

The modernization, migration and archiving of a research register

Johanna Räisä
Mikkeli University of Applied Sciences
Patteristonkatu 2
P.O.Box 181, 50101 Mikkeli
johanna.raisa@mamk.fi

Mirja Loponen
Mikkeli University of Applied Sciences
Patteristonkatu 2
P.O.Box 181, 50101 Mikkeli
mirja.loponen@mamk.fi

ABSTRACT

In this paper, we are going to describe a modernization, migration and archiving of a research register. This register contains demographic information about people in the ceded Karelia region of Finland. The migration to a database started in the 80's and now the database will be modernized. The objective is to make the modernization in a way that the data can be accessible via interfaces and devices. This allows more opportunities in the future.

Categories and Subject Descriptors

H.2.4 [Systems]: Relational databases

H.2.8 [Database Applications]: Scientific databases

H.3.5 [Online Information Services]: Data sharing and Web-based services

General Terms

Management, Performance and Standardization

Keywords

Migration, modernization, historical data, parish registers.

1. INTRODUCTION

The Karelia database foundation has been storing parish registers from the ceded Karelia of Finland since the 1980's. This data contains information for scientific research of Finnish family structure, migration, education and infant mortality. The parish registers have preserved to the Karelia database for information governance and are maintained by the Karelia database foundation. They are operating in connection with the provincial archive of Mikkeli. Now the database will be migrated to the archiving services of the Mikkeli University of Applied Science. The data will be utilized in education and scientific research.

The original database is made with dBase, an old database server. This system has too many disadvantages for preserving the data for the future. The objective is to migrate the data to a cost-effective and an open source environment. The migration will be challenging and it will have to be designed in a way that information governance is more effective.

The modernization will continue in the future and it will open opportunities for more advantage usage. The researchers and projects will get easier access to the database. The users don't have to be in one place to find data from it. The data can be linked

with other scientific registers for complementary data. In addition, the data can be updated in real time by the project operators.

2. THE HISTORY OF THE DATA

In Finland, there have been Lutheran parish registers since the 17th century. At that time, Finland was a part of the Swedish Empire and the church of Sweden had defined a law which involved parishes to keep registers. The registers had information about people living in the parish [1]. The templates and instructions for the registers came from Swedish church and priests used Swedish for writing them. This was a normal custom since Swedish was the language of literate people. An interesting feature was that the registers were huge books and the priests made the records by their own knowledge. This is why some of the information might be very unclear.

2.1 The parish registers

The parish registers are divided into three categories, confirmation registers, migration registers and history registers. The confirmation registers include information about people's Christian lives for example how a person knew the bible and how many times this person had taken part to Eucharist. The confirmation registers were written into two different books [1]. Under 15-year-old parish members were in the children's register and after a confirmation they were transferred to the confirmation register [1]. This meant that the individual had become a full member of the church. The Karelia region also had the confirmation registers of the members of the Greek Orthodox Church. This was because the Karelian people had lived close to Russia for centuries. Both religions, Lutheran and Greek Orthodox had influenced this area strongly. The migration registers have information about how people had moved from parish to parish. This information was filled every time a new member came to the parish. There might have been more than one person migrating at the same time. The priests usually wrote the name of the informer and the number of migrants. The information in history registers is more statistical. These mainly had information about child births and baptisms, deaths and burials, also banns and engagements. However the information in the history registers can change century by century. The priests didn't keep infant mortality significant before the middle of the 19th century for example.

The parish registers were a good way of having information about people. The Russian Empire used the registers for collecting taxes after incorporating Finland in the year 1809. Finland was an

autonomic part of the Russian Empire over 100 hundred years. This allowed the Finnish have their own government and they were able to use their native language.

After a peaceful time as a part of the Russian Empire, the events changed in the beginning of the 1900's. Finland saw a change of getting independent, mainly because of the revolution in the Russian Empire. Fortunately, the Finnish got their independence in the 1917.

Finland was independent only a few decades when new events started in Europe. These events sealed the destiny of the Karelian parish registers. In the year 1939 the Soviet Union declared a first war against Finland. This war lasted only few months and peace was made. A second started after a couple of months. Finland suffered two wars against the Soviet Union and the second ended in the 1944 [2]. Because of these wars Karelian people had to move farther from the Soviet border [2]. This affected also to the parish registers and the priests decided that they had to be transported to more inland for safe keeping. After the second war, the Karelian region were lost and it incorporated into the Soviet Union. Most of the registers were transported to Mikkeli to safe but some of them remained to the Soviet Union [2].

