DOI 10.15826/B978-5-7996-2656-3.03

Chapter 2. THE *BAIX LLOBREGAT* (BALL) DEMOGRAPHIC DATABASE, BETWEEN HISTORICAL DEMOGRAPHY AND COMPUTER VISION (NINETEENTH – TWENTIETH CENTURIES)[1]

Joana Maria Pujades-More,
Alícia Fornés,
Josep Lladós,
Gabriel Brea-Martínez,
Miquel Valls-Fígols

Introduction

Historical studies always need empirical evidence, and the enhancement of historical research in the «Big Data Revolution» (Ruggles, 2014), inspired additional need for large quantities of high quality data. Historical demography has always consistently used considerable amounts of data, depending on the technical developments of the moment (Billari and Zagheni, 2017). Many historical demographic databases have a long tradition, such as the Historical Samples of the Netherlands (HSN), the China Multigenerational Panel Dataset (CMGPD-LN); the North Atlantic Population Project (NAPP), the Demographic Database at Umeå University (CEDAR), the Research Program in Historical Demography at Montreal University (PRDH), the Scanian Economic Demographic Database at Lund University (SEDD), among others (Cf Ruggles et al., 2011; Lee and Campbell, 2010; Mandemakers, 2002; Edvinsson, 2000; Dillon

[1] This research has been supported by the project 'NETWORKS: Technology and citizen innovation for building historical social networks to understand the demographic past' (http://dag.cvc.uab.es/xarxes/) funded by Recercaixa program (2017–2019) (PI Joana-Maria Pujadas-Mora and Alicia Fornés). The team includes researchers from the *Universitat Autònoma de Barcelona*, the Center for Demographic Studies (CED), and the Computer Vision Center (CVC).

et al., 2018; Bengtsson et al. 2012).[2] These kind of data is particularly important to understand the demographic past (fertility, family formation, health, social stratification, social inequality, etc.) and can help to understand the present and to forcast the future. However, the digitization of historical sources is still time-consuming with high staff costs. Therefore, the recent conjunction of Historical Demography and the Computer Sciences promises to shorten the construction time for historical individual level databases and allows the building of bigger and more informative databases (Pujadas-Mora et al., 2016; Hall et al. 2000).

The *Baix Llobregat* (*BALL*) *Demographic Database* is an ongoing database project containing individual census data from the Catalan region of *Baix Llobregat* (Spain) during the nineteenth and twentieth centuries. The *BALL* Database is built within the project '*NETWORKS: Technology and citizen innovation for building historical social networks to understand the demographic past*' directed by Alícia Fornés from the Center for Computer Vision and Joana Maria Pujadas-Mora from the Center for Demographic Studies, both at the Universitat Autònoma de Barcelona, funded by the Recercaixa program (2017–2019). Its webpage is http://dag.cvc.uab.es/xarxes/.The aim of the project is to develop technologies facilitating massive digitalization of demographic sources, and more specifically the *padrones* (local censuses), in order to reconstruct historical 'social' networks employing computer vision technology. Such virtual networks can be created thanks to the linkage of nominative records compiled in the local censuses across time and space. Thus, digitized versions of individual and family lifespans are established, and individuals and families can be located spatially.

[2] For more information about each of those databases: The Historical Samples of the Netherlands (HSN) (https://socialhistory.org/en/hsn/index); China Multigenerational Panel Dataset (CMGPD-LN) (https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/27063); North Atlantic Population Project (NAPP) (https://www.nappdata.org/napp/); The Demographic Database at Ümea University (DDBPOPUM)(http://www.cedar.umu.se/english/ddb/databases/popum/); The research Program in Historical Demography at Montreal University (PRDH) (https://www.prdh-igd.com/); The Scanian Economic Demographic Database at Lund University (SEDD) (https://www.ed.lu.se/databases/sedd). This list can be extended with some novel impressive databases as the Historical Population Register (HPR) from the Norwegian Historical Data Centre (http://www.rhd.uit.no/nhdc/hpr.html) or the ones which are created for the International Demographic Unitat the Ural Ferderal University (Russia) (https://urfu.ru/en/research/international-researchcollaboration/international-research-laboratories/international-demographic-unit/). For more information on historical longitudinal databases please see: https://www.ehps-net.eu/databases.

Historical and Geographic context of the Baix Llobregat

Catalonia was one of the first places in Southern Europe to industrialize (Brea-Martínez and Pujadas-Mora, 2018; Martínez-Galarraga and Prat, 2016). The Baix Llobregat region played an important role because since the second half of the nineteenth century (Figure 1), the flourishing Barcelonese cotton industry moved towards the lower parts of the Llobregat River and its delta in search of water for the demanding production (Nadal, 1992). Then, the region changed from an agricultural to a wider occupational and social structure.

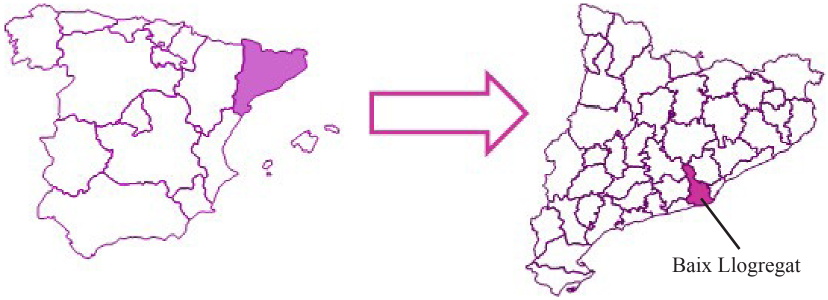


Figure 1. Map of Spain, Catalonia and the region of Baix Llobregat.
Source: Author's own elaboration (BALL Demographic Database)

The industrial growth first featured textile production and afterwards metallurgy (Carbonell i Porro, 1995). However, in the last decades of the nineteenth century the agricultural crisis in Europe affected prices and incomes. In Catalonia the *phylloxera* vineyard infection worsened the crisis (Garrabou et al., 1991). Although the agrarian crisis also disturbed the economy of the *Llobregat* riverside, its effect was limited, because the region disposed of highly fertile soils and had faced an incipient commercialization of its agricultural production since the eighteenth century. During the nineteenth century, the city of Barcelona used to be provided with fruits and other orchard products from the *Llobregat* and since the start of the twentieth century they exported agricultural commodities abroad (Tribó, 1989).

Thus, the *Baix Llobregat* region represents a highly interesting case study. On the one hand, the region played an important role in Catalan industrialization so that the BALL Database includes observations of a society facing all stages in the industrial revolution: the end of the preindustrial era and the take-off, implementation, expansion and consolidation of industrialization. On the other hand, the successful agrarian commercialization in the *Baix Llobregat* since the eighteenth century adds extra elements to the classical model of modern economic and industrial growth, which implies an increasing secondary sector and a declining primary sector (Kuznets and Murphy, 1966; Nadal, 1975; Carreras, 1990; Martínez-Galarraga and Prat, 2016; Brea-Martínez and Pujadas-Mora, 2018).

The BALL Database, therefore, offers analytical opportunities regarding the relation between industrialization and demography. For instance, a classic theory argues that declining fertility was caused by modernization and industrialization and the emergence of nuclear families (Franck and Galor, 2015; Freedman, 1979; Cherlin, 2001). However, our area of study experienced the earlier fertility decline within Catalonia - together with France the earliest fertility declines in Western Europe (Cabré, 1999; Weir, 1993; Coale and Watkins, 1986). Moreover, the *Baix Llobregat* region was traditionally featured by the strong presence of stem families. Nowadays, this area is still shaped by strong family ties (Reher, 2004; Fauve-Chamoux, 2009; Borderías and Ferrer, 2017; Esping-Andersen, 1999). Thus, the BALL Database provides an interesting and complete socioeconomic and socialhistorical "lab" for studying the demographic, familial and individual responses to a transforming world.

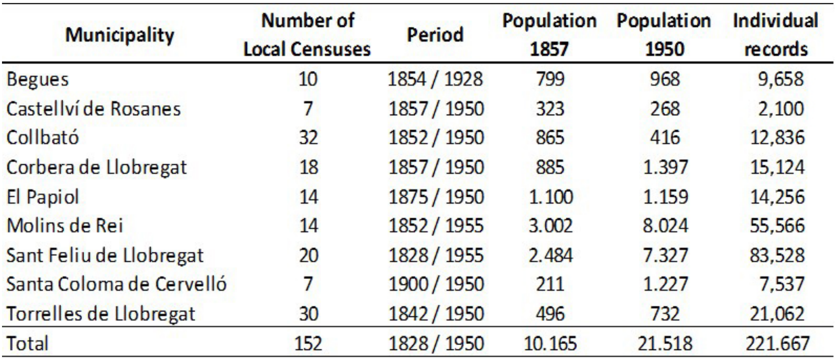
Sources: The *Padrones* (Local censuses)

Local censuses (*Padrones* in Spanish) were taken regularly in Spain since the nineteenth century. They were a result of administrative centralization and the efforts of the liberal state to increase population and wealth (Porter, 1995; Wolf, 1989). Institutionalization of state statistics was a response to a desire for quantifiable material within a framework of epistemological development of scientific objectiveness and impersonal knowledge (Porter, 1995). This was a common although not simultaneous process in nineteenth century Europe. Spanish local censuses were compulsory after 1823 (García Pérez, 2007; Reher and Valero-Lobo, 1995). Thus, they were carried out before the first modern national census (1857) or the definitive implementation of the civil register (1871).

The state progressively implemented local censuses throughout the country. Several decrees informed municipal officials of the obligation to take them. Thus, the Royal Decree of March 14th 1857 demanded all the registers to be nominative and simultaneous. Only after the enactment of the Municipal Law of August 20th 1870 was the taking of local censuses fixed to 5-year intervals. Finally, in the Municipal Statute of March 8th 1924, the formats of local censuses were standardized with respect to the type of variables that should be recorded (García Pérez, 2007; García Ruipérez, 2012). In this way, local censuses showed the sociodemographic features of each inhabitant in a particular household in an urban center or in the countryside. For each person, the register included first names and surnames, age or birth date, civil status and occupation and the family or working relationship with the household head. In some periods, the information is also available about the individuals' literacy and income. Most importantly, the local censuses contain the only data preserved at the individual level in Spain since national census manuscripts used to be destroyed once the population number was estimated and the main variables aggregated.

