

A visual data analysis for determining the geographical extent of the *cabreves*

Benito Zaragozaí^a, Pablo Giménez-Font^b

^aDepartment de Geografia, Universitat Rovira i Virgili, C/ Joanot Martorell, 15, 43480 Vila-seca, Tarragona, Spain.

^bInstituto Interuniversitario de Geografía, Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690, Alicante, Spain.

ARTICLE INFO

Keywords:

hgis
text-mining
geoparsing
data visualization
digital archives
historical cadastre

ABSTRACT

The historical cadastral archives are an important source of information to help understand our cultural heritage since they contain a trace of the activities, land uses, and buildings developed by people from different periods. However, in the era of Big Data there remain many historical documents of great value that have not been digitized or studied in depth. This is the case of the *cabreves*, which are pre-cadastral documents used for centuries in several regions of Spain to document those properties that were subject to the payment of taxes to a feudal lord. Rescuing these data would enable studying the landscape structure of relatively recent dates for which there is no cadastral cartography. However, it is difficult to establish the state of conservation, degree of accessibility, content detail, and quality of the archived *cabreves*. In recent years, progress has been made in digitizing these sources. In Spain, the Spanish Archives Portal (PARES) harmonizes and unifies the efforts of national archives, and a significant number of documents have been archived in recent years. We use text mining techniques to analyze and map the records in which *cabreves* appear. Out of the 1,752 records found, a total of 1,408 *cabreves* have been geocoded and mapped, enabling us to establish which territories and periods can be studied using these sources. From this experience, we request that digital archives maintain a geographical perspective during archival appraisal.

1. Research aims

The main goal of this work is to estimate the abundance, accessibility, and geographic distribution of the *cabreves* available in the most representative repositories. The *cabreves* are a good representation of the quality standards used to code cadastral information before the first cadastral maps of Spain were developed in the 19th century. In a previous research we demonstrated that the accuracy and the detail of the descriptive information contained in these documents could be enough to reconstruct the landscape structure and generate geographic information that could be incorporated into an Historical Geographical Information System (HGIS). To meet these objectives, the following steps are proposed:

1. Select a representative repository for the entire area of study of the Crown of Aragon. Given the great wealth of sources and their diversity, it is necessary to focus this first study on a source that is sufficiently representative.
2. Propose a reproducible and extensible analysis methodology that enables the available information to be easily visualized. This methodology should also be applicable to other repositories or archives.
3. Determine the main characteristics of the *cabreves* found. It is useful to establish the conservation status, digitalization, possibility of downloading, among other features.
4. Describe the spatial and temporal distribution of the available *cabreves*. It is necessary that the available *cabreves* can be shown on maps.

2. Introduction

Digital historical archives have grown in volume and functionality in recent decades. These new resources enable us to address research questions that had not been previously raised due to the enormous dispersion of sources in a

ORCID(s): 0000-0003-2501-484X (B. Zaragozaí); 0000-0002-1171-715X (P. Giménez-Font)

multitude of archives and databases with limited accessibility. This fact is particularly relevant in Spain—a nation with one of the largest documentary archives in the world (Conde Villaverde, 2006)—before the modern cadastres where compiled (Camarero Bullón and García Juan, 2018). In this paper, we focus on the *cabreves*, a type of notarial document, of a cadastral nature, that was widely used in the Crown of Aragon and later in the Kingdom of Spain, between the 12th and 19th centuries. In the following sections we explain the importance of this type of document in relation to the evolution of land ownership and in the reconstruction of the historical landscape. We then review the information on *cabreves* available in the Spanish digital archives and suggest the need for a methodology with which to better understand the geographical and historical scope of the *cabreves* and similar documents.

2.1. The historical landscape configured by *emphyteusis*

The historical administration of European countries, based on Roman law, has generated huge amounts of documentation since the Middle Ages. Many of the documents in the archives are related to the ownership of property and land, which implies an undoubted geographical value. For this reason, historical geography has used this type of document to conduct numerous studies related to the structure of the property and the agricultural landscape at a certain date, along with its subsequent evolution (Carrion, Migliaccio, Minini and Zambrano, 2016; Turner, Bolòs and Kinnaird, 2017; Szabó, Suchánková, Křížová, Kotačka, Kvardová, Macek, Müllerová and Brázdil, 2018; Frajer and Fiedor, 2021).