3. OPENING THE DATA

The modernization of an old data has been an ongoing process throughout the years. The Lutheran's parish registers were first microfilmed in the year 1986. In general the microfilming of various historical data started in Finland in the 1920's [3]. The main purpose of microfilming was to get materials into very small size and the Karelian parish registers were filmed into microfiches which were only readable by a magnifier. Microfilming enabled better access to the registers and minimized scuff of the originals. When the technology evolved these microfiches were scanned and the pictures were added to the digital archive, where they are accessible via internet.

The idea of storing the Karelian parish registers into a database started from Raimo Viikki, the currently retired director of the provincial archive of Mikkeli [4]. The first project started in 1985 and it was a research project. It inspected of how the database should be build and were the Karelian registers suitable for this kind of database migration. In the 1988, a few programmers were hired to carry on the project to the next level [4]. They made a test program and a manual for it. The program was tested with one parish register which was named Lavansaari. The results of the project were very successful. Due to these good results, the Karelia database foundation was founded in 1990 [4].

3.1 Storing the data

First the foundation hired some staff members for storing the data into the database. The data were saved very faithfully from the original source, row by row as it was written. The work was very slow with few workers. Fortunately, the foundation got more workers when they made a contract with Finnish public employment and business services in the year 1995 [4]. This did the saving work faster and long-term unemployed people got work for a little period of time. The co-operation is still going on and new operators are hired every year. The operators are working from home or in the office of the foundation. They are working 5 hours a day and 25 hours a week. Normally, they can make a suitable schedule for themselves [1]. The co-operation project has helped both sides. The long-term unemployed people are getting socially meaningful work and it is preventing them from social

exclusion. In addition, the foundation is getting employees for storing the data.

One operator can be working in the foundation for one year. The project has had about 500 operators in these 15 years. They are storing about 500 000 - 600 000 rows of data in a year and 90% of the work has been done by the operators [1]. Of course the turnover of the operators and teaching the new ones is reducing a little bit of the speed.

The saving process is going in a way that the operators are reading the data from the microfiches of digital archives of the national archive service of Finland [4]. They use an installed program for saving the data and it is done parish by parish. All the register types have their own programs, also including the Orthodox confirmation registers. The programs are making database files from the saved data and the files are saved in to the root folder of the software. The process is continuing in a way that the operators are saving the database files to pen drives and delivering them to the office of the foundation, or sending an email with attached files. The delivered data will be saved into the final database by staff members. In some cases, the operator might have some difficulty of reading the registers, so the staff member will help to fill the missing fields. After everything is clear, the data will be migrated into the complete database.

At the moment there are over 9 million rows saved and it is estimated that in the end the database will have more than 11 million rows. It was also estimated that the work would have finished by today but the early 18th century confirmation registers are still left. The speed has reduced very much since the confirmation registers are written with unclear handwriting or the language itself isn't understandable as we can see in the image below. These unsaved registers also include the Orthodox registers. The foundation has had difficulty of finding suitable people for saving the Orthodox registers since they are written in old Russian and these registers are saved with the Latin alphabets and not with Cyrillic. It is quite possible that the database might stay incomplete because of these difficulties.

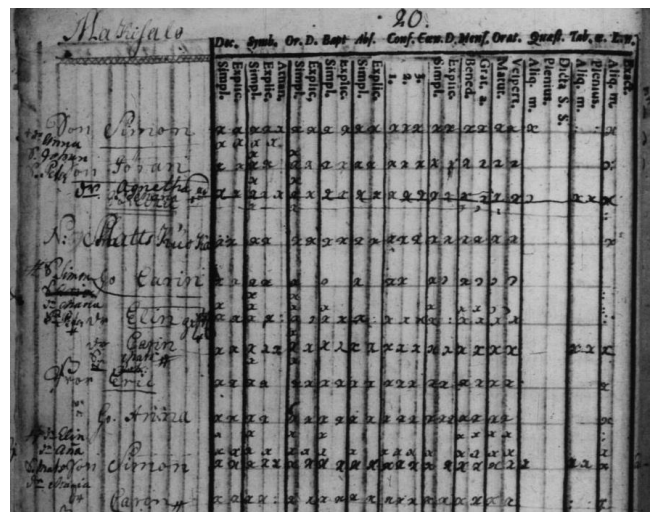


Image 1. A register page.

