Normalized database preserves radio programme information for internal users and research

Mirja Loponen, Osmo Palonen; Mikkeli University of Applied Sciences; Mikkeli, Finland

Abstract

How to archive databases is a challenging task in digital archiving. Often the systems including databases with important have to be closed; either the used technology is getting obsolete or the costs to run old systems are getting very high. In many cases the databases are still carrying valuable information. The regulations and recommendations by the archivists often advise to store information as sequential files. In those formats the data is safe but in most cases very hard to use when the original applications are no longer available.

The joint project ElkaD Mikkeli University of Applied Sciences (UAS) and national business archives in Finland (ELKA) has created and implemented a better choice. The database used by the Finnish Broadcasting Company (YLE) was transferred into a normalised SQL database using standard ISO SQL tools only. In this case the information can be transferred from one SQL engine or version to another without any re-programming.

The Mummy-Musa, as YLE calls the archive, includes a database and user interfaces with Programme search, Piece-of-music search and Reporting functions.

The Mummy-Musa was developed in the framework of the ElkaD project, co-funded by European Social Fund, Province of Eastern Finland and Southern Savo Region. Mikkeli UAS is now continuing to develop normalized databases in an EU-funded project Aton.

From paper reports to computer

YLE has collected data about its broadcasted radio programmes for decades. The information about the copyright owners, artists and journalists in each radio programme was first collected manually by using report forms that have been archived in ELKA for decades. In the late 1980s YLE started to plan a computerized system to collect this information. The journalists thought filling in the report forms was boring and the results were not very accurate, especially as the payments for the copyright owners were based on sampling. YLE paid copyright on the whole single piece of music even when music was played only for a couple of seconds in the programme. The system sought, therefore, to get accurate information and pay copyrights based on the real programme time.

The in-house software department started in 1987 to plan a new system and decided to use the database as information storage. The on-line application was using CA-ADS and the database was build based on IBM IDMS. Batch processing such as the music reports for the copyright organizations were coded using COBOL and ReportWriter. The system was running in IBM 3090-180 mainframe with 25 MIPS, operating system was MVS/ESA and disk space 50 MB. The event management system was IDMS/DC rel. 10.1.

Picture 1: Programme search. First section includes search by free text, name, journalist, between dates and broadcasting network. Programmes can be find also based on tunes by using similar kind of search fields.

Picture 2: Results from a programme search.

The Musa system was launched in June 1990 and during autumn 1992 all Finnish radio channels used it to send reports to the national copyright organizations TEOSTO (creators) and GRAMEX (performers). The system was discontinued in 2005. Today a new Musa2 is in use, but the pre-2005 information was not transferred there. However, as the old information is still needed when re-sending programmes, a method to continue with it had to be developed.

Picture 3: Programme report

Picture 4: Tune search. The name of the tune, performers, genre, the name of the recording, label and some specific fields can be used in ther.

Around the same time as YLE was planning to stop their Musa system Mikkeli UAS was looking for a partner with born digital material to join the project ElkaD, where the national business archives – officially the Central Archives for Finnish Business Records (ELKA) – and the UAS were implementing a digital archive for ELKA. YLE decided to join in. Besides having a shared goal, the ElkaD and YLE shared a continuity between the traditional paper archive and digital archive: the paper-based music reports that pre-dated the Musa system were already archived in the ELKA's repository.

Normalized database

The project to specify and implement sustainable database started with a study for a far more long-lasting solution, still keeping the functionality. The project crew had read about normalized databases, but had not much practice. Based on the study it decided to create a SQL database based on the ISO-SQL only. That decision dictated that none of the SQL-specialties like procedures and triggers could be utilized. Those are widely used in active database-based applications build using different commercial database products such as SQL Server and Oracle.

In the case of digital long-term preservation there was a need to save the information in two different formats. The project crew found out that it is a good practice to keep all material in two formats and content: in the normalized archival database and as structured XML-files taken from the database.

One big challenge for the project was that all the planning and implementation documentation was not found. During the active life cycle of Musa system the product and the whole in-house software department was handed over to a joint company Tebit Oy (Tieto Enator Broadcasting IT) founded by YLE and a big Scandinavian software operation Tieto Enator.

Another problem was typical for the IT-products getting close to the end of life-cycle: the changes in organization and technology had created the situation where only few people who knew the system were still available. In this case luckily just enough. Those who were involved tried to find documentation on the system, but even when most of the people knew that these once existed and were in use, nothing was found. One user manual for the system was found by Lasse Vihonen, who was specifying and implementing the system in the late 1980s and early 1990s. He is today the chief of Radio archives in YLE.