Table 1

Number and time range of local censuses in BALL Demographic Database, by municipality.



Municipality	Number of Local Censuses	Period	Population 1857	Population 1950	Individual records
Begues	10	1854 / 1928	799	968	9,658
Castellví de Rosanes	7	1857 / 1950	323	268	2,100
Collbató	32	1852 / 1950	865	416	12,836
Corbera de Llobregat	18	1857 / 1950	885	1.397	15,124
El Papiol	14	1875 / 1950	1.100	1.159	14,256
Molins de Rei	14	1852 / 1955	3.002	8.024	55,566
Sant Feliu de Llobregat	20	1828 / 1955	2.484	7.327	83,528
Santa Coloma de Cervelló	7	1900 / 1950	211	1.227	7,537
Torrelles de Llobregat	30	1842 / 1950	496	732	21,062
Total	152	1828 / 1950	10.165	21.518	221.667

Source: Authors' own elaboration (BALL Demographic Database, version July 2018).

The BALL Demographic Database until now contains nine municipalities (Figure 2): *Begues, Castellví de Rosanes, Collbató, Corbera de Llobregat, El Papiol, Molins de Rei, Sant Feliu de Llobregat, Santa Coloma de Cervelló* and *Torrelles de Llobregat*, altogether representing a total number of 152 local censuses comprising the period 1828 to 1950 (Table 1). The medium and small size of the communities registered in the BALL local censuses is far from an inconvenience for the study of networks. It is in reality a strong feature, making it easier to assess the mechanisms that individuals and households used at the micro level in order to adapt to and interact with socioeconomic changes at the macro level (Coleman, 1986).

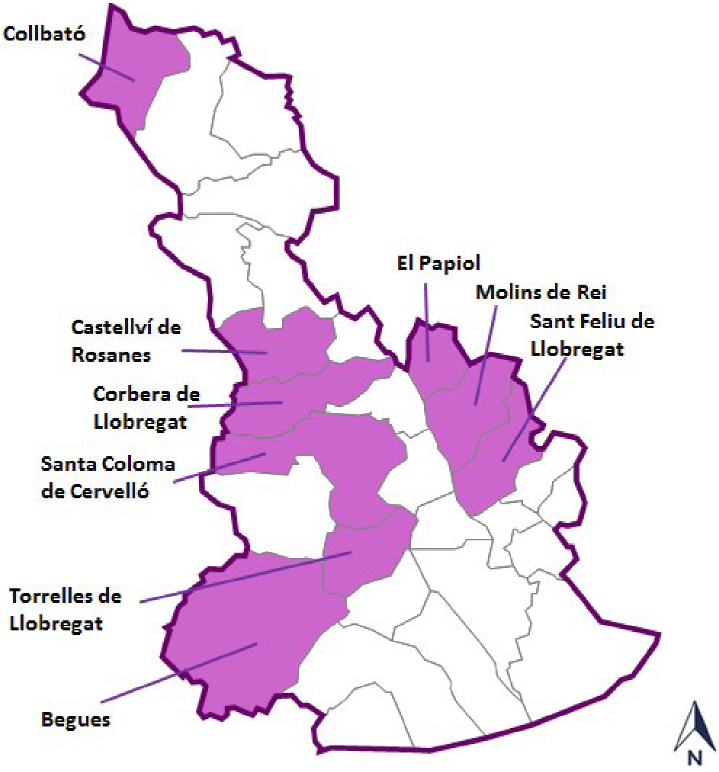
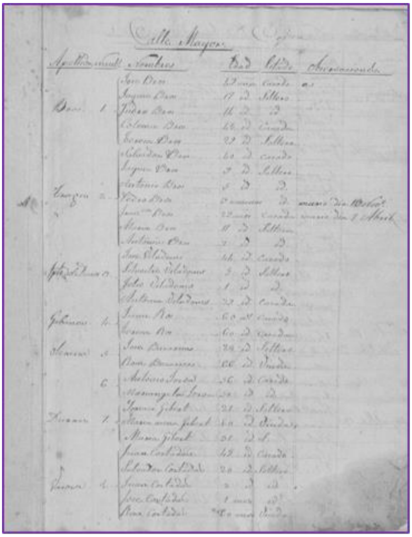
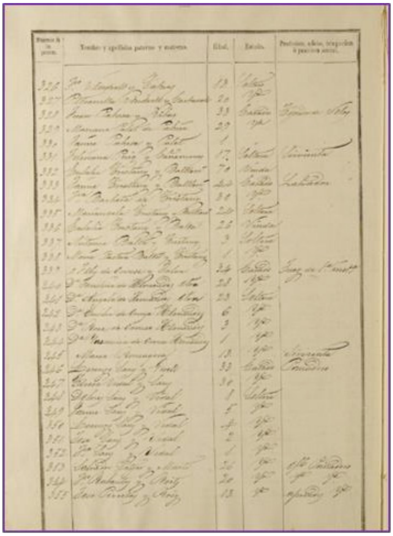


Figure 2. Map of BALL Demographic Database (Baix Llobregat county).
Source: Authors' own elaboration (BALL Demographic Database)

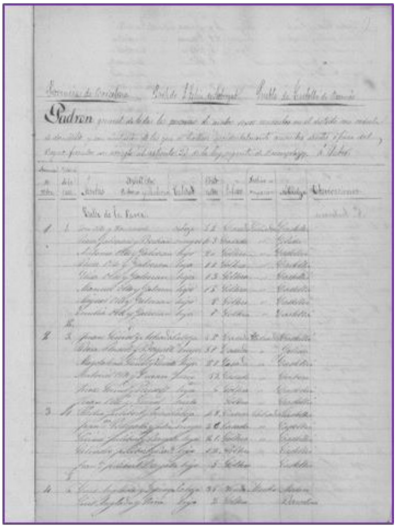
The pre-1880 local censuses were taken before the standardization of census formats. They are basically nominative lists of the inhabitants gathered in the census (Figure 3). For this reason, the number of variables registered usually includes only first names, surnames, ages and civil statuses.



Collbató, 1842



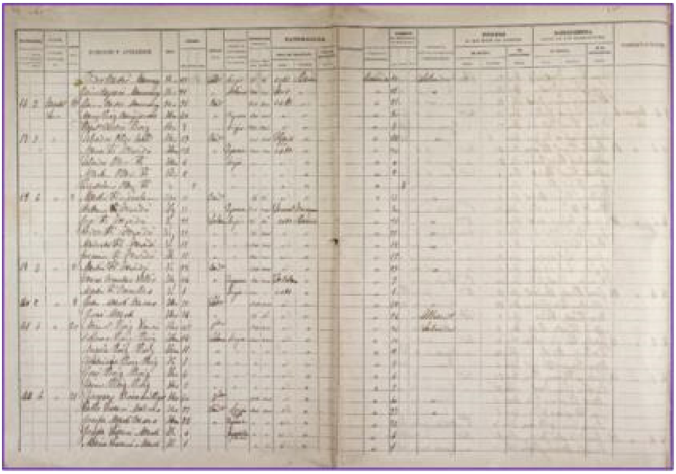
Sant Feliu de Llobregat, 1857



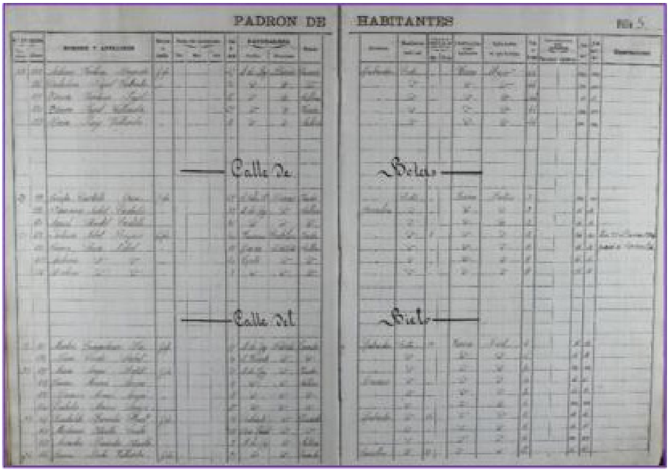
Castellví de Rosanes, 1866

Figure 3. Local Censuses previous to 1880. *Source: Arxiu Comarcal del Baix Llobregat (ACBLL)*

From 1881 the number of variables registered in local censuses increased to around 15 in each census (Figure 4). In this way, the recorded variables were first names, two surnames, occupations, marital status, age and/or birthdate, address (street and house number), birthplace as well as time of residence. From 1906 onwards, other variables are the relationship to the household head, literacy (reading and/or writing), and fiscal contributions (see Annex 1). Nevertheless, in spite of the rich detail in these sources, the local censuses masculinize the data, leaving out many women's occupations. Female work was usually underrecorded or registered with labels such as *su sexo* (tasks of her sex/gender) or *sus labores* (her own tasks) (Borderias, 2012). The only exception to this female invisibility in the local censuses is found in those from 1936 during the Spanish II Republic (1931–1939), which mostly registered women's occupations as well as which factories or enterprises employed the wage earners.