The process of property consolidation that took place in Europe between the Middle Ages and the beginning of the Modern Age generated various types of property records based on a simple descriptive structure (owner-boundaries) without many spatial references but of considerable geographical interest. Although there are several types of relationship between feudal rights and “fine-paying tenants”, the use of *emphyteusis* or copyhold lease in several regions of Spain, Italy, and France stands out (Béaur, Congost and Luna, 2018). This type of contract (or tenure regime) was often used in the Middle Ages in areas in which the Christian kingdoms had advanced over the territories of Al-Andalus, and especially in the Crown of Aragon, which included the kingdoms of Valencia, Aragon, Mallorca (Balearic Islands), and the Principality of Catalonia (Figure 1). It was a formula that favored the colonization of the territory, as it represented an attractive contract for settlers that was also beneficial for the feudal lord. Specifically, the property was divided into two parts: the lifetime property (freehold) was reserved for the Lord (*dominium directum*) and the copyhold (*dominium utile*) was reserved in perpetuity for the settler by means of an annual payment, and could be transmitted to their descendants (Congost, 2003). This forced a strict control of assets and payments through property records known in the Crown of Aragon as *cabreves* (in Spanish), *Capbreus* (in Catalan) or *Capibrevi* (in Latin). During the 17th century and, especially the 18th century, the generation of documents relating to property was multiplied by the bureaucratization of the manors and by the multiplication of notaries (Peris Albentosa, 1995). With the Liberal Revolution and the dismantling of the regime in the mid-19th century, the *cabreves* became basic documents in the costly process of property consolidation, carried out by hundreds of thousands of *emphyteutic* copyholders who managed to become absolute owners. The magnitude of the process was enormous, especially in the regions of Valencia and Catalonia. In the case of the Kingdom of Valencia, for example, the properties of the nobility are proportionally now very scarce, but at the beginning of the 19th century three-quarters of Valencia – now a regional state – were under the jurisdiction of a noble (Gil Olcina, 1998). As a result, in many places a land ownership similar to the original is maintained, as occurs in Catalonia, where several studies have shown how the families of current owners have descended from the ancient *emphyteutic copyholders* represented in the *cabreves* (Congost, 2015).

2.2. Importance of the *cabreves* as a source of geographical data

The *cabreves* are public deeds linked to the *emphyteusis* or copyhold lease, the predominant legal formula in the Crown of Aragon – one of the most important crowns of the European Mediterranean and founded in the 12th century. *Cabreves* are pre-cadastral sources in which the *emphyteutic* copyholder declares the copyhold (*dominium utile*) on real estate and land, recognizing the *dominium directum* of the feudal Lord (Gil Olcina, 2012). The information contained in the *cabreves* is very detailed when describing the properties of each *emphyteutic* copyholder with surface size, location, and limits data, as well as annual taxes paid for each of the properties. In many cases, there is a succession of books over time (decades or centuries), which enables historical studies on the structure of the property, types of crops, or hereditary transmission (Benítez, 1984). The *cabreves* also contain an abundant number of place names. Some may still be valid, although sometimes they may have changed in their pronunciation or spelling. Some place names have disappeared for different reasons, such as substitution related to changes in ownership or changes in land use, such as the urbanization of a rural area. Toponymy is a fundamental element for localization and also provides precise

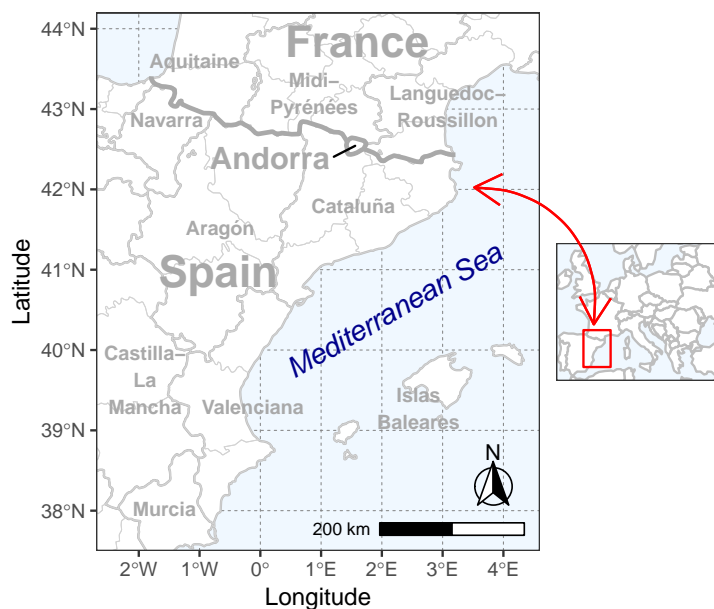


Figure 1: Location map of the study area. The region of interest includes the borders of the former Crown of Aragon, between 37.5° and 44.2° north latitude and -2.7° and 4.6° in longitude. The scale bar is for reference only.

information about local history and enables the reconstruction of a historical landscape (Conedera, Vassere, Neff, Meurer and Krebs, 2007; Antonson, 2009). Figure 2 shows one page of the *Cabreve* of Algerri (1,779), a municipality in the province of Lleida, in the Catalan region of La Noguera. In this Figure, the typical structure of a *Cabreve* is highlighted using different colours.

