month).

Table 2. Main cattle tables in Oracle system and data contents.

Table name	Content	Number of entries
Owner	Our clients (MR+AI+IR)	40 622
Animal	Animal pedigree	1 430 000
Owners' animal	Location of animals	606 000
Milking	Milk recordings	3 560 000
Lact	Lactations	3 200 000
Insemination		740 000

References

Claus J. Seminar on Animal Model Breeding Values: How to understand and how to make use of them. 19-20 November 1996, Pirgu, Estonia.

Groeneveld E., 1995. Konzept zur Organisation der Datenbank und Datenwerarbeitung am Jõudluskontrolli Keskus, Tartu.

International Committee for Animal Recording (ICAR). Constitution. Rules for the Granting and for the use of the Special Stamp. International Agreement of Recording Practices. 02-06-1994.

Reents R., Uba M., Pedastsaar K., Vares T. Implementation of animal model for production traits of dairy cattle in Estonia. INTERBULL meeting, June 23-24, 1996. Veldhoven, The Netherlands.