“BABEȘ-BOLYAI” UNIVERSITY FROM CLUJ-NAPOCA
FACULTY OF HISTORY AND PHILOSOPHY

PhD THESIS
SUMMARY

COORDINATOR
Res. Ass. Prof. LUMINIȚA DUMĂNESCU

PhD CANDIDATE
ANGELA CRISTINA LUMEZEANU

CLUJ-NAPOCA
2019

“BABEȘ-BOLYAI” UNIVERSITY FROM CLUJ-NAPOCA
FACULTY OF HISTORY AND PHILOSOPHY

DIGITAL INFRASTRUCTURE FOR SOCIAL HISTORY.
BUILDING HISTORICAL DATABASES

COORDINATOR
Res. Ass. Prof. LUMINIȚA DUMĂNESCU, PhD.

PhD CANDIDATE
ANGELA CRISTINA LUMEZEANU

CLUJ-NAPOCA
2019

Contents

Keywords: Transylvania, relational databases, historical population database, source-oriented model, method oriented model, Historical Population Database of Transylvania (HPDT), Historical Data Grinder (HDG), War Invalids, Orphans and Widows Database (IOVR), Probate Database of Transylvania (PDT)

INTRODUCTION

The present work aims to provide historians in Romania with an in-depth and applied discussion of the tenets of historical databases construction, with a focus on databases involved in demographic and social-historical research. This overview will maintain a widely intelligible approach, without awarding too greater emphasis to the technical aspects of database construction, but rather insisting upon the structural and conceptual issues that historians (like researchers in the humanities, in general) are not always equipped to deal with. Such an overview is currently direly necessary, when an increasing number of databases, online document portals and repositories, or projects with a digital component emerge almost yearly. Meanwhile, the level of general knowledge concerning the process of translation of historical source to data (in database form) remains at best, basic, a fact which hinders this development. Database models and the principles underlying them remain understood only partially, and therefore used only to a fraction of their real potential (Erickson 2013, 134). To remedy this situation, works such as the present one aim to offer applied, in-depth, and accessible discussions of what database construction involves, and how such an endeavor might transform the research fields for social, demographic, or even political historians focused on modern Transylvania.

The present work arose from two different needs: the first, already noted, was that of making current database conceptualization known in an accessible form to a new category of beneficiaries, namely historians; the second motivation for this work is closely related to this, and concerns the necessity of documenting the construction and development of the Historical Population Database of Transylvania, the most complex of project of creating digital infrastructures ever implemented in Romania to date. I will discuss first the structure of the work, and then provide a brief and more technical overview of the language and environment which were used for the construction of all the databases documented in this thesis. The latter section is included as a final part of the introduction in order to highlight the functionalities and characteristics shared by all databases. This structure is also meant to keep separate the technical aspects to the database design, which might be of interest to a more specialized audience, from the purely historical and descriptive/analytical sections to follow.

The work is divided into six main chapters, which address both broader, theoretical or conceptual issues, as well as the application of the principles of database design to the construction of various digital infrastructures.

The HPDT (Historical Population Database of Transylvania) was contains data drawn from parish registers of baptism, betrothals, marriages, deaths, change in denomination, confirmation, blessings, vaccination. The HDG (Historical Data Grinder) is also based on archival or edited sources in a more narrative format, like press, correspondence, memoirs, yearbooks of institutions and so on. The IOVR (War Invalids, Orphans and Widows) Database harnesses its information from the files of individuals registered for receiving pensions, which include a large variety of documents, among which the main types were the following: civil status papers, reports of personal characteristics of enrolled soldier, information about war participation, testimonies proving the disappearance/ death, medical records, data regarding the situation of the beneficiaries' family and the amounts of money they received and pieces of evidence attesting changes in the situations of pension beneficiaries. The PDT (Probate Database of Transylvania) contains testaments and probate registers. For an adequate understanding of the complexity of databases each chapter of this work is introduced by a complete description of the sources used for building the databases.

The historical sources are complemented by the secondary literature on how to build and manage historical databases. The databases for HPDT, HDG, IOVR and PDT were built in MySql, an Oracle relational database management system (RDBMS) based on Structured Query Language (SQL). MySQL is the most popular open source database, as it also runs on multiple operating systems including Windows, Linux and the Unix platform. In most cases MySql is associated with web applications and was developed to manage large databases very quickly.

The web applications used for creating the User Interface are the Ruby language and the Ruby on Rails framework. Rails is an open source *model-view-controller* (MVC) framework that allows the developer to write less code while accomplishing the same goals (the HPDT has almost 71 000 lines of code).

All the applications were deployed on an Ubuntu server using Ruby Version Manager (RVM). RVM helps managing different installed versions of ruby and rails, while each application has its own package of small routines that need to be installed. The connection between the applications and the Apache web server is made through Phusion Passenger, a powerful web and application server.