The importance of the *cabreves* as a geohistorical source of the Crown of Aragon is widely known (Congost, 2009). However, as a documentary source, there are two main drawbacks or limitations. Firstly, although it was recommended to perform a *cabreve* every 20 years, this was rarely achieved and there was no standardized periodization. In most cases, only one or two *cabreves* have been preserved for the same manor, which limits comparative studies in different periods, other than the year of the *cabreve* and the present. *Cabreves* are a textual source without cartographic references, although spatial references are useful enough to be used as a fundamental source in the reconstruction of the historical organization of a given territory and its landscape (García Juan, Álvarez Miguel, Camarero Bullón and Escalona Monge, 2012). In Zaragozaí, Giménez-Font, Belda-Antolí and Ramón-Morte (2019) the geographical potential of this document source has been reviewed with the application of new information management techniques, verifying that the *cabreves* can be analyzed using semi-automatic techniques based on the graph theory. These techniques are useful for reconstructing the structure of the landscape and provide new evidence on the location of archaeological sites. However, it was also demonstrated that even if many tasks can be automated, the process can be very laborious and disproportionate when analyzing even a small manor. Consequently, it is worth asking to what extent these methodologies can be extrapolated to the study of *cabreves* from other areas, or to the study of documents with a structure similar to that of the *cabreves*. This issue can only be resolved if the number and availability of the documents kept in the archives is known exactly. The more documents available, the more beneficial it will be to continue developing new software applications such as those proposed in Zaragozaí et al. (2019).

2.3. The *cabreves* in the digital archives

The development of the internet has caused an unprecedented increase in the accessibility of all types of information (Zeng, 2017; Anderson, Eaton and Schwartz, 2015). This growth has led to the emergence of new digital archives and new lines of work, such as family history studies (Bainbridge, 2018) or the dissemination of information on new platforms (Kelly, 2018). In Europe, the greater dissemination of historical archives has as a prominent example the case of the Europeana portal – which includes more than 50 million archived elements (from books and archives, to multimedia elements and 3D models) from more than 3,500 European institutions (Petras and Stiller, 2017). However,