Torrelles de Llobregat, 1887

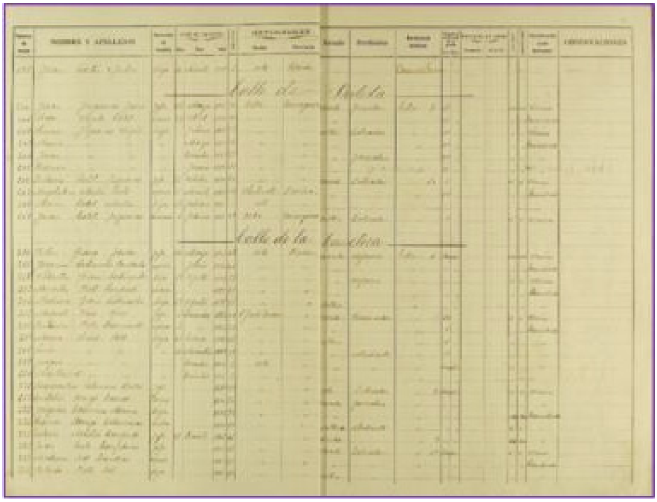


PADRON DE HABITANTES

Calle de

Calle del

Molins de Rei, 1899



Sant Feliu de Llobregat, 1910

Figure 4. Local Censuses from 1880. *Source: Arxiu Comarcal del Baix Llobregat (ACBLL)*

Type of database

As mentioned above, the aim of the project is to create historical 'social' networks using census data. Therefore, we use a social network model to transcribe the nominative census records. This data model also allows using the powerful techniques of graph based data analytics for querying the database. This representation is scalable, allowing not only to integrate and link data from other censuses but also other demographic sources like birth, marriage, death records. Figure 5 illustrates how the census data is structured according to a social network or graph representation, with the time dimension as a stack of linked graphs. Each individual graph is a static representation of the population at time *t* (a particular census) where the nodes are observations of people. The corresponding households where they were registered and the graph edges represent relations (genealogic or other relevant affinities deduced from the source documents like occupations, household neighbours etc). In dynamic, time varying graphs, the life courses of individuals are constructed by linking the corresponding record of their observation at time t_n with the observation at time t_{n+1}.

Technically, we implemented this in a relational database according to the data model in figure 5. Current database functions allow us to implement graph structures in a native format, while providing query methods to analyse the graphs (community detection, record linkage, centrality-based node detection etc). As future work, we plan to migrate to such a format. The proposed design is inspired by the *Intermediate Data Structure* (IDS), a standard proposed by the EHPS Network (Alter & Mandemakers, 2014). This should ensure the interoperability among data sets, and be the basis for the implementation of data analytics.

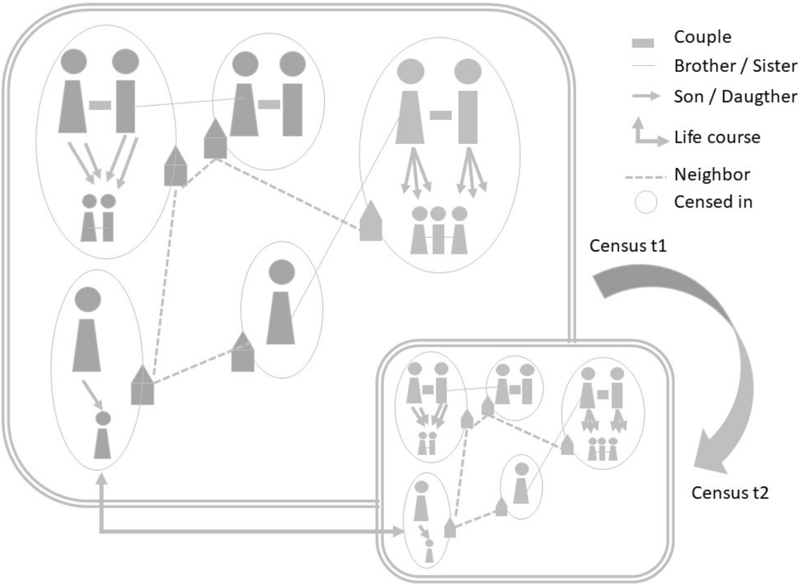


Figure 5. Social network view of census data.
Source: Authors' own elaboration

The design proposed in Figure 6 has the central entity *Person,* containing the information that is permanent over time. Each person has different *Observations*, belonging to the corresponding *censuses* that have been analyzed. Like in the census, observations are grouped in *Households*. A census record is considered an *Event*, which is a class with different subclasses (according to the different event types: census record, marriage record, birth record or death record). This structure allows future users to reconstruct the life course of a person based on the extraction of his/her events.

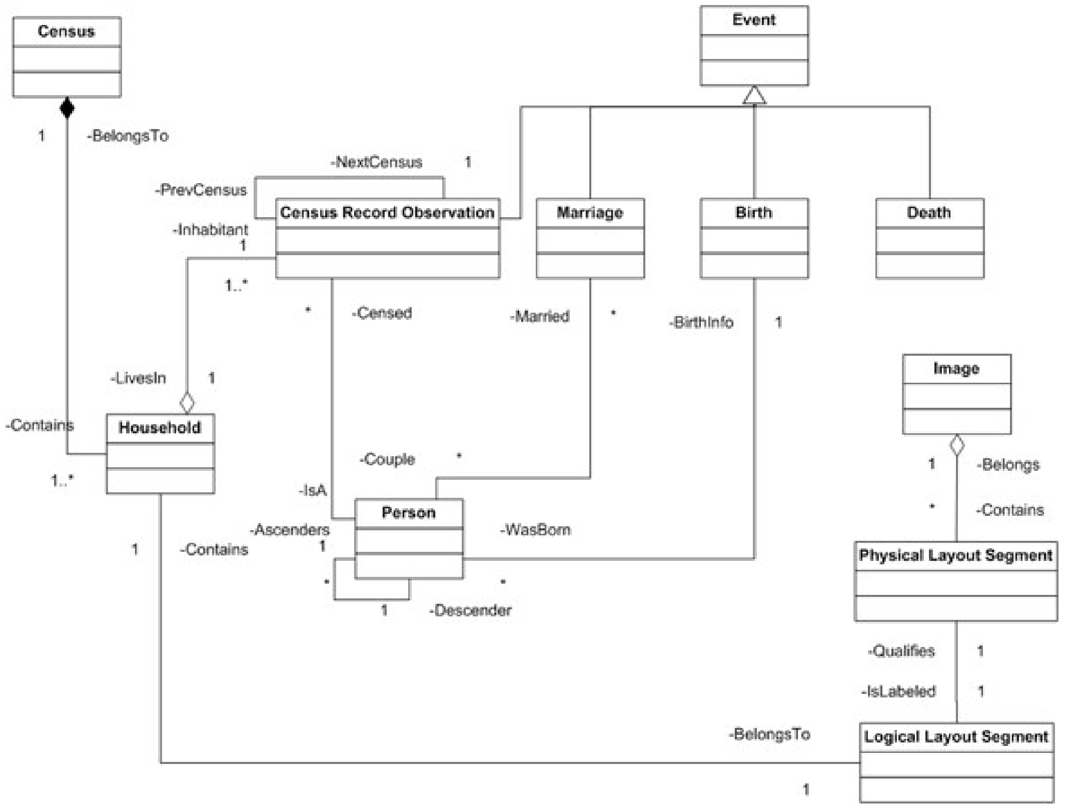


Figure 6. Data Model of the database. *Source: Authors' own elaboration (in unified modeling language)*

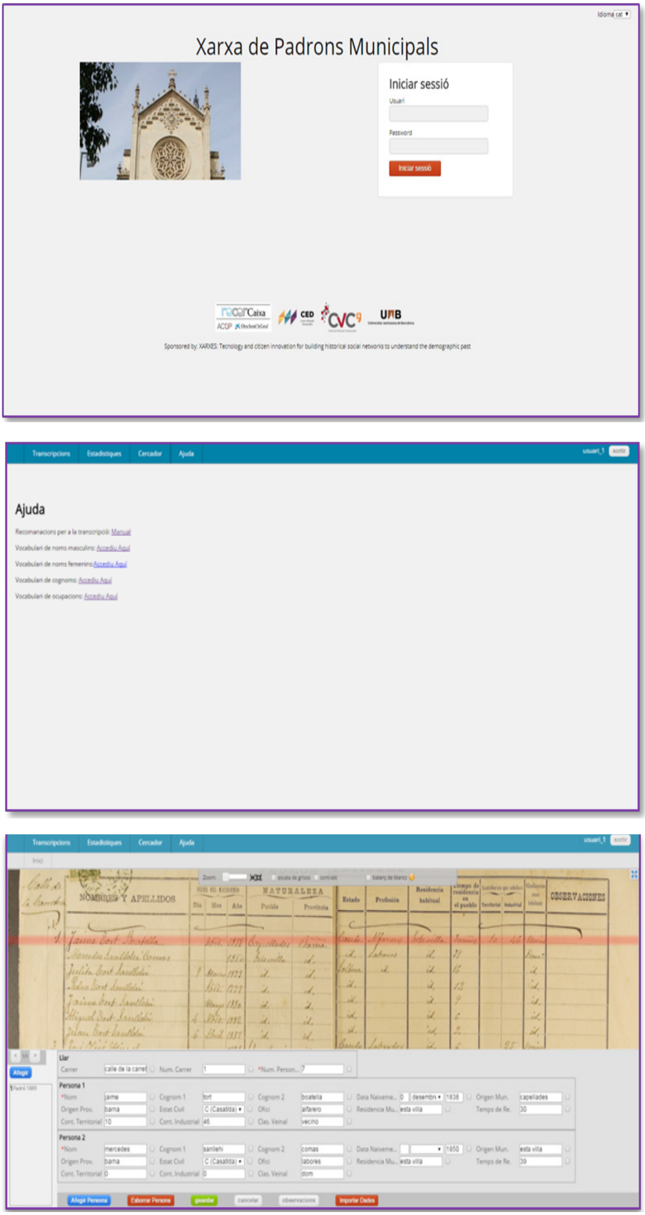
This design is flexible enough to incorporate other event types, not necessarily related to demographic data, but relevant in the life course. Finally, the database architecture contains entities for storing the images from which the information is extracted and the image segments that contain the relevant information as cropped by the computer vision algorithms.

The transcribed data are transferred into a digital format compatible with different formats (Excel, SPSS, STATA, R, etc), downloadable from a crowdsourcing platform used also for collaborative transcription of the original sources.

Transcription:
Crowdsourcing platform, videogame, computer vision & volunteers

The transcription of the manuscripts can be performed either manually or through computer vision techniques. Given that paper degradation and the high variability in handwriting styles in historical manuscripts impose many difficulties, the existing handwriting recognition techniques still need more development before we shall trust automatic transcription completely. For this reason, the local censuses have been transcribed by a community using a web-based crowdsourcing platform, and validated through a gamesourcing mobile application, both assisted with computer vision techniques.