CHAPTER 1. Databases in Historical Research: A Transformative Instrument

The first chapter of this thesis was dedicated to an overview of the existing databases, with an emphasis on those exploiting historical sources. While databases can serve many needs in historical research, we can point out a number of functions in which the historian is always interested in, like the management of data, record linkage and data analysis.

Kees Mandemakers and Lisa Dillon have created a guideline with the best practices and stated the importance of selecting the right kind of sources from which to draw a dataset to be included in a database, emphasized the means of data input, their integration, storage, and dissemination. They recommended that the sources to be transcribed literally in the computer, creating a digital clone of the written document. Further on, there should be a database for the sources, one for the standardised data, and a third one for public release. The standardised database allows corrections to the data, cleaning and coding the values, as well as the integration of different data sources. This type of database is called in literature *source-oriented database.* Its advantages are clear: it offers a very large variety of research themes, the data are reliable and can be counted, linked, studied to a high level of detail. But it also has its limitations. If there are limited resources or the time is not enough at some point some information must be excluded which is contrary to the principles established by Mandemakers & Dillon. (Mandemakers & Dillon, 2004)

Another model in use is the so-called *method-oriented database.* It is easier to design and build, and the data collection takes less time. For this kind of approach, it is essential to know from the beginning what is the research question, and even what kind of queries to run. Once the design is in place it is very hard to diverge from the functionality of the database, for example to pursue another research question.

For the historians who deal with the study of historical population, the source-oriented database remains the best approach. To include a full population from a region into the database can however take a lot of time: some of the historical population databases were started decades ago and are still under development. Most of the databases focused on demography and population studies are centralised online on the European Historical Population Samples Network (EHPS-Net), created in 2011 to bring researchers from all over the world together in order to create a common format for the various databases on population. Among them, we can count the HPDT.

CHAPTER 2. What is a database? Basic Tenets of Database Functionality

The second chapter is devoted to the models used for database construction, and the means of data storage and data access. Both are relevant for the models employed for the Romanian databases that I have created. During the research developed for the present PhD Thesis I have noticed that there are a lot of misunderstandings and misuse of the term database. It was one of the present work's aims to provide historians with basic applied knowledge of what precisely constitutes a database, and how historical sources might be translated into the relational database format. There are a multitude of file types and mediums in which data can be stored. However, in order for them to be considered databases, they must provide not only storage of data, but also of the relations between the bits and pieces of information included in a historical source, the ties provide them with meaning. Moreover, a database needs to provide efficient ways of accessing and retrieving data (insert, update, delete, and visualisation).

The relational database model was first theorized by E.F. Codd in a paper published in 1970, when he developed a series of basic rules. According to his rules, a database is a "well-organised collection of data that are related in a meaningful way, which can be accessed in different logical orders" (Codd, 1970). All the raw data are stored in an organised manner, but the information can still be immense. There is a software program known as *data management system* (DBMS) that manages the database and translates between user requests and the data storage. Numerous DBMS are available today, like Microsoft Access, FileMakerPro, Oracle, Mysql, PostgreSql, DB2 etc. The DBMS need to have the following capabilities: to provide a way to enter, modify and delete data, to find a way to retrieve the information, e.g. the queries, and should offer restricted access to data by creating users and passwords.

The graphical database is an online DBMS that works on a graph data model. A graph data model consists of two elements: nodes (entities) and relationship. A node is an entity (person, place, category, profession etc.). Each relationship represents how these nodes are associated with each other. This kind of database offers a great flexibility especially for adapting documents, texts and literature to fit into a system. The graph database is not suitable for large amount of data and for handling queries that covers the entire database.

A spatial database is a collection of information that is optimised to store objects defined in a geometric space. Usually these are associated with geographic locations, topographic features etc. Data for spatial databases are stored in lines, polygons, points, coordinates.

I have emphasized the details of a relational database, where objects are organised into a collection of relations which are depicted in a tabular form. A *schema* has mainly two components: *Entity* and *Relationship*. *Entity* is an object for which we want to store the information into the database. There are usually easy to recognize like Clients, Students, Employees. They are correspondent the tables in the relational model. *Relationship* define how the entities stored in the database interact with each other. In general the database is organised in several tables with columns and rows identified by a unique key. Each column represent an attribute of the entity. The logical connection between these tables` rules is the relation: each table represents a relation; each column on a row has a single value; every row has a unique identifier called primary key; there is no identical column (they need to have different name); all the entries on a column have to be on the same type.

There are three types of relationships that define the connection between entities: one-to-one, one-to-many and many-to-many relationship. Relational database model normally do not permit to apply directly the many-to-many relationship. In order to avoid confusion, the relationship is divided in one-to-many relationship with the help of a third table, usually called a *join table.* A join combines two or more tables from the database based on a common relationship (foreign_keys). An important factor that affects the performance of the database is the join operation, they are the connecting factor between the tables in a relational model, everything can be retrieved if the join statements are correctly applied. One of the processes used in building a relational database is normalisation, in order to avoid redundancy.