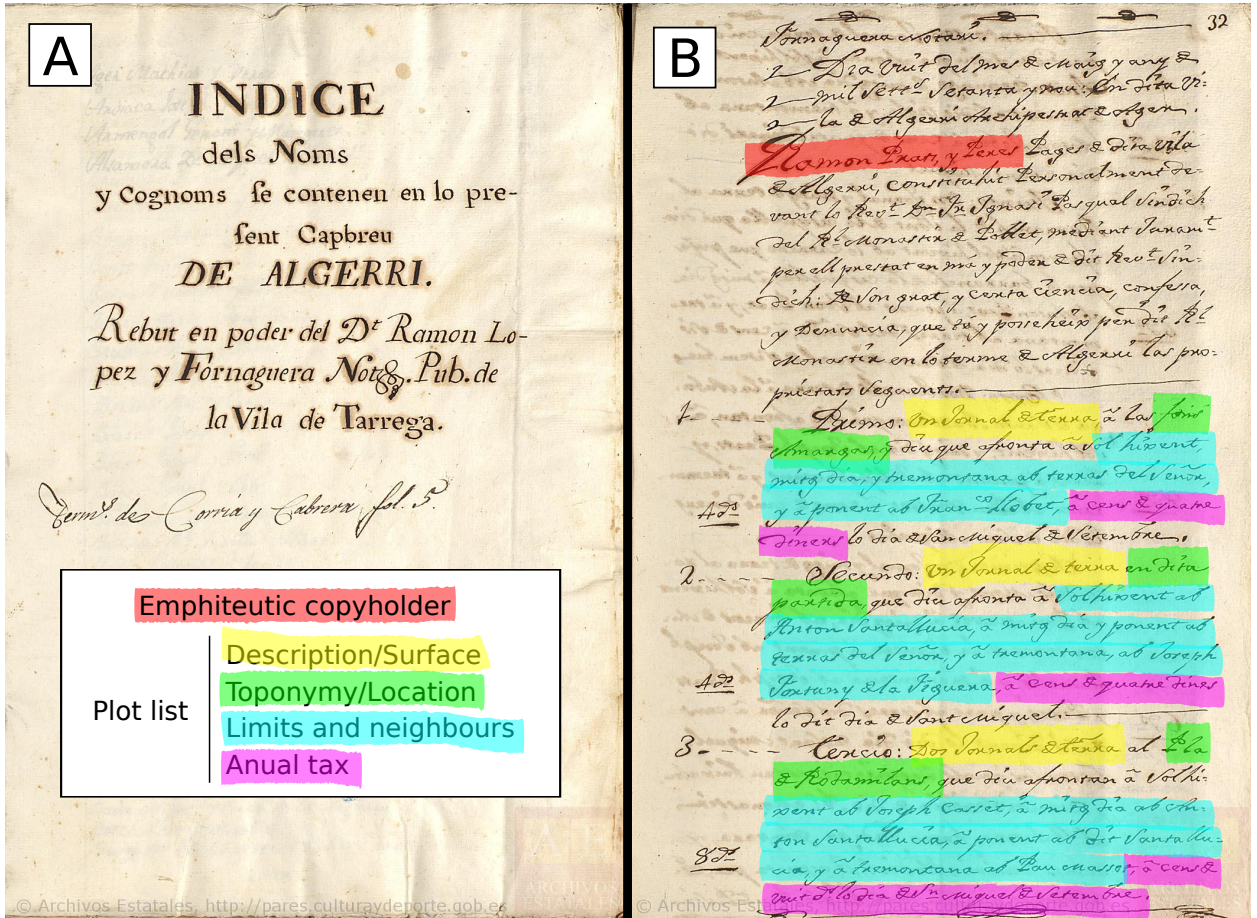


Figure 2: Sample of the information contained in the Capbreu de Algeri of 1779. Excerpts from A) the cover and B) page 32, which shows the three plots held by Ramon Pran y Peres, their basic description, location and taxes. Source: Archivo Histórico Nacional through the Spanish Archives Portal (PARES).

despite the large volume of documents already preserved and indexed, it is widely recognized that classic archival problems persist. In short, there are not enough resources to keep all the records, but strategies are needed to decide what should be kept (Hodder, 2017). Archivists must ensure in an objective way that the records provide evidence of the activities of the society as a whole, or that the records kept may have some kind of utility (Couture, 2005). Once the archivists have made decisions, the most important items are preserved and made accessible – even if only to announce that they exist – and at that point it is worth asking whether these resources are used and what benefit they generate. Such documents may be used in new historiographic studies (Sinn, 2012; Dascher, 2015; Anderson et al., 2015). Naturally, all these issues also directly affect the archiving, digitization, and publication of the *cabreves* on the internet.

Digital archives are usually built with a “bottom-up” philosophy, so the most important platforms add information from other smaller portals that usually have fewer resources, whether they are portals of private foundations or regional portals maintained by the administration. In Spain, there are several portals on the internet with references to *cabreves*. Generally, these archives follow international standards, regardless of their internal structure, such as the General International Standard Archival Description – ISAD (G). There is some basic data that all archives usually collect.

There are several relevant digital archives in which we could find *cabreves*, but in this work we focus on the most important archives in terms of the number of indexed documents. In the Valencia region, we can highlight several digital archives in which we find *cabreves* that have already been digitized because a major regional effort has been made to manage the most important public archives. The Sistema Arxivístic Valencià En Xarxa (SAVEX; <https://savex.gva.es>) has a census of documentary heritage that includes the main Valencian municipal and

regional archives. From this platform it is possible to access 185 references to *cabreves*, of which only six use the word *capbreu* (the term in Catalan).

Secondly, in the region of Catalonia we find an online archive that adds a large number of *cabreves* available for its territory (see <http://arxiusenlinia.cultura.gencat.cat>). The regional nature of this web portal is very marked, and it becomes clear when searching for *cabreves* in different languages. In Catalan (*capbreu*) there are up to 2,982 documentary units belonging to 290 documentary funds. When using the Spanish terms the number of documents found is significantly smaller (16 documentary funds and 53 documentary units). In addition to the richness of this repository we can highlight that many of the documents are digitized in the form of images, which are available for direct download. For example, indexed from the Arxiu Comarcal Baix Camp, we found a *Cabreve* of Savassona – in the province of Barcelona – dated between 1142 and 1143.

Thirdly, we can highlight the PARES (Spanish Archives Portal; <https://pares.mecd.gob.es>). This is intended for the dissemination of Spanish documentary historical heritage through access to the digitized funds of the main national archives. Unlike previous databases, PARES covers all of Spain and to achieve its main objective it requires the unification and harmonization of the large Spanish archives: Archivo de la Corona de Aragón; Archivo Histórico Nacional; Archivo General de Indias; Archivo Histórico de la Nobleza; Archivo General de la Administración; Archivo Central de Cultura; Archivo General de Simancas; Archivo de la Chancillería de Valladolid; along with other regional archives and documentary centers. Taking into account all possible search terms, the number of entries containing *cabreves* in the PARES currently reaches 1,752 – and is the largest set of items found in the archives. It is worth mentioning that PARES is the main source of these documents for several European platforms such as Archives Portal Europe (<https://www.archivesportaleurope.net>) or Europeana (<https://www.europeana.eu/portal/en>). In Figure 2 we show an extract from a well-documented *cabreve* indexed in PARES. However, despite the great integration effort undertaken at PARES, there are still many other regional, local, or private archives whose resources are not listed on this platform.