Since manual transcription is tedious and time-consuming, we have developed a crowdsourcing platform for data entry (Figure 7). The idea of crowdsourcing is to split the work into many micro-tasks and ask contributions from a large group of people, especially from the online community. Thus, the task is shared among many users and finished in less time. The graphical user interface of our web-based platform offers a data entry tool with a user friendly environment, integrating both the original source and the data entry form in the same view.



Xarxa de Padrons Municipals

Iniciar sessió

Usuari

Password

Sponsored by: XARXES: Tecnology and citizen innovation for building historical social networks to understand the demographic past

Ajuda

Recomanacions per a la transcripció: Manual

Vocabulari de noms masculins: Accediu Aquí

Vocabulari de noms femenins:Accediu Aquí

Vocabulari de cognoms: Accediu Aquí

Vocabulari de ocupacions: Accediu Aquí

Llar

Carrer: calle de la carret □ Num. Carrer: 1 □ *Num. Person...: 7 □

Persona 1

*Nom: jaime □ Cognom 1: tort □ Cognom 2: boatella □ Data Naixeme...: 0 desembre ▾ 1838 □ Origen Mun.: capellades □
Origen Prov.: barna □ Estat Civil: C (Casatda) ▾ □ Ofici: alfarero □ Residencia Mu...: esta villa □ Temps de Re.: 30 □
Cont. Territorial: 10 □ Cont. Industrial: 46 □ Clas. Veinal: vecino □

Persona 2

*Nom: mercedes □ Cognom 1: sanllehi □ Cognom 2: comas □ Data Naixeme...: ▾ 1850 □ Origen Mun.: esta villa □
Origen Prov.: barna □ Estat Civil: C (Casatda) ▾ □ Ofici: labores □ Residencia Mu...: esta villa □ Temps de Re.: 39 □
Cont. Territorial: 0 □ Cont. Industrial: 0 □ Clas. Veinal: dom □

Figure 7. Crowdsourcing platform: http://dagapp.cvc.uab.es/PadronsXarxes/.
Source: Authors' own elaboration

We also use the transcribed data to teach the computer vision algorithms how to interpret further manuscripts. Concretely, our handwriting recognizer, based on deep learning, uses document images with their corresponding transcriptions to improve the transcription system and to adapt itself to every new handwriting style. As a result, these algorithms can be used to speed up the transcription in two different scenarios: first, to assist the transcription via information transfer, and second to perform an automatic transcription with manual validation.

In the first scenario, we use the redundancy in censuses to automatically transfer repeated information from one census to the next. As was stated before, local censuses in Spain were recorded in intervals of a few years and the nominative information about individuals recorded within households, was quite stable. This redundancy is used to assist the transcriber and speed up the transcription. Once one census has been manually transcribed, the redundant information (names, surnames and addresses) is transferred to the next one, so it is only necessary to update the changes manually: adding new members or deleting those who left or died in each household. For this purpose, household records from consecutive censuses are automatically aligned using the street address. Then, the individuals are automatically located in the image of the census manuscript using visual word search, «word spotting». Concretely, the individual names and surnames from a census are searched in the corresponding home's records in the next census (figure 8). Since the process is based on a focused search, the accuracy is high. This redundancy has been used to assist the transcription of the 1886 census, once the 1881 census was transcribed (Mas et al., 2016), which led to a 70% reduction in the transcription time.

In the second scenario, the transcription is performed automatically, followed by manual validation. First, snippets corresponding to word images are segmented from the manuscripts. Then, these word images are clustered to find high frequency words that can be jointly transcribed using a small percentage of representative instances. Afterwards, the software sends the words in each cluster to the deep learning based transcription system, which provides the most plausible transcriptions for each word image (Figure 9). Finally, clusters that contain words with the same transcription can be transcribed at once. In this way, we avoid the validation of every single word, speeding up the transcription while maintaining high quality performance.

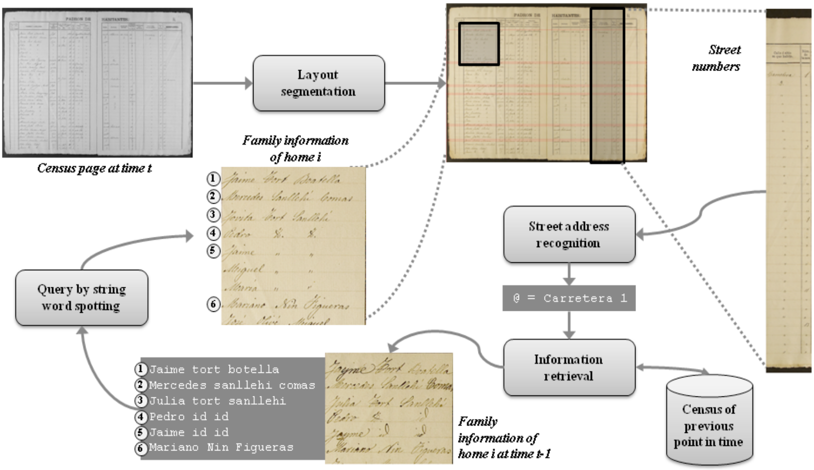

Census page at time t
Layout segmentation
Street numbers
Family information of home i
Street address recognition
@ = Carretera 1
Query by string word spotting
Information retrieval
1 Jaime tort botella
2 Mercedes sanllehi comas
3 Julia tort sanllehi
4 Pedro id id
5 Jaime id id
6 Mariano Nin Figueras
Family information of home i at time t-1
Census of previous point in time

Figure 8. Architecture of the proposed system in first scenario.
Source: Mas et al., 2016

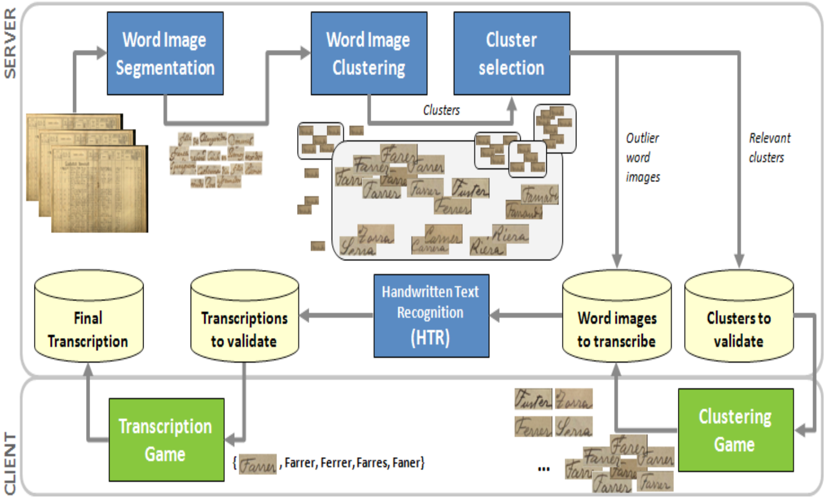


Figure 9. Architecture of the proposed system in second scenario.
Source: Chen et al., 2018

The manual validation has been done through gamesourcing, understood as crowdsourcing via gamification which consists in the application of game-design elements and principles in a non-game context. Concretely, two Android games have been developed (Chen et al., 2018). The first game is designed to validate the word clustering algorithm. It shows some instances of a given cluster, and asks the user to confirm that those words are the same, thus belonging to the cluster. The second game validates the output of the transcription algorithm. When the player selects one word, the system shows the most probable transcriptions according to the algorithm, and the user selects the correct answer among these possibilities (Figure 10). Experiments demonstrate that the transcription effort can be significantly reduced, and that user engagement is higher than with the traditional crowdsourcing web-based application.

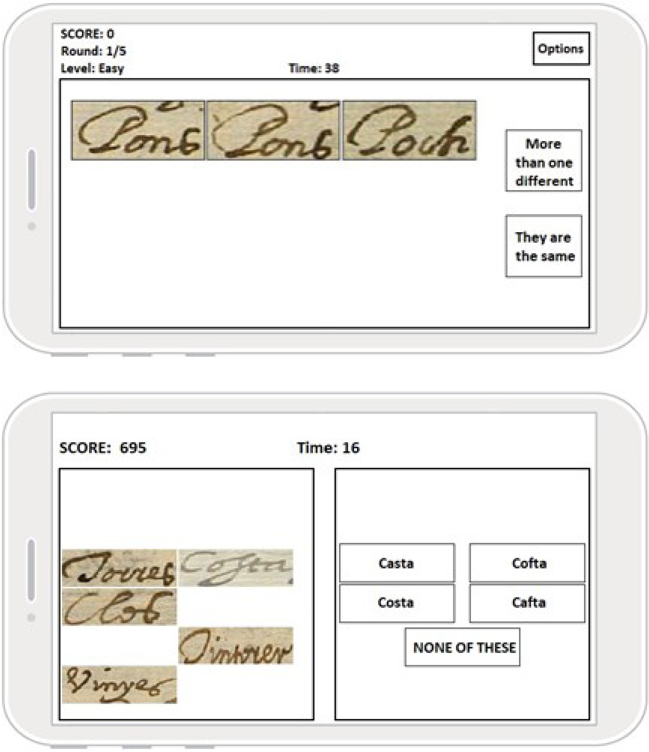


Figure 10: Images of the two games. *Source: Chen et al., 2018*

Thanks to the coordination and collaboration with the Regional Archive of the *Baix Llobregat*, local study groups, town halls and local archives, it was possible to gather groups of transcribers, volunteering in each municipality (figure 11). Collaboration of volunteers in gathering data was already used in the "Cambridge Group for the History of Population and Social Structure" (Wrigley & Schofield, 1989). Until July 2018, 119 transcribers participated (58 females and 61 males), whose mean age is around 60 (Table 2). The transcribers had diverse cultural backgrounds, but enthusiasm about local history and genealogy in common.