When developing a database, it is very important to follow a stepwise cycle of activities. Jan. L Harrington referred to this as the *structured design life cycle,* and it includes the assessment of needs, the design of the database, and the development and implementation of the database. After the implementation there is a testing phase and then the cycle repeats itself until the database reaches the optimum configuration.

CHAPTER 3. Historical Population Database of Transylvania

The third chapter was dedicated to the **Historical Population Database of Transylvania (HPDT)**. HPDT was born as part of a project with the main goal of building a population database covering the period between 1850 and 1914 in Transylvania. It is the largest chapter, because it deals with the first historical population database created in Romania. The overview includes a discussion of the HPDT's three separate databases and its many versions, as it can serve as a model for other databases implemented in the Romanian historical field.

The chapter opens with an overview of the sources used for the database creation. The main sources are vital registrations (baptisms, marriages, burials) which were filled in and kept by the church. As a characteristic for Transylvania is the fact that the church registers were written in several languages (Romanian, Hungarian, Latin, German etc.) with different alphabets (Latin, Cyrillic) and scripts (Kurrentschrift etc.). The registers are also from diverse denominations: Orthodox, Greek-Catholic, Catholic, Reformed, Lutheran, Jewish etc. These proved to be a challenge for the database architect and for the operators, since the database had to be permanently adjusted to reflect the sources. Parish registers are some of the most widely used sources for historical demography.

The information in the HPDT respects with great strictness the original sources, following the source-oriented model, even though during the data entry process the operator might have noticed various errors of writing, registration, or simply omissions which could have easily been filled by logical deduction. Thus, the information from the database faithfully follows the character of the source to the letter, with all its potential errors or alterations. Wherever needed, the operator noted in a separate rubric the observations or the issues related to any particular situation. The entire database is in English (including the observations made by the operator), but the information *per se* was transcribed in the language of the original registry, without diverging from the source.

The third chapter of the thesis then analyses the dataset construction and follows with the description of the triple database that HPDT encompasses: the source database, the standard one, and the open access one. From the beginning the database was developed to have a research component and a public open access one. The research database, available only by user login has a source database, a standard database and a table with linked individuals. The public database (http://hpdt.ro:4080) is an open access website that offers to the broad public an insight on the dynamics of Transylvanian population from 1850-1914. The information presented in the open database is part of the source database.

All database components of *Historical Population Database of Transylvania (HPDT)* are created in the relational database model using MySql database management system, an open source framework and have a User Interface created in Ruby language program and Rails framework. The user does not have a direct connection to the database itself, but it relates to it through the User Interface that enables the insert, edit and deletion of records. It follows the best practices model for constructing the population databases.

Each table details a single type of vital event from the perspective of the religious ceremonies accompanying it: Baptism (*Tbirths*), Engagement (*Tbetrothals*), Marriage (*Tmarriages*) and Burial (*Tdeaths*). In addition to the three main tables there are several complementary tables, related to each other by a corresponding column in order to handle complex queries, to search and find data in a logical manner and to eliminate redundancy. The logical connection between the tables is made by the foreign key. To the foreign keys are added constraints, functions that ensure the database will never have orphaned records in a child table.

All the additional tables are related to one or all 4 of the main tables by one column or several columns. There are 24 look-up tables that help with maintaining the information in a normalized form. Presently, the *Tbirths* table is connected to 21 complementary tables, *Tmarriages* has 44 connections, *Tdeaths* has 20, and *Tengagements* has 7 connections. Some of these connections are made to the same secondary tables.

In order to maintain a balance between the principles of best practices in historical databases and the principles of a good relational database separate tables for Priests and

Midwives were created. To eliminate redundancy, the table *Tpriests* records the priest just once and is linked to the tables of major events by a foreign key – *priest_id*. Thus, it manages to eliminate the manual entry of the same priest in thousands of rows. The table for midwives – *Tmidwives* - is constructed in exactly the same fashion. The table is linked to the Baptism table by a foreign key – *tmidwife_id.*

There is a table that documents the original source (*Tsource)* – where it can be found, the language used, the alphabet, the parish and denomination, the dates of the first and last record.

Next, the third chapter delineated the Standard Database. In order to have a better quality of the Record Linkage process the recorded information was first standardized. The tables from the Source database were duplicated in another database called Standard Database.

In addition to the tables originating from the source database several tables have been created. The most important table is the table for names (*Tnames*). Standardization has been done for first name, last name and nickname. A second independent table is *Tlocation,* which contains all the localities mentioned into the database with the standard correspondent.

Another standard table is the occupation table – *Toccupation.* All the occupations mentioned in the database were added to a table and encoded using the Historical International classification of occupation HISCO.