Finally, it is worth mentioning that digital archives are not the only source where *cabreves* can be found on the internet. In addition to the above-mentioned portals, numerous historical and historiographic studies of a local nature have been published in which *cabreves* are analyzed in some detail. There are many local studies published in magazines and books in which various aspects of the landscape and society are analyzed – the Sibhill-la portal (<http://sibhilla.uab.cat/>) of the University of Barcelona contains more than 200 works based on *cabreves*. These types of studies often contain images or transcripts made by the authors themselves. Among these types of studies we can highlight research based on *cabreves* that date back to 1291 (Puig Costa, 2013), toponymic studies rescuing *cabreves* that were believed to have disappeared (Corredor Plaja, 2001), or even landscape reconstruction studies where it was possible to draw the map of a municipality in the 17th century solely on the information contained in a *cabreve* (Torres Gros, 1996).

In our work, we aim to analyse the metadata contained in the aforementioned digital repositories. Therefore, we extract some useful information contained in the records containing references to *cabreves*, analyse this information, and evaluate its availability for being introduced in a HGIS. This approach shares similarities to some studies on processing old cadastral maps in HGIS (Harvey, Kaim and Gajda, 2014; Frajer and Fiedor, 2021), building 3D models from 2D cartographic or photographic sources (Tobiáš, Cajthaml and Krejčí, 2018; Gatta, Ariotti and Bitelli, 2017). However, in this work we are building 2D models from simple textual references.

3. Sources and methods

Given the dispersion of sources mentioned above, we decided to focus this study on the Spanish Archives Portal (PARES), which is a reference portal in Spain and is described in detail in the scientific literature (Sánchez Mairena, 2014; Pacios and La Torre Merino, 2018). Our methodology consists of locating all the *cabreves* in which it is possible to identify at least one place name and the dates (start-end). At the same time we also extracted other data of potential interest (such as the number of pages, conservation status, language, etc.). We then geocoded the place-names found to create a geographic information system (GIS) database with which to generate statistics, graphs, and maps that help us better understand the characteristics, availability, accessibility, and spatio-temporal distribution of the *cabreves*. The processes and analyzes performed in this work have been developed in the form of R scripts (R Core Team, 2017), and we have used free software and open data in containerized reproducible workflows Zaragozí, Trilles and Navarro-Carrión (2020).

3.1. Spanish Archives Portal (PARES)

PARES was inaugurated in 2007 and currently offers descriptive records of documents that may or may not be digitized. Of the 5,335,059 indexed descriptive records, 31.1% have been digitized – and users are encouraged to offer the maximum descriptive information so that they can discover the existence and locate the maximum number of documents. Within these records the number of images or digital objects amounts to 34,187,941 (<http://pares.culturaydeporte.gob.es/estadisticas.html>). The digitization process is selective and documentary collections are chosen by virtue of specific criteria such as the age of the manuscripts, demand by researchers, or the importance of the information recorded (Sánchez Mairena, 2014). The organisation in charge of digitization is the Document Reproduction Service of the State Archives (SRDAE). However, it is important to note that the database of description units and signatures are created and managed by the source archives. Together with the aforementioned “bottom-up” philosophy, the resources indexed in the Spanish Archives Portal are located in international repositories such as APEX (Archives Portal Europe network for excellence) or Europeana (www.europeana.org). It is necessary to remember that not all the information available in PARES is available through its public search engine. The more professional IntraPARES platform holds a large percentage of the records that are still being processed by the archivists.

3.2. Web crawling and data scraping

PARES publishes useful information to achieve the objectives proposed in this work and this information is presented in such a manner that it can be extracted using web mining techniques to create a structured database that can be analyzed in various ways. These techniques have been applied in many contexts and there are numerous examples in social sciences of the use of crawlers to gather data to research an aspect of web use or web publishing (Thelwall and Stuart, 2006).

There are numerous tools to perform the web-crawling and web-scraping processes. In this study we use a GNU R library called RCrawler (Khalil and Fakir, 2017). RCrawler can crawl, parse, manage pages, extract contents, and produce a dataframe that can be directly employed for analyzing the PARES webpage contents. The manual provided with this package facilitates its use, but tasks performed with crawlers or ‘bots’ can cause problems if misused in a variety of ways. For example, without prior planning, there is a risk of overloading the service on which consultations are carried out and degrading its operation (Thelwall and Stuart, 2006). In our case, the advanced search engine of the Spanish Archives Portal (PARES) enables easily listing all the archives that contain information about *cabreves*. The archives are only crawled once, and this generates a minimum workload for the PARES portal.