Documentation is needed

The case of YLE demonstrates the demand to archive proper documentation of each database system including the data structure and formats of the database. In fact this is something that should be defined in the first enquiry of a system and included in the contract. In the project the first step was to find out the data formats, descriptions of the values used and the relations between the tables and data. It was quite a lot work when all the required information had to be studied from the old application.

When the specification process for the Mummy Musa, as the archival database is called, was started, the project team found out that the old system held a lot of information that was no longer needed. One had to make value analysis and weeding to limit the amount of data. The project team consisting of the users and designers of the old system in YLE and the developers in Mikkeli UAS, decided to go backwards from reporting to the information held in the system, i.e. the information and data included in the radio's broadcasting request form. The form holding general information about the programme and the music played held all the information valuable for the further use.

The information gathered from the old system was taken into the tables in the new archival relational database, which was designed as simple as possible. The change in technology demanded that data conversion was needed: fixed long data fields that had only few characters were changed to flexible fields. In practice the data was collected from the old system into ASCII files and those were loaded in the new archival database. The verification of the data was done by sampling.

The new archival application consists of three main elements: Programme search, search by tune and report generation. The interface to the system was planned to be browser-based and using open standards J2EE and JSP. In programme search the user can search broadcasted programmes by selecting from the information fields or pieces played. In search-by-tune option music can be searched by composer, lyricist, etc. The results can be saved or printed out as a report.

The archival database application keeps the database constant; however it is possible to add information from the successors. In fact it is also possible to transfer the information from the paper

144

reports to the archive database by digitizing the reports, if needed. From technical point of view the latter is possible, but the problem is, who can pay it.

For time being, the current SQL product (MS SQL Server 2000) will lose support and become obsolete. The normalized archival database can easily be carried into a new SQL product when no product and version related features are used. The easiest way is to just use the functionality build in the SQL-database product to transfer the data. When using this method close to emulation, there is a need to verify that everything is converted completely. When considering the life cycle of archival database, we cannot know how long SQL database format will be supported. Then datasets can be transferred into a new database by using XML files. The precondition still is that the new product can handle XML transfer structure.

Based on our experience the database management systems have often been build so that the new ones have the feature to ingest data from the old database structures. In the ElkaD project the old information of a Lotus Notes based photo archive were successfully transferred to RDF-based system using SQL. The authors of this document have also the experience of gathering information from dBase to SQL without any bigger problems.

Nine million events

The Mummy Musa has 9 million events in the database and is in daily use in the Finnish Broadcasting Company (YLE). There is a secure connection form YLE WAN to Mikkeli UAS digital repository, where the database and application is located. Also the researchers can use it in the ELKA as one application of the digital archive services of the business archives as the research room PC's have access to the Mikkeli UAS.

After the archival database has been opened to YLE personnel and researchers in ELKA there is an option to build specific SQL Queries for specific research purposes. One idea is also to create services for the public. As an example some might be interested to know what kind of music was played when they were born or how many times their favourite song played in their wedding has been played in public.

References

Järn-Kokkinen-Palonen: ELKAD Puheenvuoroja sähköiseen arkistointiin – Publications of Mikkeli University of Applied Sciences, Series A: Research Reports 7, Mikkeli 2006

ElkaD Loppuraportti, Mikkeli UAS, Mikkeli 2006

Picture 5: PDF-Report.

Author Biography

Mirja Loponen studied system design in ATK Institute in Helsinki the 1970s. Since1973 she has been working in specialist and managerial positions in Tietotehdas and Honeywell Industrial Automation in Finland. She has been working with Mikkeli University of Applied Sciences as database specialist form 2004. She also is the chairman of the Union for Municipal Archives in Finland.

Osmo Palonen studies history and IT in University of Tampere and has made FKT diploma including graphic arts, IT and management in Helsinki Institute of Marketing. He worked from the early 1970s to the 1990s in newspaper industry as a journalist, including project and systems management for the editorial IT-systems. Before joining the Mikkeli UAS in 2003 he worked for Honeywell Industrial Automation. Palonen is now in charge of the digital repository, archiving projects and service contracts as well as being the MD of the Disec Oy. He is a member of the board in Union of Business Archives and the Editor of the quarterly archiving specialist magazine Faili.