After the standardization takes place, then the record linkage can be implemented. Record linkage is the process of finding datasets that are referring to the same entity across multiple sources, to find double records and merge them Thus, every piece of information about the entity is bound together in order to have a complete record of that entity. It is a method of standardizing information which can be found in sources that do not share a common denominator (id, national identification number, social number etc.) It has begun to be regularly employed in historical research dealing with big data, especially in the past decade. In Romania, the record linkage process had never been used for historical research, as until now there was no database for historical population large enough to permit this kind of research. The building of HPDT created the necessary structure for trying the methods of record linkage.

The algorithm for record linkage is based on the mathematical formula of calculating the Jaro-Winkler distance. This formula measures the similarities between two strings of characters (names, places, occupations etc.). In order to apply the automated Record Linkage a new table was created – *Individuals.* The table was populated with data extracted from the three event tables of the sample.

The open access database was presented in the last part of the third chapter. The Historical Population Database of Transylvania (HPDT) also has an open access component available to the wider public (http://hpdt.ro:4080). It contains all the events recorded in the original database regarding Baptisms, Marriages and Burials. Not all the columns from the original database were displayed, but rather only a selection deemed relevant for non-specialist users.

Designing and building a database, especially when using a Source oriented approach is a time-consuming task. When working with complex sources, with heterogeneous information the task will prove extremely challenging, although some of the results are quite promising for the historians with a focus on population studies. All the results, and especially the record linkage process depend very much on the quality of sources. Transylvania has a unique position where the different denominations offers a variety of sources with a multitude of information fields. The problem occurs when all this variety must be integrated in source-oriented model.

CHAPTER 4. The Historical Data Grinder

The next chapter of the thesis was dedicated to the development of the next historical database, namely the **Historical Data Grinder.** Its primary merit it is that it acknowledges the different research systems employed by historians, therefore offering an alternative to the relational database model.

The historical sources include a high number of biographical and prosopographic instruments, like books, dictionaries, various lists of names, schematismi. Unfortunately, they are available only in printed format or in electronic versions, therefore they are not so easily accessible, and they do not encourage comparative approaches.

Considering these factors and the need of a common framework where the information from multiple sources can be stored and connected with the already existing databases, we believe that the best suited model is the Entity-Attribute-Value (EAV). The EAV model is also in accordance with the computer knowledge of the Romanian historians.

The main advantage of the EAV system is its flexibility. The simplest form of this design contains only three tables for entity, attribute and value, and the metadata table. Adding a new attribute to the entity does not require creating new columns and redesigning the database by the programmer. It is a simple task of adding another row to the entity table which can be done by the researcher. In the same manner the attributes are included in Attribute table. The data format is very clean, the relations are very clear, and it is much more suitable when new attributes need to be stored in the database on a regular basis.

For the historian, the main challenge is understanding how the EAV design works. We created a model named *Historical Data Grinder* (HDG) where the information is decomposed according to the EAV design. There are two ways of recording the information. First the information can be processed offline. The historian can organize the data into spreadsheets following the design of the database entry form. Then the information is imported into the database in bulk through .csv files. The second method is by inserting the records one by one into the database using insert/edit form. This requires a permanent internet connection and more time.

In the current structure, the database can accept theoretically any kind of historical information, from the reconstruction of family relations in small circles or in local or geographical clusters, to the reproduction of the evolution of administrative changes, with their modifications in time and even with the related population if desired (the latter in turn divided into categories as needed). It has an increased ability to act both as an aggregator, and as a stand-alone database.

The graphical interface of HDG lists the records in a tabular form. Through the web interface, the user can also download records in .csv or .xls file formats based on search filters in order to extract only the information needed for research.

The Entity-Attribute-Value (EAV) model fits much better with the needs of the historians who use various sources and need to integrate them into a complex analysis. Flexibility is the main advantage proposed by the EAV model, while the main challenge is understanding how the EAV design works and how to join multiple data into a coherent model that allows complex analysis.

CHAPTER 5. War Invalids, Orphans and Widows Database: IOVR

The fifth chapter was devoted to the **War Invalids, Orphans and Widows Database**. After the end of the First World War, in order to support the affected people, the transitional authorities of Transylvania and then the central ones in Bucharest issued a series of regulations and laws regarding the invalids, widows and orphaned children. The files of people registered for receiving pensions contain a variety of documents the main type being: civil status papers, reports of personal characteristics of enrolled soldier, information about war participation, testimonies proving the disappearance/ death, medical records, data regarding the situation of the beneficiaries' family and the amounts of money they received and evidences attesting changes in the situations of pension beneficiaries.

The database was structured into three main tables (Invalids, Widows, and Orphans), two secondary tables (Children, Events) and five auxiliary tables (Counties, Places, Relation, Ranks, Units). Three tables are used in common with HPDT (Denomination, Marital Status and Occupation). There other three joint tables are connecting the Occupation table with the main tables in order to create a many-to-many relationship enabling the multiple selection. The IOVR model has presently 16 tables and it is connected with HPDT database, but it can also be separated and built as an independent database if necessary.