Each PARES record contains different fields that are useful for achieving our objectives. Table 1 shows the main fields extracted during the process. It should be noted that the description and date fields are of special interest because they can be processed to derive new information such as place names or the accuracy of dates.

3.3. Geoparsing resources from PARES

Dates, descriptors, and categories, as well as other data fields extracted from PARES, can be easily standardized for incorporation into any type of analysis. However, the most interesting information for this study are the place names that usually appear in the description field of each record. These place names are used to study the geographic distribution of the archived *cabreves* and their characteristics. These place-names appear within texts whose analysis may present several difficulties, the most important being that: (1) the texts may be written in various styles in various languages – Spanish, Catalan, or Latin; (2) the place-names may be confused with other places outside the study area; (3) texts may contain references to one, two, or more *cabreves*; and (4) a *cabreve* may encompass several locations – so it will also have more than one associated place name. These types of problems are relatively frequent in the study of historical sources and can be solved by text mining techniques that are commonly known as Geoparsing (Clifford, Alex, Coates, Klein and Watson, 2016; Gregory, Donaldson, Murrieta-Flores and Rayson, 2015). Geoparsing is the process of looking for place names in free text to unequivocally geolocate the referred elements. Thus, geoparsing goes beyond simple geocoding, since geocoding is based on clear geographical references – such as postal addresses – while in geoparsing there is an important component of interpretation.

Geoparsing the descriptions provided in PARES has the advantage of processing relatively small text fragments. However, there are two drawbacks that may make it difficult to extract place names: (1) the variety of sources added in the PARES – understood as the diversity of standards used in each archive; and (2) the aforementioned multilingual character entries related to *cabreves* (Spanish, Catalan, Latin, and their dialectal varieties), which also affect place names, since each place can have a different name in each language.

There are several alternatives for geoparsing historical documents. For example, the Edinburgh Geoparser (Alex,

Table 1

Attribute table produced in the content scraping phase, with an example of a real crawled record.

Field	Description	Example
id	PARES resource ID. May be repeated depending on the description number of text lines.	1491619
description	One line extracted from the PARES item description (which may be composed of one or more paragraphs).	"capbreu de biure"
date	Raw dates in multiple formats (YYYY, YYYY-mm-dd, etc)	1613
signature	Source archive signature	ACA, REAL PATRIMONIO, BGC, Volúmenes, NÚM.302
digitized	One of: (1) fully, (2) partially or (3) N/A	N/A
described	One of: (1) fully, (2) partially or (3) N/A	N/A
level	One of: (1) Unidad Documental Simple, (2) Unidad Documental Compuesta or (3) other types.	Unidad Documental Simple o Elemento de Descripción Asociado
archive link	Source archive URL of the resource	Archivo de la Corona de Aragón http://pares.mcu.es/ParesBusquedas20/catalogo/description/1491619

Grover, Tobin and Oberlander, 2019), the Pleiades Gazetteer (Paterson and Gregory, 2019), or other platforms such as the World-Historical Gazetteer (Mostern, 2017), are very effective in certain contexts. However, due to the specific needs of this work, it was decided to perform a custom analysis using Nominatim (<https://nominatim.openstreetmap.org/>), which is a free and general open source software nomenclator. If the number of queries is high, Nominatim enables installation on a computer or server, and this offers great flexibility. In addition, if the results are unsatisfactory, it is possible to expand the list of place names as necessary.

The proposed analysis consists in filtering a list of stop words as a previous step to geocoding the remaining words. Usually, the term 'stop words' refers to the most common words in a language that are considered less important words in different applications (Nothman, Qin and Yurchak, 2018), but in this work the stop words are all those words that could never be a place name in the study area – or be part of a place name. The creation of these lists is a semi-automatic and iterative task in which the counts of all occurrences of words that appear in the PARES descriptions are reviewed.

Nominatim enables defining a search box – a geographical area – that avoids retrieving place names that are outside the region of interest. Blurred by the passage of time, the study area of this work would be one in which at some point in history, *cabreves* were used to administer the rights associated with *emphyteusis*. This area will be difficult to define precisely as it strongly depends on the available sources and the methodology used. However, if we cannot approximate a plausible study area, we run the risk of making serious mistakes when geoparsing place names. For example, among the records of PARES we find several *cabreves* referred to San Juan de Jerusalem – The Sovereign Military and Hospital Order of San Juan Jerusalem, Rhodes, and Malta, better known as the Order of Malta. If the list of stop words is not sufficiently complete, it is likely that the term Jerusalem would be geocoded, offering a clearly erroneous geolocation. This study has tried to define a search area wide enough so that no place is excluded, but tight enough to avoid cases like the one just mentioned. The search box is shown in the map in Figure 1 and in the other maps included in this work.