4. THE DATABASE

The original database is dBase -server and mostly kept in the hard drives of computers and CD-ROMs. The dBase was a cheap and efficient choice in the 80's, it suited this kind of research register. It is basically a spreadsheet type of a database and there hasn't been any user interface. The backup of the database has been handled with floppy disks and they are still in use. This hasn't been very agile for future development. The server is causing a lot of work and managing is very time consuming.

First there was one database but this changed during the years. The database had to be divided into two different versions. This solution was made when the database was accessible via internet. The Finnish law has defined that personal data will be public when it is over 100 years old from the birth. However the data of deceased people will be public in 50 years from the death [4]. The foundation decided to solve the privacy issue by dividing the database in two. The public data was accessible via internet and private data via the provincial archive of Mikkeli.

4.1 Use of the database

The first access to the database was on the computers of the provincial archive of Mikkeli. This access was mostly for the researchers and use of genealogy. In the year 2008 the foundation released a web service, called Katiha. The data was migrated to a MYSQL-server from the public data of the original database. It also has statistical information from the registers such as child births. Still, the complete database could be seen in the provincial archive of Mikkeli and this choice was more suitable for that time.

To get access to the complete database, researchers can apply it with signed application. A staff member of the foundation will accept these applications and when they are approved the researchers can go to the provincial archive of Mikkeli to browse the complete database. Otherwise Katiha is a very convenient tool for browsing the public data and find statistical information.

5. MODERNIZING PROCESS

The original database will be migrated to a new server managed by the Mikkeli University of Applied Sciences. In addition, it will have a web service. The purpose is to make the database accessible in the future. The foundation will be disbanded when the data is stored and this might happen in a year or two. Some actions have to be done before this happens. It is natural to migrate the data to the Mikkeli University of Applied Sciences. The original parish registers are still in Mikkeli and the university has had other co-operation projects with the provincial archive of Mikkeli. In addition, the university has a good knowledge about migration and archiving data, also utilizing scientific data on education.

5.1 Migration to a new server

To make the database cost-effective, it will be migrated completely into MariaDB which is open source SQL-server, developed by MariaDB foundation from MySQL. Basically, it is more developed version of the Oracle's MySQL-server [5]. The original database has over 2000 tables. Every 70 parishes have their own tables. In the figure number one, we can see the structure and how the tables are linked together. This was done to make the search faster. In the new database, these tables will be joined to around 30 tables. Also the column names of the database will be translated into English. This change will be more

informative for developing the new web service and making an interface for other services.

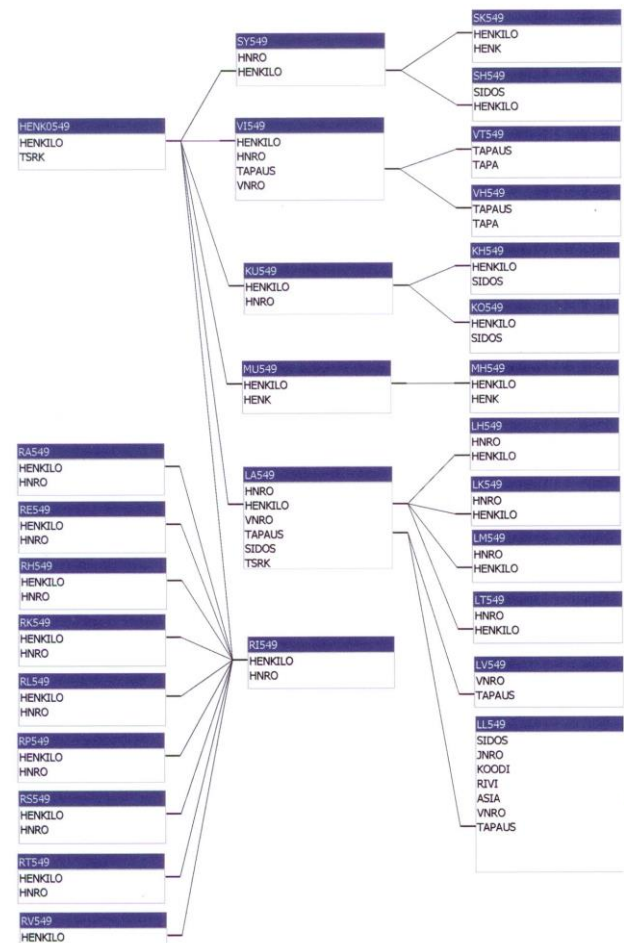


Figure 1. The table structure of one parish.