There are entry forms for the invalids, widows, and orphans. In the distant future, we hope that the data can be linked to information from the Historical Population Database of Transylvania (HPDT), which would allow the identification and tracking of individuals in the past of Transylvania, both before and after the twentieth century wars.

CHAPTER 6. The Probate Database of Transylvania

The sixth and final chapter focuses on the **Probate Database of Transylvania (PDT).** Built with a different purpose in mind than both the Historical Population Database of Transylvania and the Historical Data Grinder, PDT was designed to host information from several well-defined types of sources dealing with social, professional, and economic attributes. While similar to the HPDT in conceptualisation (source-oriented rather than research-oriented or EAV), it nevertheless adopts a more fragmented approach to the informational content of the sources it breaks down than in the case of parish registers.

The purpose of the Probate Database of Transylvania is to collate data at individual level from several types of social-economic and juridical sources. Romanian research has been late to discuss probate inventories and estate divisions, or fiscal conscription/taxpayer lists as sources.

The Probate Database of Transylvania is currently structured into twelve tables, and it is hosted on the same server as the HPDT, with which it shares the design of the General User Interface. Like the HPDT, it was also constructed in .sql, its architecture evolving to cope with the variation in the sources. Nevertheless, as opposed to the HPDT, it does not yet follow the original source with faithfulness, instead breaking it down into its main components.

The structure of the Probate Database of Transylvania currently follows two major pathways, determined by the two types of sources it incorporates. Each type of sources has its own source table – *Wills_Sources* and *Register_Sources* respectively. The *Individuals* table is also linked to several other smaller tables beside the *Civil Status* table, namely the *Ethnicity,* the *Role,* and *Profession* tables. Another table that deserves mention is the *Relation* table, in which the relations between individuals mentioned in the documents entered into the database are codified.

While it adopts certain principles and guidelines from both the source-oriented model and the EAV model, it departs from both of these in certain aspects: only main entities such as individuals, divisions, or last wills and testaments are singled out as such, while other sections follow the source more faithfully, given the great variety of attributes and values present in the sources.

Because the linkage process will ultimately involve primarily individuals and pieces of real estate (to the extent that it is possible), information is entered gradually, focusing primarily on collating data on the persons noted in the sources, where they owned property, and what it was worth. Although this contravenes some of the principles on database construction of historical population databases, as outlined by Mandemaakers and Dillon (2004), it is much easier and less time-consuming than the faithful reproduction of the historical sources used in their entirety, prior to any kind of standardisation.

CONCLUSIONS

The fast development of information technology (IT) has had considerable impact in all scientific fields. In the past decades, this impact has begun to make itself felt in the humanities as well, with the crystallization of sub-disciplines such as social science history or the emergence of journals such as *Historical Computing* or *Historical Methods.* When historians as a branch became increasingly familiar with the computer, a lot of information that was otherwise processed manually could be analyzed automatically, thus freeing time to explore a myriad of research techniques and theoretical orientations. The research job became a lot easier especially when it comes to storage space, easy ways to access and retrieve data, allowing the historian to perform complex analyses on large quantities of data in a matter of minutes.

Nowadays, we cannot imagine our life without a computer. The information we gather as researchers is ever-growing and requires a platform in order to be studied and shared among scholars. Therefore, all the information must be organized in an efficient manner that enables the researcher to manage, process and share the results. The best solution is a data management system – database – which has experience wide-scale development due to the necessities of research. The most common used system of this type is the relational model.

Since Codd theorized the first relational database model (Codd, 1970) the use of databases has increased exponentially. Now the use of databases is part of everyday life, having also changed the concept of research methods. The creation of vast databases by researchers, as well as the use of modern statistical techniques, and the focus on the social foundation are defining traits for the new history. The development of databases is central in the transformation of methods in research.

All the information from the above-presented four Romanian databases will be at some point in the future linked and inter-connected. All of them display a conceptual model that deals with different kinds of sources, reflecting the flow and content of actual historical research, where information must be drawn from multiple directions. What the constructions of these digital infrastructures has shown is that the quality and heterogeneity of the sources is the main feature dictating the architecture of the database, beside historians' particular research questions or the costs and benefits of a method-oriented vs. a source-oriented approach. The experience of building the HPDT, the HDG, the IOVR and the PDT databases has amply demonstrated that the ideal solution for Transylvania is a remodelling of classical source-oriented database that accommodates the sources and combines the standardization of data with the original information.

Moreover, the experience of creating these digital infrastructures has also shown the importance of informed communication between historians and the database developer, the first and key step in ensuring that the best possible overall solution is implemented. While neither party can be expected to be a specialist in the other's field, historians wishing to include a digital component into their research projects in the form of a database need to have at least a basic understanding of the main tenets of database construction and its functionalities, in order to structure their sources in the most appropriate manner. While the database architect's expertise with historical sources is invaluable, it also falls to the historians to translate what it is their projects need to achieve, the specificities of their particular sources, and to have a clear overview from the beginning of the amount of heterogeneity present in the future data. The present work has therefore also tried to constitute a first stepping-stone for Romanian historians interested in constructing or working with a database, basing its accessible account on current digital infrastructure projects developed in Romania.