The workflow with these tools is represented in Figure 3. Once all stop words in the description field have been eliminated, several cases are raised that affect the geocoding process. Figure 3 shows an activity diagram detailing the most common cases. Mainly, we can specify three cases: (1) there has been a single word that can be a place name; (2) there have been several words that can be part of a place name or refer to several locations within the same *cabreve*; and (3) there is no word left. The latter case may occur for various reasons, for example, it may be the case that the description refers to the *cabreves* collected by an individual and no place is explicitly mentioned. As seen in the diagram in Figure 3, cases 1 and 3 are quite straightforward. Case 1 can be consulted directly in a geogazetter, while

case 3 would be automatically ruled out. In contrast, the entries in case 2 relate to more than one possible place name. This means that the words found can: (1) be part of a single compound toponym (e.g. “Capbreu of the monastery of Sant Pere de Camprodon”); (2) that several are used (e.g. “Capbreu de Prats, All, Ger i altres”); or (3) that only one of the words is a place name and the others have not been filtered as stop words. In any case, even though they are real place names belonging to a historic period, it is possible that they are not included in the geogazeteer. According to each case, the entries of *cabreves* found in the PARES will have a spatial representation in the form of a point (or multipoint in cases such as “Capbreu de Prats, All, Ger i altres”). The result will be a spatial database with which to determine the main characteristics of the *cabreves* found and describe their spatio-temporal distribution.

3.4. Visualization tools

In this study, several common thematic charts and maps have been created, but it has been appropriate to use some specific visualization methods to analyze categorical data. Specifically, alluvial diagrams have been used (Rosvall and Bergstrom, 2010). These visualizations provide a better picture of the interrelation between proportions and interactions between variables (Bezerra, Silva, Guedes, Silva, Leitão and Saito, 2019). These methods are based on a tabulation of data values that show the frequencies of each value or group of values in the data set. The *ggalluvial* library (Brunson, 2017) – also available on the R GNU platform – greatly facilitates the creation of this type of graphic from the tools already indicated. We have not been able to find historiographic studies in which this type of visualization has been applied to analyze a digital archive. However, there are some examples in other disciplines that suggest its considerable potential for analyzing the content of PARES. For example, these diagrams have been used in studies of digital innovation ecosystems (Chae, 2019), bibliometric analysis (Yeung, 2018), evaluating clusters of public transport users Gutiérrez, Domènech, Zaragozaí and Miravet (2020) or – more related to the present study – in an exploratory data analysis (Bezerra et al., 2019).

4. Results and discussion

When the web crawling was finished, the PARES had 1,752 entries labeled with the alternative terms for *cabreve*–*Cabreo* (Spanish), *Capbreu* (Cat), *Capibrevi* (Lat) – of which it turned out that 152 entries contained more than one *cabreve*, while 241 did not contain the word *cabreve* in the description. Finally, 483 different place names related to 1,511 *cabreves* from different dates were found. Among the different attributes found in PARES there are several that are found in all entries, and so they are very useful for analyzing this database as a whole (see Table 1). Logically, there may be a significant number of *cabreves* pending review and publication (IntraPARES), so the number of available entries could change relatively quickly.

Regarding the methodology, one of the main results of this work is the stop word list. In total, 764 terms specific to the case study of the *cabreves* have been listed – summarizing those of different languages – and 66 people names. In addition, standard *stop word* lists have also been used in the three main languages: Spanish (734 words), Catalan (280 words), and Latin (50 words). The most common deleted words are usually full names of people, professions, noble and ecclesiastical titles, courtesy formulas, vague geographical references, types of legal documents, Roman numerals, archival codes, and abbreviations.

In Figure 4 it can be clearly seen that all of the *cabreves* found are from four archives. The Archivo de la Corona de Aragón (ACA) is the most productive, and the Archivo de la Memoria Histórica (AMH) provides a single *cabreve* that turns out to be an out of context mention. As explained above, the most recent *cabreves* date from the middle of the 19th century, with the beginning of the process of dissolution of the old regime and the liberal revolution. However, the remarkable increase in the number of the *cabreves* from the 16th century must be related to the development and consolidation of hundreds of manors. This translates into the generation of numerous lawsuits between members of the family that own the manor, and between the manor and an increasingly solid and complex administration. Not forgetting the rise of some *emphyteutic* copyholders, who from the 18th century, and especially during the 19th century, confronted the owners of the manors. In 25 cases, *cabreves* have been found for which there is no specific date in the description provided by PARES.

Figure 5 shows the degree of accessibility of these *cabreves*, if they are archived, described, or even digitized. Although there is a significant number of described and/or digitized *cabreves*, it should be noted that in most cases this information does not appear on the website (*not specified* in Figure 5).