In the modernization process the database will also have other tables, such as temporary tables and metadata tables. These tables include data which is helping the access to the content. The Parishes-table has parish index and parish name for example. This table is connected to the register tables with the parish index. These metadata tables are also used for listing information in the user interface. This way all the data doesn't have to be queried at once.

5.2 Using new and faster techniques

We are using an open source search platform called Apache Solr. It is making the search faster from a large amount of data. The platform is written in Java and it runs on standalone full-text search server [6]. It has Lucene Java library as a core and the data can be output as JSON which is a lightweight data-interchange format. Humans can read this format easily and machines can parse and generate it. This allows Solr to be used with many programming languages. Solr will power up the search and it can be configured easily for any kind of data [6]. In addition, the data can be indexed to Solr straight from the SQL-server using simple SQL-scripts. With Solr indexes the data can be accessed fast and then connect to the database for more information. In the Solr-

server, there will be only an indexed data such as names, parishes, years of birth.

The new modernized database will be used through a web service and the service will connect to MariaDB-server with JDBC connection. This is a Java-based API which defines how a client can access a database. This API is used to get the full information from the database. The JDBC connection is also used for saving the data to the database. This access is only enabled to the operators and administrators. The service will have user accounts and the access rights are defined that way. Although there are accounts, everyone has access to the search and it can be used for browsing the public data.

Now everything is handled in the same place. Storing the data can be done with the service. The researchers can find data from it. The administrators are only creating users to the service. An operator will get account for managing the data. A researcher will have account after the approval of the application. With the account he or she can have access to the private data. However, the researcher has to go to the provincial archive of Mikkeli to get access to the private data. This is made to make sure the data doesn't get in to wrong hands.

This new database and web service are designed in a way that information governance is going to be easier. It is migrated to a server which uses MariaDB, an open source SQL-environment. Web service is done with Java. These will give the register a chance to survive for the future generations in a format that can be read and modified easier. When, the database is modernized there are possibilities to use it in other projects and develop an interface for versatile access.

6. PROSPECTS

The work on the foundation will finish in a few years. This enables that the data structure can be modified in to more lightweight. In addition, there can be experiments of making the data anonymous. This would help to open the data for everyone. Due to some of the registers remained on the other side of the border. This would be a good base for an international project which could make the database fully complete.

The parish registers are only one part of the information from ceded Karelia. There is more information in personal records and associations. Some of this information has been digitised and there might be a chance to get access to it. Linking all this data would be important but it would need new projects and help of voluntary work. Some plans have already been made.

1. Integrating geographic information from different eras to the data.
2. Connecting the statistical information with the geographic information.
3. Integrating different kind of data to the database, such as photographs, drawings and stories.
4. Connecting the data with the information in the social media.

A lot of Finnish population moved to Australia, Canada and United States during the wars against the Soviet Union. There could be chance to integrate the data globally and get other than Finnish linguists and scientists interested in the data.

7. CONCLUSIONS

During the modernization process it was interested to see the course of the register's life. At first, the research register were in huge books. Then these books were digitised into microfiches. Finally, the research register was stored to a database and now there are more developing opportunities.

Today there are a great number of devices but still the most important thing is the information. The 18th century priests wrote the data by not knowing what will happen in the future. Although they linked the data to dates of birth and residence. They made metadata by accident. With this information there is a question to be thought. Do we have to think other than this metadata for linking the data to other sources? There are the timestamps and places which can be used, or are there?

8. REFERENCES

- [1] The Karelia Foundation. 2014. <http://karjalatk.fi>.
- [2] Ripatti, Jarkko. *Karjalan luterilaiset seurakunnat evakossa*. Paino Livonia Print. 2014.
- [3] Lybeck, Jari. *Asiakirjahallinnon oppikirja*. The National Archive of Finland. 2006. Helsinki
- [4] Vartiainen, Minna. 2013. Karjala-tietokantasäätiön Karjala-tietokanta: Tietokannan haasteet. *In archiving seminar 2013* (Finland, May, 2013)
- [5] MariaDB foundation. 2014. <https://mariadb.org/>.
- [6] The Apache Software Foundation. Apache Lucene. 2011-2012. <http://lucene.apache.org/solr/>.