SELECTED REFERENCES

Electronic Sources:

http://port.sas.ac.uk/mod/book/view.php?id=75&chapterid=132).

http://port.sas.ac.uk/mod/book/view.php?id=75&chapterid=133

http://s3.thinkaurelius.com/docs/titan/0.5.4/getting-started.html

http://soc.kuleuven.be/web/staticpage/6/30/eng/819

http://www.cedar.umu.se/english/ddb/public-access-databases/

http://www.censusmosaic.org/data

http://www.karjalatk.fi/index.php?lang=en

http://www.rhd.uit.no/sources.html

https://katiha.mamk.fi/

https://socialhistory.org/en/hsn/index

https://www.ehps-net.eu/databases

https://www.ehps-net.eu/databases/antwerp-cor-database

https://www.ehps-net.eu/databases/historical-population-database-transylvania-1850-1914

https://www.ehps-net.eu/databases/historical-sample-netherlands

https://www.ehps-net.eu/databases/hungarian-historical-demographic-database

https://www.ehps-net.eu/databases/karelian-database

https://www.ehps-net.eu/databases/mosaic-project

https://www.ehps-net.eu/databases/norwegian-historical-population-register-1800-1964

https://www.ehps-net.eu/databases/poplink

https://www.ehps-net.eu/databases/popum

https://www.linkedin.com/pulse/comprehensive-analysis-nosql-vs-rdbms-rassul-fazelat https://www.mysql.com

https://www.ed.lu.se/databases/sedd

https://www.ipums.org/

https://usa.ipums.org/usa/

https://international.ipums.org/international/

https://www.nappdata.org/napp/

https://www.ehps-net.eu/databases/scanian-economic-demographic-database

https://www.ed.lu.se/databases/sedd

https://www.ed.lu.se/databases/sedd/sedd-public-access

https://www.studytonight.com/dbms/first-normal-form.php

https://www.tutorialspoint.com/difference-between-schema-and-database-in-mysql

http://dev.mysql.com/doc/refman/5.7/en/create-procedure.html)

http://dev.mysql.com/doc/refman/5.7/en/select.html

http://ec.europa.eu/eurostat/web/nuts/local-administrative-units

http://www.mysqltutorial.org/mysql-boolean/

http://www.mysqltutorial.org/mysql-foreign-key/

https://dev.mysql.com/doc/refman/8.0/en/char.html

https://dev.mysql.com/doc/refman/8.0/en/date-and-time-functions.html#function_date-format

https://dev.mysql.com/doc/refman/8.0/en/date-and-time-types.html

https://port.sas.ac.uk/mod/book/tool/print/index.php?id=75

https://www.britannica.com/topic/dispensation

https://www.tutorialspoint.com/mysql/mysql-data-types.htm

Bibliography

1. “Regulament pentru funcționarea Oficiului Național Invalizi, Orfani și Văduve și pentru organizarea și dezvoltarea operelor de asistență națională” / “Regulation for the Functioning of the National Office for the Invalides, Orphans, and Widows, and for the Organization and Development of the National Assistance Works”(1922). *Monitorul Oficial*, 29 aprilie: 874-901.
2. ”Historical Sample of the Netherlands (HSN)”. (2015). *Annual Report*. International Institute for Social History, Amsterdam.
3. ”Intoxicants & Early Modernity, England, 1580-1740”, Project information available at https://www.intoxicantsproject.org/publications/database/.
4. **Alter**, G., **Mandemakers** K. (2014). ”The Intermediate Data Structure (IDS) for Longitudinal Historical Microdata, version 4”, *Historical Life Course Studies*, vol. 1, 1-26, published 26.05.2014. Available at http://www.ehps-net.eu/article/intermediate-data-structure-ids-longitudinal-historical-microdata-version-4/hislives_vol_1_03_07_14.pdf.
5. **Alter**, G., **Mandemakers**, K. and **Gutmann** M.P. (2009), Defining and Distributing Longitudinal Historical Data in a General Way through an Intermediate Structure, *Historical Social Research / Historische Sozialforschung*, 3 (109), 34 : 78-114.
6. **Bouchard**, G. (1998). ”The Content of Occupational Titles: What Can Be Learned from a Population Register?”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 31(2): 75-86.
7. **Chelaru**, R.D. (2018). ”The Virtual Genealogical Archive (ArhGenVirt). A Pilot Digitization Project of the Parish and Civil Status Registers in the Bucharest and Brasov County Archives (Romania)”, *Transylvanian Review*, 27(3): 140-154.
8. **Codd**, E.F. (1970), “A Relational Model for Large Shared Data Banks”, *Communications of the ACM*, 13(6): 377–387.
9. **Codd**, E.F. (1985a), “Is Your DBMS Really Relational?”, *ComputerWorld*, 14 October.