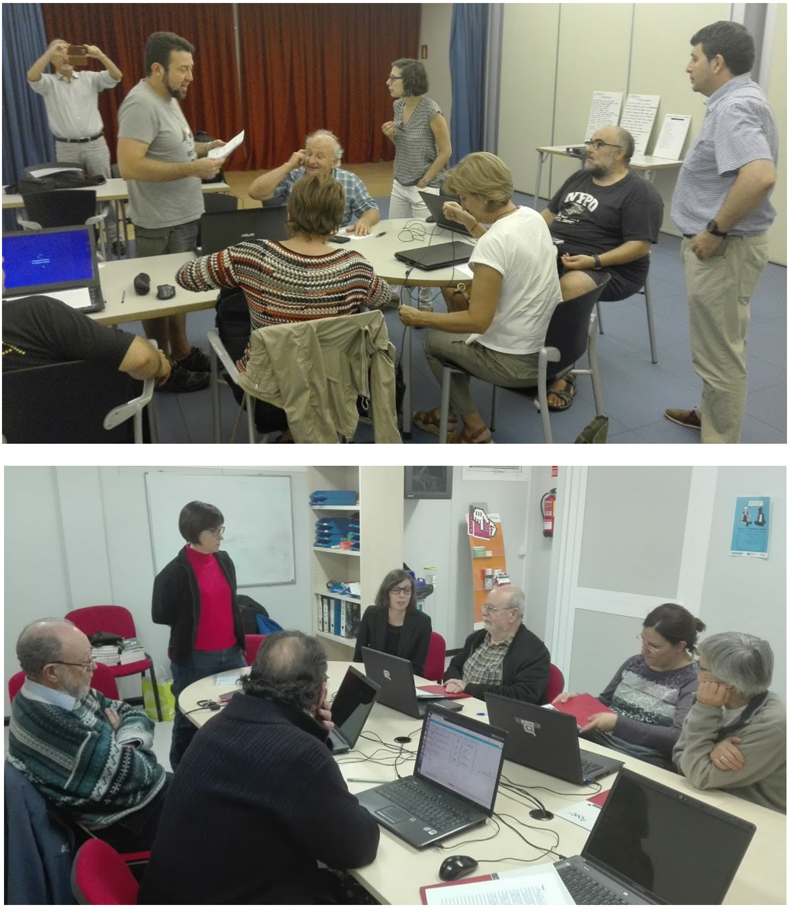
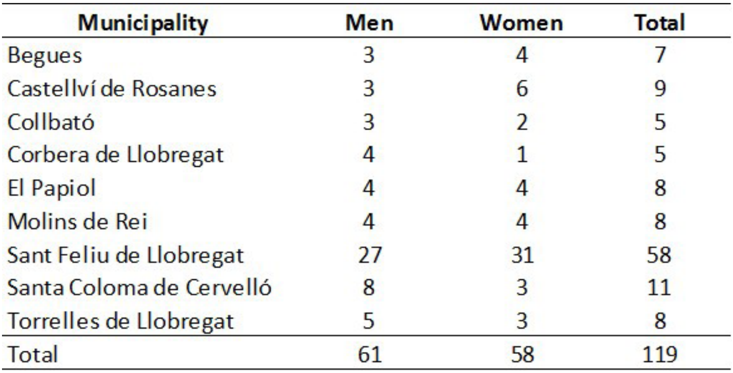


Figure 11. Pictures of some groups of transcribers. *Source: Authors' own elaboration.*

In order to enhance transcription quality and continuously engage the volunteers, we have performed several informative sessions for presenting the NETWORKS project and how the online application of transcription works as well as raising people's awareness about the local censuses and their importance. These informative sessions succeed to engage and train the volunteers, who afterwards received authorization for conducting online transcription.

Table 2

Number of volunteer transcribers by municipality



Municipality	Men	Women	Total
Begues	3	4	7
Castellví de Rosanes	3	6	9
Collbató	3	2	5
Corbera de Llobregat	4	1	5
El Papiol	4	4	8
Molins de Rei	4	4	8
Sant Feliu de Llobregat	27	31	58
Santa Coloma de Cervelló	8	3	11
Torrelles de Llobregat	5	3	8
Total	61	58	119

Source: Authors' own elaboration.

In summary, these technological developments allow citizens and archivists to participate in the extraction of demographic information through web-based crowdsourcing platforms and gamesourcing applications, which incorporate handwriting recognition algorithms to assist the transcribers. This collaboration can be understood as a way of popularizing science for the public and helps to install critical thinking. We shall develop new user experiences in the near future, like geoprojections on interactive maps. Thus, the project facilitates the consumption and dissemination of the historical knowledge in an illustrative and pedagogical way.

Quality checks

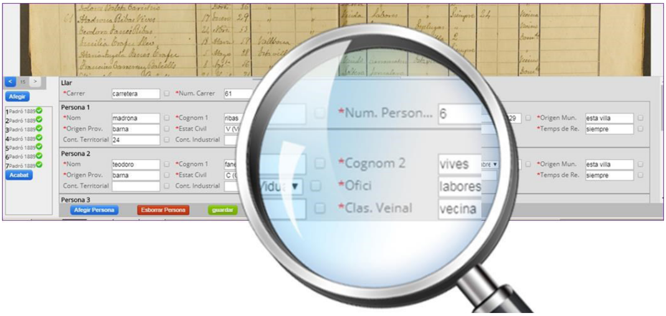
Having online applications for the purpose of transcription enhances supervision and quality control in real time, as well as interplay with the transcribers. In addition, functions for correction are also implemented, as are periodical controls in order to ensure better transcription:

– Pre-transcription:

Recommendations for transcription were delivered to all transcribers in face-to-face training sessions, in order to ensure a common and systematic approach to transcription.

– Control during the transcription:

The data entry tool has compulsory fields for the most important variables, blocking the submission of an incomplete transcription (figure 12). Additionally, in order to avoid leaving out any household member, the total number of household members has to be declared when beginning transcription of every new household. The application includes a "Help" menu with different tools such as dictionaries, lists of names as well as geographic and occupational guides to help the transcribers when in doubt. Moreover, all the transcribers can continuously address their queries via email or social networks to enhance the communication between researchers, technicians and transcribers. Every transcriber had a reviewer in charge of verifying transcription quality and solving doubtful cases. Finally, we released statistics about the transcription and database progress to the community of transcribers in order to dynamise the transcription process.



Llar
*Carrer carretera □ *Num. Carrer 61
*Num. Person... 6

Persona 1
*Nom madrona □ *Cognom 1 ribas
*Origen Prov. barna □ *Estat Civil V (V
Cont. Territorial 24 □ Cont. Industrial
*Cognom 2 vives
*Ofici labores
*Clas. Veinal vecina
29 □ *Origen Mun. esta villa □
*Temps de Re. siempre □

Persona 2
*Nom teodoro □ *Cognom 1 fane
*Origen Prov. barna □ *Estat Civil C (
Cont. Territorial □ Cont. Industrial
*Origen Mun. esta villa □
*Temps de Re. siempre □

Persona 3

Figure 12. Control during the transcription. *Source: Authors' own elaboration*

– Post-transcription control:

Once the transcription of a given local census was finished, several analyses are conducted regarding the frequencies of different variable values for identifying likely inconsistencies.

Harmonization and codification of data

One central characteristic of the BALL Demographic Database is that instead of transcribing data in a harmonized way, the volunteers have been told to transcribe all records literally in order to avoid inconsistent interpretations of the nominative data. In this way, once the sources have been transcribed it is necessary to apply linguistic harmonization. The same individual appears with different names and surnames, or apparent variants because they were recorded originally with different spelling, or abbreviations. One combination of a common first name and surname might refer to several individuals (Goiser and Christen 2006; Herzog et al., 2007, Schürer, 2007). These questions need particular attention in the case of the Catalan language, which was not standardized until 1913. Also in occupation titles, locations, relationships with the household head or even marital status there is vast variability. Apart from reducing such variability, harmonization is important to make variables and outcomes more accessible for analysis and comparisons at the national or international level. For this reason, we recode the variables in each local census in order to remove phonetic variations caused by the different dialect of Catalan and influences from other languages which favored many written variations (Peytaví 2010; Rubió and Lizondo, 1997). For instance, we found the surname *Ferrer* (Smith) written as *Ferré, Farré, Farrer*, and so on.

Linguistic harmonization helps overcome these obstacles and clusters anthroponyms in order to compile dictionaries of names and surnames (Bloothooft, 1998; Christen, 2012). Thus, nominative data have been harmonized according to language criteria to facilitate the record linkage that identifies the same individual in different censuses (Jordà, 2016; Jordà et al., 2013). In addition, places have been geolocated, and occupations have been coded using the Historical International Classification of Occupations (HISCO) (van Leeuwen et al., 2002).

Once all occupations were codified, we ranked and classified them according to sociooccupational position by means of HISCLASS (van Leeuwen and Maas, 2011) and HISCAM (Lambert et al., 2013), respectively. HISCLASS differentiates individuals in consonance with the social group to which they belonged according to dimensions like manual/non-manual division, skill level, degree of supervision and economic sector, which gives 12 different classes going from unskilled rural workers at the bottom to higher managers and professionals at the top. HISCAM is a different occupational stratification scale based on the Cambridge Social Interaction and Stratification scheme, using marriage data from Belgium, Britain, Canada, France, Germany, the Netherlands, and Sweden (Prandy, 2000). The main idea behind this scaling is that individuals who interact more (in terms of occupational and social relations) are closer in terms of social position, assuming that these interactions represent the occupational stratification structure. The result is a ranking of occupations (theoretically from 0 to 99), showing not only similar social standing but also differences between occupations. Finally, we have grouped each occupation into economic sectors, following an adaptation of HISCO for the Catalan historical labor market (Pujadas-Mora et al., 2014).

As it was mentioned before, the BALL Database, in July 2018, is composed of 9 municipalities (figure 2) and 152 local censuses that correspond to a total of 221,667 individual observations (see table 1). The harmonization of its names, occupations and places uses the reference tables created during the harmonization of the Barcelona Historical Marriage Database (Jordà, 2016). In this way, we rely on a reference table for the names (female and male names) with 27,434 unique records, the surnames with 101,238 unique records, the occupations with 24,399 unique records and the place names with 47,556 unique records, which facilitate the work of linguistic normalization and harmonization.

