National Genealogical Repository
Developing a Central Database

Antero Ferreira

anterof@gmail.com

Carlota Santos

csantos@ics.uminho.pt

CITCEM-Grupo de História das Populações (GHP)
Universidade do Minho
Guimarães - Portugal

Abstract

This paper presents the National Genealogical Repository project which intends, in a first phase, merge all the existing genealogical databases on the Portuguese population and, in a second phase, aims to broaden its focus nation-wide[1]. We will also discuss some methodological questions related to this process of integration.

1. Introduction

Over the last three decades, the University of Minho has developed a line of research centred on studying historical communities from a micro-analytical perspective. This research is based on databases built from baptism, marriage and death records, organised according to a methodology developed by Maria Norberta Amorim (1991), in which it is possible to follow the life path of each resident in the community in genealogical series.

These databases currently include over 1 million individuals, covering in geographical terms primarily the North of Portugal and two islands of the Azores Archipelago, as well as smaller nuclei in the districts of Porto, Aveiro, Lisbon and Evora. Such a volume of information requires that a centralised system be built, which is able to gather the data on the different communities and allows the researcher to follow the life paths of individuals from broader geographical areas. This need becomes more pressing when we talk of studies on urban areas over several centuries, given the high mobility of individuals and families.

2. Genealogical Databases

In the computer system we use, the SRP – Sistema de Reconstituição de Paróquias (Parish Reconstruction System), the researcher manually enters several events, starting with births, and creates right from this step all family relationships. A typical birth record creates three files on the individuals (child, father and mother) and family, in cases where this family or these individuals, father and mother, do not yet exist in the database.

The researcher enters the information and creates the relationships, in a process that is greatly assisted by the application's capacities to search for individuals.

The system's prime advantage derives from the fact that reconstruction is centred on the individual and not the family. This approach means that all the individuals mentioned in the records, regardless of their having family ties in the parish , are included in the database. Thus, highly useful information is collected in the subsequent exchange process of the several parish databases, given the significant mobility of individuals in the societies of the past.

The genealogical database's structure is essentially rooted in two tables:

- INDIVIDUALS – includes all the relevant information that identifies the individual, collected from baptism, marriage and death records, such as name, date of birth and death, family of origin, titles, etc.. Each individual is given a unique identification number.

- FAMILY – includes information that identifies the family, namely, father and mother.

All the known family relationships are reconstructed based on these two tables. There are also other tables with relevant information, such as, PLACE OF RESIDENCE, PROFESSION, SIGNATURES, etc.. The fact that each individual has a unique ID number means that we can nominatively exchange all types of sources in which that individual is mentioned. Other tables can be added to the database according to the

source and researcher's particular interests, providing the basis for subsequent studies in the area of Demographics and History of the Family, Social History, Biodemographics and other sciences.

The GHP's website hosts about 80 of these genealogical databases, which are openly available to anyone who wishes to consult these communities' family relationships and reproductive dynamics.

3. The National Genealogical Repository and Building a Central Database (CDB)

Since 2001, our research team has been working to build a Central Database covering the national territory which could serve to study mortality and mobility. However, several obstacles have been encountered in this task, some of which have not yet been completely resolved:

3.1 Merging of several databases. If it is true that some vital events are easy to identify, because they are unique, such as birth and death, for example, others, such as marriage or the birth of children, can take place at a variety of locations, making their identification very difficult. This task becomes even more complex because there are no clear rules on the transmission of surnames in Portugal. To overcome this difficulty, we are developing a semiautomatic information exchange procedure: for each new individual included in the CDB, the system searches for similar individuals, comparing names, gender, filiation (legitimate, illegitimate), date and place of birth, name of father and mother. If no similar individuals exist, the new one is automatically recorded. In a next stage, the researcher is presented with a list of similar individuals, and it is up to him/her to decide the procedure to follow.

3.2 Data entry interface. The interface used provides the researcher with high output when entering data. The researcher enters a new individual based on a baptism record, and checks for a family of origin. If there is such a family, the researcher establishes the family kinship between the new individual and the family. If it does not exist, the researcher must record the parents on the individuals table, creates a new family and, finally, establishes the relationship with the baptised individual.

Only one user can use the DB on the system we have implemented, which significantly reduces work efficiency. We are taking the first steps to enable the DB's simultaneous use via an internet platform. We are still debating new search possibilities in the DB, which can improve the identification of individuals during the process of recording information.

3.3 Data retrieval and building genealogies. A number of routines are currently available which allow researchers to conduct traditional analyses in terms of historical demographics. We have also developed data conversion routines from the CDB to the GEDCOM format, which allows us to use GeneWEB , a web-based genealogy software tool.

4. Conclusion

The GHP has long experience in using computer resources to study historical demographics, population history and biodemographics. The aspiration for a more systematic approach to urban spaces and the study of mobility has made the development of a central database an imperative. This goal has led us to face the challenge of computing the exchange of records, although at this stage in the project, we can only present tendencies in our research.

5. References

Amorim, Maria Norberta, (1991), Uma metodologia de reconstituição de paróquias, Braga, Universidade do Minho.

Amorim, Maria Norberta, et al., (2001), "Reconstituição de paróquias e formação de uma base de dados central.", Reconstrução de famílias, fogos e estratégias sociais – Actas do IV Congresso d Associação de Demografia Histórica, vol. II, Castelo Branco, Instituto Politécnico de Castelo Branco, pp. 57-56. Available online at http://hdl.handle.net/1822/4347

Ferreira, Antero (2004), "Sistemas Informáticos para análise de dados demográficos: uma abordagem histórica", SIAD'04 : Sistemas informáticos para a análise de dados demográficos. Guimarães: Universidade do Minho. Available online at http://hdl.handle.net/1822/1925.

Rodrigues, M. de Fatima; Carlos Ramos (1999); Pedro R. Henriques; "Intelligent system to study demographic evolution". Proc. SPIE 3695, Data Mining and Knowledge Discovery: Theory, Tools, and Technology, 161; doi:10.1117/12.339978.

Rodrigues, Maria de Fátima, Pedro Rangel Henriques, (2002), "Integration of heterogeneous knowledge", Submission to 2002 International Conference on Artificial Intelligence (IC-AI'02), Las Vegas, EUA. Available online at http://www.ipb.pt/~pmatos/fct/artigo3.pdf

Websites

http://www.ghp.ics.uminho.pt/

http://www3.di.uminho.pt/~gepl/SEED/