10. **Codd**, E.F. (1985b), “Does Your DBMS Run By the Rules?”. *ComputerWorld*, 21 October.

11. **Codd**, E.F. (1990), *The Relational Model for Database Management*, 2nd ed. Addison Wesley Publishing Company.

12. **Crăciun**, B., **Holom** E.C., **Popovici** V. (2015). „Historical Population Database of Transylvania: Methodology Employed in the Selection of Settlements and Micro Zones of Interest”, *Romanian Journal of Population Studies*, 2(9): 17-30.

13. **Dillon**, L., **Amorevieta-Gentil**, M., **Caron**, M., **Lewis**, C., **Guay-Giroux**, A., **Desjardins**, B. & **Gagnon**, A. (2018). ”The Programme de recherche en demographie historique: past, present and future developments in family reconstitution”, *The History of the Family*, 23(1): 20-53.

14. **Dinu** V. & Nadkarni P. (2007). ”Guidelines for the Effective Use of Entity–Attribute–Value Modeling for Biomedical Databases”, *International Journal of Medical Informatics*, 76(11-12), 769-779.

15. **Edvinsson**, S. (2000). “The Demographic Data Base at Umeå University - a resource for historical studies”. In P.H. Hall, R. McCaa, & G. Thorvaldsen, *Handbook of International Historical Microdata for Population Research*. Minnesota Population Center.

16. **Farhat**, Y., **Martins**, M. & **Swanson**, D.A. (2014). *Methods of Demographic Analysis*, Springer.

17. **Feneșan**, C., ed. (1986). *Izvoare de demografie istorică*, Vol. 1, *Secolul al XVIII-lea. Transilvania*. București: Direcția generală a arhivelor statului din RSR.

18. **Gavrilă**, I. (2011). *Baze de date istorice. Cercetarea istorică asistată de calculator. Rezultate statistice ale comparării automate a bazelor de date istorice*. Oscar Print, București.

19. **Glavatskaya**, E., **Thorvaldsen**, G., **Fertig**, G. & **Szoltysek**, M. (2019). *Nominative Data in Demographic Research in the East and the West*, Ekaterinburg: Ural University Press.

20. **Harrington**, J.L. (2009). *Relational Database Design and Implementation: clearly explained*, 3rd Ed., Morgan Kaufmann Publishers-Elsevier, Burlington MA.

21. **Harrington**, J.L. (2010). *SQL Clearly Explained*, Third Edition, Morgan Kaufmann Publishers-Elsevier, Burlington MA.
22. **Helgertz**, J. (2014). ”Introduction to SEDD, IDS and STATA”, Lecture presented at Lund Summer School Course.
23. **Holom**, E.C., **Crăciun**, B. (2014). “Invalizi, orfani, văduve. Consecințe sociale ale „Marelui Război” în Transilvania. Studiu de caz”/ Invalids, Orphans and Widows: The Social Consequences of the “Great War” in Transylvania. Case Study”. In I. Bolovan, G. Cojocaru, O.M. Tămaș (Eds). *Primul Război Mondial. Perspectivă Istorică și istoriografică/World War I. A Historical and Historiographical Perspective*. Cluj Napoca: Academia Română/Centrul de Studii Transilvane, Presa Universitară Clujeană, 559-568.
24. **Holom**, E.C., **Sorescu-Iudean** O., **Hărăguş** M. (2018), ”Beyond the Visible Pattern: Historical Particularities, Development, and Age at First Marriage”, *The History of the Family*, 23(2): 329-358.
25. **Mandemakers**, K. & **Dillon**, L. (2004). ”Best practices with large databases on historical populations”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 37 (1): 34-38.
26. **Mandemakers**, K., *et al*. (2018). ”HSN standardized, HISCO-coded and classified occupational titles, release 2 2018.01” (IISG Amsterdam 2018). Available at https://datasets.iisg.amsterdam/dataset.xhtml?persistentId=hdl:10622/MUZMAL.
27. **Matthijs**, K., & **Moreels**, S. (2010). “The Antwerp COR*-database: a unique Flemish source for historical-demographic research”. *The History of the Family*. 15(1): 109-115.
28. **Maudsley**, E. & Munch T. (1993). *Computing for historians. An introductory guide*. Manchester University Press, Manchester and New York.
29. **Moldovan**, L. (1958). ”Registre confesionale de stare civilă din Transilvania”. *Revista Arhivelor*. 1: 159-185.
30. **Moldovan**, L. (1968). ”Registrele parohiale din Transilvania”, *Revista de statistică*, XVII, 11: 52-60.