Towards longitudinal data: The *Sant Feliu* life course database

Among potential derivations from the BALL Demographic Database is the possibility to create longitudinal databases of individuals and households by linking local censuses. For this, we have used one of the most populated towns in *Baix Llobregat*, one with a large time coverage, which is also the capital of the county, namely *Sant Feliu de Llobregat,* to create a sub-database. Thus, the longitudinal database of *Sant Feliu* is based on reconstruction of individual life-courses using local censuses. The town of *Sant Feliu de Llobregat* was one of the most important in the region, in economic and administrative terms, being the judicial district capital, with the arrival of new economic activities such as textile and metallurgical industries since the second half of the nineteenth century and the railway station in 1855. The database contains the information in all the 15 censuses recorded in *Sant Feliu* from 1828 to 1940. This information has benefited from computer-assisted data transcription through crowdsourcing in which 58 volunteers have collaborated (27 men and 31 women) for a period of 2 years. As explained above, the local censuses in Spain (including *Sant Feliu*) were taken at short intervals.

The Longitudinal Database of *Sant Feliu* contains 59,084 observations of individuals, which increased chronologically, mainly from the 1920 onwards when *Sant Feliu de Llobregat* started having important inmigration. The individuals are distributed in 12,748 households across the entire period with a mean number of 4.6 individuals per household, levels that decreased over time from 5.2 persons per household in 1828 to 4 in 1940 (table 3). The nominative data has been harmonized, addresses have been geolocated, and occupations have been encoded.

Table 3

Individuals and households by census year in Sant Feliu de Llobregat 1828–1940

Years	Individuals Households		Individuals per Household
1828	2,209	426	5.2
1833	1,470	313	4.7
1839	1,946	377	5.2
1857	2,472	533	4.6
1878	2,762	610	4.5
1881	3,005	598	5.0
1889	3,118	644	4.8
1906	3,606	805	4.5
1910	3,809	866	4.4
1915	4,330	936	4.6
1920	4,353	918	4.7
1924	5,575	1081	5.2
1930	6,392	1459	4.4
1936	7,023	1458	4.8
1940	6,727	1675	4.0
Total	59,084	12,748	4.6

Source: Authors' own elaboration.

Harmonization and codification significantly reduced the number of spelling variations, erroneous or misleading data. From the initial 1,976 given names transcribed literally, the harmonization dropped the number to 1,033 variants. Since Spanish individuals usually receive two surnames (paternal and maternal), the surname harmonization was multiplied by two. The 4,269 different paternal and 5,641 maternal surnames, became 2,767 and 3,312 after harmonization, respectively. Finally, at the geographic level, the 3,743 different addresses observed were harmonized to 1,552 different geographic points. Nevertheless, as the database includes local censuses spread over a wide period, not all contain the same information. For instance, local censuses in *Sant Feliu* prior to 1906 did not have information on the specificity of relationship with the household head, i.e. who were spouses, children and/or other relatives. We have used other variables such as surnames, marital status and age in order to reconstruct family relationships within the same household. Finally, the procedure of harmonizing and coding the occupations resulted in 1,611 different occupational titles that we simplified into 292 HISCO codes.

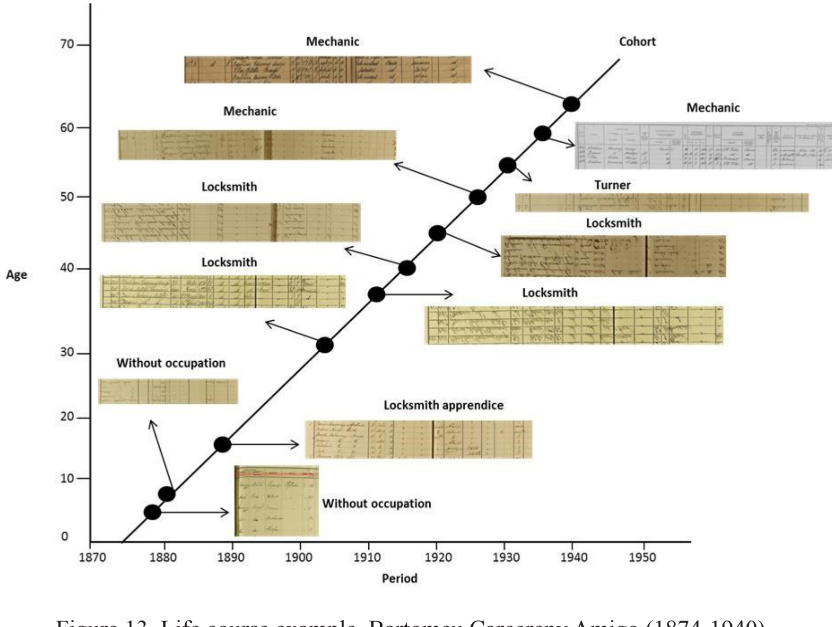


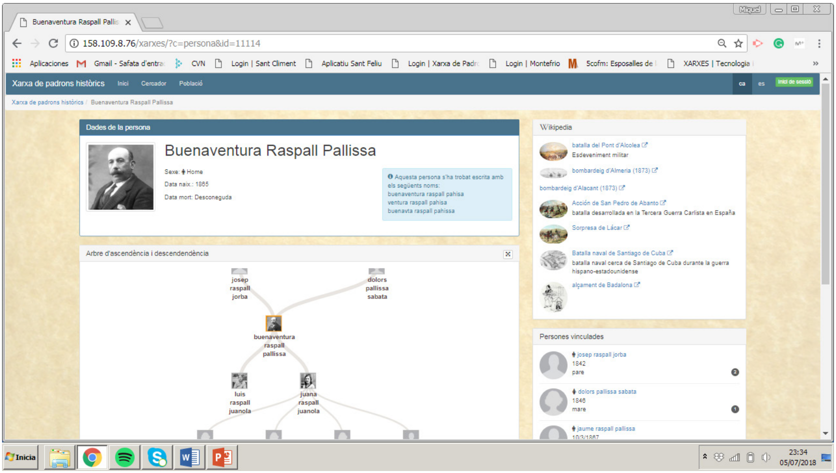
Figure 13. Life course example. Bartomeu Carcereny Amigo (1874-1940).
Source: Authors' own elaboration.

After standardization, nominative record linkage was performed. The procedure uses deterministic (two string distances, the Bag distance and the Levenshtein distance) and probabilistic criteria, together with a language model that avoid similarities in pronunciation to be considered as the same surname (e. g. Piera & Riera, Pons & Pous, etc.) to and penalized typographic errors (Villavicencio, Jordà, and Pujadas-Mora, 2015). This allowed us to follow 10,405 individuals at least in two different censuses. In the best cases, we reconstructed up to 12 different observations for individuals, which signifies assessing the entire life-cycle of a person. For instance, consider the life course of Bartomeu Carcereny Amigó (figure 13). This individual was a native of *Sant Feliu* and born on 1874. From then on, the longitudinal database allowed us to follow him in 11 different enumerations across his life until the last observation found in 1940, when he was 66 years old.

As can be seen from this example, one possibility of analysis with longitudinal data, is the work career of individuals. The first two observations of Bartomeu were without occupation. Around 14 years old, we found Batomeu's first occupational observation as locksmith apprentice, which he continued until the first decade of the twentieth century. However, from the 1920s until the 1940s the four last observations of Bartomeu informs us of two different intermittent occupations, mechanic and turner. Thus, as can be observed through a single case, the longitudinal database of Sant Feliu informs us not only about changes at the individual micro level, but also how this individual reacted to macro level transformations such as industrialization, given he turned from an artisan position to an occupation that emerged with a new mode of production.

Popularizing science: Onomastics browser & demographic visualizations

The NETWORKS project has an important commitment to citizen science that goes beyond their participation in census transcription. The project also aimed at reaching the public by contributing to society through the creation of a browser constructed to display onomastics and interactive demography (figure 14). In May 2017 the onomastics application of the Sant Feliu database was launched: http://158.109.8.76/xarxes/. This also shows the genealogical tree (with a varying number of generations) for each individual in the town.



158.109.8.76/xarxes/?c=persona&id=11114

Xarxa de padrons històrics · Inici · Cercador · Població

Xarxa de padrons històrics / Buenaventura Raspall Pallissa

Dades de la persona

Buenaventura Raspall Pallissa

Sexe: Home

Data naix.: 1865

Data mort: Desconeguda

Aquesta persona s'ha trobat escrita amb els següents noms:
buenaventura raspall pahisa
ventura raspall pahisa
buenavta raspall pahissa

Arbre d'ascendència i descendència

Wikipedia

batalla del Pont d'Alcolea
Esdeveniment militar

bombardeig d'Almeria (1873)

bombardeig d'Alacant (1873)

Acción de San Pedro de Abanto
batalla desarrollada en la Tercera Guerra Carlista en España

Sorpresa de Lácar

Batalla naval de Santiago de Cuba
batalla naval cerca de Santiago de Cuba durante la guerra hispano-estadounidense

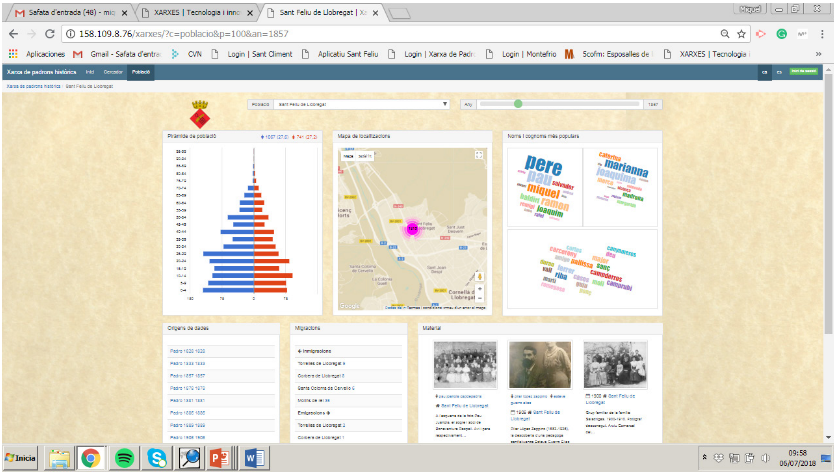
alçament de Badalona

Persones vinculades

josep raspall jorba
1842
pare

dolors pallissa sabata
1846
mare

jaume raspall pallissa
10/3/1867



158.109.8.76/xarxes/?c=poblacio&p=100&an=1857

Xarxa de padrons històrics · Inici · Cercador · Població

Xarxa de padrons històrics / Sant Feliu de Llobregat

Població: Sant Feliu de Llobregat

Any: 1857

Piràmide de població

1067 (27,6) 741 (27,2)

Mapa de localitzacions

Noms i cognoms més populars

Orígens de dades

Padro 1828 1828

Padro 1833 1833

Padro 1857 1857

Padro 1878 1878

Padro 1881 1881

Padro 1886 1886

Padro 1889 1889

Padro 1906 1906

Migracions

Inmigracions

Torrelles de Llobregat 9

Corbera de Llobregat 8

Santa Coloma de Cervello 6

Molins de rei 35

Emigracions

Torrelles de Llobregat 2

Corbera de Llobregat 1

Material

pau joanola capdegadris
Sant Feliu de Llobregat
A l'esquerra de la foto Pau Juanola, el sogre i soci de Bonaventura Raspall. Avi i pare respectivament...

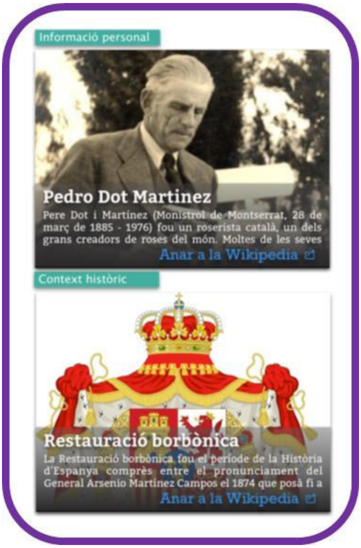
pilar lopez zappino · esteve guarro elias
1906 Sant Feliu de Llobregat
Pilar López Zappino (1882-1936), la descoberta d'una pedagoga santfeliuenca Esteve Guarro Elias

1900 Sant Feliu de Llobregat
Grup familiar de la família Saladrigas. 1900-1910. Fotògraf desconegut. Arxiu Comarcal del...

Figure 14. Onomastic browser. *Source: Authors' own elaboration.*

Moreover, it includes interactive visualizations of population pyramids and the town's most common names and surnames during the nineteenth and twentieth centuries. Additionally, we linked the web site with Wikipedia to provide the historical context of the database and biographical information about individuals recorded in the censuses. Wikipedia articles are also presented through an app (figure 15).

Another mobile app is being developed to enrich the census information with pictures, kept in historical archives or family archives.



Pedro Dot Martinez

Pere Dot i Martínez (Monistrol de Montserrat, 28 de març de 1885 - 1976) fou un roserista català, un dels grans creadors de roses del món. Moltes de les seves

Restauració borbònica

La Restauració borbònica fou el període de la Història d'Espanya comprès entre el pronunciament del General Arsenio Martínez Campos el 1874 que posà fi a

Figure 15. Wikipedia connection.
Source: Díaz, 2018

Concluding remarks

Citizen science, crowdsourcing, historical demography, computer vision and gamification are the cornerstones of the *Ball demographic database*. Formally, the Ball database is a relational database containing census data at the individual level from the Catalan county of *Baix Llobregat* in Spain for the nineteenth and twentieth centuries with more than 221,667 individual observations built by diferent interdisciplinary projects in Digital Humanities. These projects are a joint venture of the Center for Demographic Studies and the Computer Vision Center, both at the Universitat Autònoma de Barcelona. Hence, this database is a proof that activating volunteer citizens through an *ad hoc* online crowdsourcing platform as a data entry tool and games for touch-screen devices, and Handwritten Text Recognition (HTR) techniques into data collection of primary sources can meaningfully reduce the time for building individual-level historical demographic databases. This integration has been possible thanks to recent advances in Handwriting Recognition, the expansion of information technologies, the popularization of handheld devices, the assimilation of internet in everyday life and the massive digitalization of historical documents. However, the current state-of-the-art of Handwriting Recognition still requires some human intervention.

Thus, two scenarios have been designed in BALL database to speed up trancription using computer vision algorithms, based on deep learning techniques. First, assisted transcription by information transfer from one document to another and second, automatic transcription with manual validation. These scenarios have proved to reduce progressively the human effort. In the first scenario, the system benefits from the redundant information (names, surnames and addresses) shared by consecutive censuses only needing the manual transcription (through a crowdsourcing platform) of one census, whose information is transferred to the next census. Using visual word search, so-called word spotting, each transcribed household in the first census is linked with the image of the same household a few years later, assisting the manual updating of information from one census to the next one. Moreover, the use of gamesourcing experiences for the validation of automatic transcription carried out by handwriting recognition algorithms is a twist in the integration of computer vision methods and keeping the human intervention. At the same time, the engagement of citizens through gamification appears to be higher, as can be expected, which has a beneficial effect on the total number of transcribed words. Overall, we have set the stage for effective semi-automatical processing of large document collections in order to create databases in a faster and more effective way as part of the Big Data revolution.

Additionally, the BALL database is more than a database for demographic and computer vision research. The possibility of browsing this database through names and surnames using an open access and user friendly webpage boost greatly their use, mainly for the public at large, who is usually not familiar with this kind of sources. Queries can be also done through genealogical trees, that have been generated using record linkage techniques, facilitating extremely the search task for many users, who are looking for their ancestors. At the same time, visualizations of the population's evolution and onomastics through graphics are offered as tools to popularize demography among the general audience. In this way, citizens who are a fundamental part in the building of the database can also take advantage of the final result through the above-mentioned webpage, which is an import asset in terms of knowledge transfer to society. Besides, in spite of pursuing scientific objectives, these citizen-centered projects have an important potential social impact in terms of literacy or shortening the technological gap in terms of social inclusion and cohesion.

Variables	Local Censuses																					Total
	1828	1833	1857	1857	1878	1881	1886	1889	1890	1896	1900	1906	1910	1915	1920	1924	1930	1940	1945	1950	1955	
Name	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	21
Surname 1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	21
Surname 2			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	19
Occupation			X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	18
Sex																X	X	X	X	X	X	6
Number of Men in the household	X																					1
Number of Women in the household	X																					1
Number of household																X	X	X	X	X	X	6
Relation to the head of household							X			X	X	X	X	X	X	X	X	X	X	X	X	13
Civil Status		X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	19
Age		X	X	X	X	X	X		X	X	X	X	X	X	X	X						14
Street Name		X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	18
Number of the household		X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	18
Citizenschip																X	X	X	X		X	5
Municipality of Birth						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16
Province of Birth						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16
Person Number			X	X		X			X	X	X	X	X	X	X	X	X	X	X	X	X	16
Date of Birth						X	X	X	X	X	X	X	X		X			X	X	X	X	13
Month of Birth						X	X	X	X	X	X	X	X		X			X	X	X	X	13
Year of Birth						X	X	X	X	X	X	X	X		X		X	X	X	X	X	14
Municipality of usual residence						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16
Province of usual residence																X	X	X	X	X	X	6
Years living in the municipality						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16
Months living in the municipality						X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16
Territorial contribution								X	X	X	X	X	X	X								7
Industrial contribution								X	X	X	X	X	X	X								7
Classification of inhabitant								X	X	X	X	X	X	X		X	X	X	X	X	X	13
Literacy							X		X	X	X	X	X	X	X	X	X	X	X	X	X	14
Residential Status																	X	X	X	X	X	5
Wages and Salary																		X			X	2
Total	4	6	7	7	6	17	18	18	21	22	22	22	22	19	19	21	22	25	24	23	25	370

Annex 1. Variables of Local Censuses of Sant Feliu de Llobregat, 1828–1955.
Source: Authors' own elaboration.

References

Alter, George and Mandemakers, Kees. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical life course studies*. 1

Bengtsson, Tommy, Martin Dribe, and Patrick Svensson. (2012). *The Scanian Economic Demographic Database: Version 3.0* (Machine-readable database). Lund, Lund University, Centre for Economic Demography.

Billari, Francesco C. and Emilio Zagheni. (2017). Big Data and Population Processes: A Revolution? In *SIS 2017. Statistics and Data Science: new challenges, new generations. 28–30 June 2017 Florence (Italy) Proceedings of the Conference of the Italian Statistical Society*, edited by Alessandra Petrucci, and Rosanna Verde. Firenze University Press. 67–178.

Bloothooft, Gerrit. (1998). Assessment of systems for nominal retrieval and historical record linkage. *Computers and the Humanities*. 32 (1): 39–56. https://www.jstor.org/stable/30200450

Borderías, Cristina. (2012). La reconstrucción de la actividad femenina en Cataluña circa 1920. *Historia Contemporánea*. 44:17–47. http://www.ehu.eus/ojs/index.php/HC/article/view/6600/6038

Borderías, Cristina, and Llorenç Ferrer-Alòs. (2017). The stem family and industrialization in Catalonia (1900–1936). *The History of the Family*. 22(1): 34–56. Doi.10.1080/1081602X.2016.1242083

Brea-Martínez, Gabriel, and Joana Maria Pujadas-Mora. (2018). Transformación y desigualdad económica en la industrialización en el área de Barcelona, 1715–1860. *Revista de Historia Económica-Journal of Iberian and Latin American Economic History*. 36(2): 241–273. Doi:10.1017/S0212610917000234

Cabre, Anna. (1999). *El sistema català de reproducció. Cent anys de singularitat demogràfica*. Barcelona: Editorial Proa.

Carbonell i Porro, Joan Anton. (1995). Desenvolupament de la indústria tèxtil a Sant Feliu de Llobregat (1861 – 1923). In *El pas de la societat agrària a industrial al Baix Llobregat*, edited by Angel Calvo i Calvo. Barcelona, Centre d'Estudis Comarcals del Baix Llobregat, Publicacions de l'Abadia de Montserrat, 405–425.