31. **Nadkarni**, P., & **Marenco**, L. (2007), ”Database Architectures for Neuroscience Applications”, *Methods in Molecular Biology*, vol. 401: *Neuroinformatics*, 37–52.
32. **Nadkarni**, P., **Marenco**, L., **Chen**, R., **Skoufos**, E., **Shepherd**, G., **Miller**, P. (1999), ”Organization of heterogeneous scientific data using the EAV/CR representation”, *Journal of the American Medical Informatics Association*, 6 (Nov / Dec): 478-493.
33. **Popovici,** V., (2015), ”Parish registers from Transylvania – sources for the History of Medicine (late 18th – early 20th centuries)”, *Acta Medico-Historica Adriatica*, 2(13): 287-302.
34. **Räisä**, J., & **Loponen**, M. (2014). “The modernization, migration and archiving of a research register”, in *DLM 2014 [Recurso eletrónico]: making the information governance landscape in Europe / 7th Triennial Conference of the DLM Forum*; ed. José Borbinha, Zoltán Szatucsek, Seamus Ross. - Livro eletrónico. - Lisboa : Biblioteca Nacional de Portugal, Available at http://purl.pt/26107/1/DLM2014_PDF/13%20%20The%20modernization,%20migration%20and%20archiving%20of%20a%20research%20register.pdf
35. **Rassat**, S., **Ruggiu**, F. J. (2019). ”Big data, 3D models and digital heritage as instruments for demographic and historical studies, the example of the 'Charleville' project”, *In Situ – Revue de Patrimoines*, Issue 39, Available at https://journals.openedition.org/insitu/22296#text.
36. **Shorter**, E. (1971). *The historians and the computer*. Englewood Clifs.
37. **Sigmirean**, C. (2000), *Istoria formării intelectualității românești din Transilvania și Banat în epoca modern*, Cluj-Napoca: Presa Universitară Clujeană.
38. **Sigmirean**, C. (2007), *Intelectualitatea ecleziastică. Preoții Blajului (1806-1948)*, Târgu-Mureș: Editura Universității ‘Petru Maior’.
39. **Sigmirean**, C. (2011), *Formarea elitelor militare ale Imperiului Austro-Ungar. Studenți transilvăneni la Academia Militară „Ludovika” din Budapesta*, Sibiu: Astra Museum.
40. **Sorescu-Iudean**, O. (2017). „The Permeability of Borders: Drafting Romanian Last Wills in the Seat of Sibiu during the Eighteenth Century”. *Banatica*, 27: 463-484.

41. **Sorescu-Iudean**, O. (2019). “Mai mult decât suma părților. O viziune integrativă asupra inegalităților sociale, de venit și avere în Sibiul secolului al XVIII-lea prin Probate Database of Transylvania”. Conference Presentation at the National Conference of the Centre for Population Studies, „Studii de Populație. Provocări și perspective”, 1-2 July 2019, Cluj-Napoca.
42. **Sorescu-Iudean**, O., **Lumezeanu**, A. (2018). “Un model digital pentru surprinderea succesiunii în Transilvania premodernă”, Conference Presentation at the National Congress of Romanian Historians, 29 August – 1 September 2018, Iași.
43. **Sumathi**, S. & **Esakkirajan**. S. (2010). *Fundamentals of Relational Database Management Systems*, Springer, Berlin.
44. **Szołtysek**, M., & **Gruber**, S. (2016). “Mosaic: recovering surviving census records and reconstructing the familial history of Europe”. *The History of the Family*. 21(1): 38-60.
45. **Teorey**, T.J., **Lightstone**, S.S., **Nadeau**, T., **Jagadish,** H.V. (2011), *Database Modeling and Design: Logical Design*, Fifth Edition, Morgan Kaufmann Publishers-Elsevier, Burlington MA.
46. **Thorvaldsen**, G. (2000). “The Norwegian Historical Data Centre”. In P. H. Hall, R McCaa, & G. Thorvaldsen, *Handbook of International Historical Microdata for Population Research*. Minnesota Population Center.
47. **Thorvaldsen**, G. (2011). “Using NAPP Census Data to Construct the Historical Population Register for Norway”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 11(1): 37-47.
48. **Van Leeuwen**, M.H.D, **Maas**, I., **Miles**, A. (2002). *HISCO. Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.
49. **Varga**, R. (2017). “Romans 1by1 v.1.1. New developments in the study of Roman population”. Digital Classics Online 3(2): 44-59.
50. **Varga**, R., **Pázsint**, A.I., **Boda**, I., **Deac** D.A. (2018). “Romans 1by1. Overview of a research project*”. Digital Classics Online*, 4(2): 37-63.

51. **Willigan**, J. D., **Lynch**, K. A. (1982). *Sources and Methods of Historical Demography: Studies in Social Discontinuity*, Academic Press.

52. **Winchester**, I. (1992). “What every historian needs to know about record linkage for the microcomputer era”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History,* 25(4):149-165.

53. **Wisselgren**, M., **Edvinsson**, S., **Berggren**, M., & **Larsson**, M. (2014). „Testing Methods of Record Linkage on Swedish Censuses”. *Historical Methods: A Journal of Quantitative and Interdisciplinary History,* 47(3): 138-151.