Carreras, Albert. (1990). *Industrialización española: estudios de historia cuantitativa*. Madrid, Espasa Calpe.

Chen, Jialuo, Pau Riba, Alicia Fornés, Joan Mas, Josep Lladós, and Joana Maria Pujadas-Mora. (2018). Word-Hunter: A Gamesourcing Experience to Validate the Transcription of Historical Manuscripts. In *International Conference on Frontiers in Handwriting Recognition (ICFHR).*

Cherlin, Andrew J. (2001). *Public and private families: A reader*. New York, McGraw-Hill

Christen, Peter. (2012). *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection. Data-centric systems and applications*. Berlin, Springer Science & Business Media.

Coale, Ansley J., and Susan Cotts Watkins. (1986). *The decline of fertility in Europe: the revised proceedings of a Conference on the Princeton European Fertility Project*. Princeton, Princeton University Press.

Coleman, James S. (1986). Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*. 91(6): 1309–335. http://www.jstor.org/stable/2779798.

Díaz, Rafael. (2018). *Connection of a historical social network with the Wikimedia community*. End of degree in Computer Engineering directed by Alicia Fornés and Joana Maria Pujadas-Mora. Barcelona, Autonomous University of Barcelona.

Dillon, Lisa, Marilyn Amorevieta-Gentil, Marianne Caron, Cynthia Lewis, Angélique Guay-Giroux, Bertrand Desjardins, and Alain Gagnon. (2018). The Programme de recherche en démographie historique: past, present and future developments in family reconstitution. *The History of the Family*. 23(1): 20–53. Doi: 10.1080/1081602X.2016.1222501

Edvinsson, Sören. (2000). The Demographic Data Base at Umeå University–a resource for historical studies. In *Handbook of international historical microdata for population research*, edited by Hall, Patricia Kelly, Robert McCaa, and Gunnar Thorvaldsen.Minneapolis, Minnesota Population Center, 231–248.

Esping-Andersen, Gosta. (1999). *Social foundations of postindustrial economies*. OUP Oxford.

Fauve-Chamoux, Antoinette. (2009). *The stem family in Eurasian perspective: Revisiting house societies, 17th–20th centuries*. Vol. 11. Peter Lang.

Franck, Raphael and Oded Galor. (2015). Industrialisation and Fertility Decline. *Working Paper, No. 2015–6*. Brown University, Department of Economics.

Freedman, Ronald. (1979). Theories of fertility decline: A reappraisal. *Social forces*. 58(1): 1–17. DOI:10.2307/2577781

García Pérez, María Sandra. (2007). El padrón municipal de habitantes: origen, evolución y significado. *Hispania Nova: Revista de historia contemporánea*. 7:1–8. http://hispanianova.rediris.es/7/articulos/7a005.pdf

García Ruipérez, Mariano. (2012). El empadronamiento municipal en España: evolución legislativa y tipología documental. *Documenta & Instrumenta-Documenta et Instrumenta*. 10:45–86. Doi: 10.5209/rev_DOCU.2012.v10.40485

Garrabou, Ramon, Josep Pujol Andreu, and Josep Colomé i Ferrer. (1991). Salaris, ús i explotació de la força de treball agrícola (Catalunya 1818–1936). *Recerques: història, economia, cultura*. 24:23–51. https://www.raco.cat/index.php/Recerques/article/view/137681

Goiser, Karl, and Peter Christen. (2006). Towards automated record linkage. In *Proceedings of the fifth Australasian conference on Data mining and analystics-Volume 61*:23–31. Australian Computer Society, Inc.

Hall, Patricia Kelly, Robert McCaa, and Gunnar Thorvaldsen, (ed.). (2000). *Handbook of international historical microdata for population research*. Minneapolis, Minnesota Population Center.

Herzog, Tamar. (2007). Nombres y apellidos: ¿cómo se llamaban las personas en Castilla e Hispanoamérica durante la época moderna? *Jahrbuch für Geschichte Lateinamerikas*. 44(1):1–35. Doi:10.7767/jbla.2007.44.1.1

Herzog, Thomas N., Fritz J. Scheuren, and William E. Winkler. (2007). *Data quality and record linkage techniques*. New York. Springer Science & Business Media.

Jordà, Joan Pau. (2016). *Aproximación a las migraciones históricas a través del estudio de la información nomina*l. Doctoral Thesis. Barcelona. Universitat Autònoma de Barcelona.

Jordà, Joan Pau, Miquel Valls, and Joana Maria Pujadas-Mora. (2013). Apellidos y migraciones. Estudio a través de los fogatges catalanes de 1497 y 1553. *Revista de Demografía Histórica*. 31(1):105–130. https://ddd.uab.cat/record/166318

Kuznets, Simon, and John Thomas Murphy. (1966). *Modern economic growth: Rate, structure, and spread*. Vol. 2. New Haven: Yale University Press.

Lambert, Paul S., Richard L. Zijdeman, Marco HD Van Leeuwen, Ineke Maas, and Kenneth Prandy. (2013). The construction of HISCAM: A stratification scale based on social interactions for historical comparative research. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 46(2):77–89. Doi.org/10.1080/01615440.2012.715569

Lee, James Z., Cameron D. Campbell, and S. Chen. (2010). China multigenerational panel dataset, Liaoning (CMGPD-LN). In *User guide*: 1749–1909. Inter-university Consortium for Political and Social Research.

Mandemakers, Kees. (2002). Building life course datasets from population registers by the Historical Sample of the Netherlands (HSN). *History and Computing*. 14(1–2):87–107. Doi:abs/10.3366/hac.2002.14.1–2.87

Martínez-Galarraga, Julio, and Marc Prat. (2016). Wages, prices, and technology in early Catalan industrialization. *The Economic History Review*. 69(2):548–574. DOI: 10.1111/ehr.12127

Mas, Joan, Fornés, Alícia and Lladós, Josep. (2016). An interactive transcription system of census records using word-spotting based information transfer. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*: 54–59. IEEE. DOI: 10.1109/DAS.2016.47

Nadal, Jordi. (1975). *El fracaso de la revolución industrial en España: 1830–1914*. Barcelona: Ariel.

Nadal, Jordi. (1992). *Moler, tejer y fundir: estudios de historia industrial*. Barcelona: Ariel.

Peytavi Deixona, Joan. (2010). *Antroponímia, poblament i immigració a la Catalunya moderna: l'exemple dels comtats de Rosselló i Cerdanya (segles XVI–XVIII)*. Vol. 20. Barcelona: Institut d'Estudis Catalans.

Porter, Theodore M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton: Princeton University Press.

Prandy, Kenneth. (2000). Class, the stratification order and party identification. *British Journal of Political Science*. 30(2):237–258. https://www.jstor.org/stable/194274

Pujadas-Mora, Joana Maria, Alicia Fornés, Josep Lladós, and Anna Cabré. (2016). Bridging the gap between historical demography and computing: tools for computer-assisted transcription and analysis of demographic sources. In *Future of historical demography: upside down and inside out*, edited by Matthijs, Koenraad, Saskia Hin, Jan Kok, and Hideko Matsuo. Leuven ACCO: 222–231.

Pujadas-Mora, Joana Maria, Juanjo Romero Marín, and Conchi Villar. (2014). Propuestas metodológicas para la aplicación de HISCO en el caso de Cataluña, siglos XV–XX. *Revista de Demografía Histórica*. 32(1):181–220. https://ddd.uab.cat/record/189774

Reher, David S. (2004). Family ties in western Europe. In *Strong family and low fertility: A paradox?*, edited by Gianpiero Dalla Zuanna and Giuseppe A. Micheli. Dordrecht: Springer, 45–76.

Reher, David S. and Angeles Valero Lobo. (1995). *Fuentes de información demográfica en España*. Vol. 13. Madrid: Centro de Investigaciones Sociológicas.

Rubio, Agustín and Mateu Rodrigo Lizondo. (1997). *Antroponímia valenciana del segle XIV: nòmines de la ciutat de València (1368–69 i 1373): estudi i índexs*. Vol. 38. València: Universitat de València.

Ruggles, Steven. (2014). Big microdata for population research. *Demography*. 51(1):287–297. Doi: 10.1007/s13524-013-0240-2

Ruggles, Steven, Evan Roberts, Sula Sarkar, and Matthew Sobek. (2011). The North Atlantic population project: Progress and prospects. *Historical Methods*. 44(1):1–6. Doi:10.1080/01615440.2010.515377

Schürer, Kevin. (2007). Focus: Creating a Nationally Representative Individual and Household Sample for Great Britain, 1851 to 1901-The Victorian Panel Study (VPS). *Historical Social Research/Historische Sozialforschung*. 32(2):211-331. https://www.jstor.org/stable/20762213

Tribó, Gemma. (1989). *Evolució de l'estructura agraria del Baix Llobregat, 1860–1931*. Doctoral Thesis doctoral. Departament d'Història Contemporània. Barcelona. Universitat de Barcelona.

Van Leeuwen, Marco HD, Ineke Maas, and Andrew Miles. (2002). *HISCO: Historical international standard classification of occupations*. Leuven, Leuven University Press.

Villavicencio, Francisco, Joan Pau Jordà, and Joana M. Pujadas-Mora. (2015). Reconstructing lifespans through historical marriage records of Barcelona from the sixteenth and seventeenth centuries. In *Population reconstruction*, Bloothooft, Gerrit, Peter Christen, Kees Mandemakers, and Marijn Schraagen, (eds). Springer, Cham, 199–216.

Weir, David R. (1993). Family reconstitution and population reconstruction: two approaches to the fertility transition in France 1740–1911. In *Old and New Methods in Historical Demography*, edited by David Reher and Roger Schofield. Oxford, Clarendon Press: 145–58.

Woolf, Stuart. (1989). Statistics and the modern state. *Comparative Studies in Society and History*. 31(3):588–604.

Wrigley, Edward Anthony and Schofield, Roger. (1989). *The population history of England 1541–1871*. Cambridge, Cambridge University Press