This type of chart (see Figures 4 and 5) may have some utility to quantify a specific aspect of the information extracted, but the truth is that they have limitations when it comes to categorical data. For example, with this type of

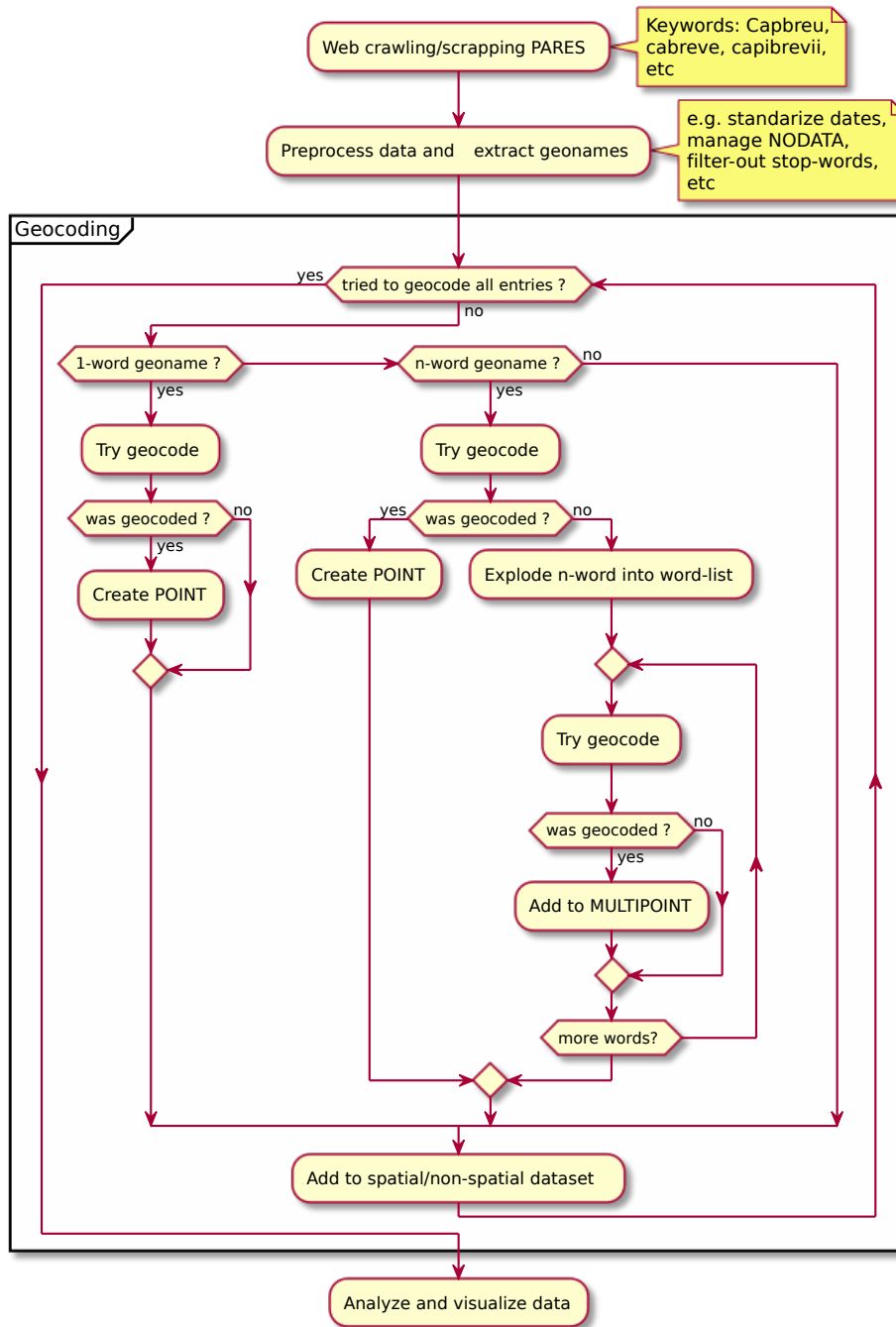


Figure 3: Activity diagram of the methodology with an expanded description of the geocoding process.

representation it is difficult to understand whether the *cabreves* that have been fully digitized are those that have been fully described. For this reason, to understand the characteristics of the archives being analyzed we have decided to use alluvial diagrams to show the relationships between the columns of the database (see Figure 6).

An alluvial diagram is very suitable for analyzing the data extracted from PARES. Figure 6 shows the relationships between some of the attributes obtained during the web crawling phase. In an alluvial diagram, the variables are presented in parallel vertical axes, while the values are represented in the form of blocks within each axis. Finally, the

A visual data analysis for determining the geographical extent of the *cabreves*

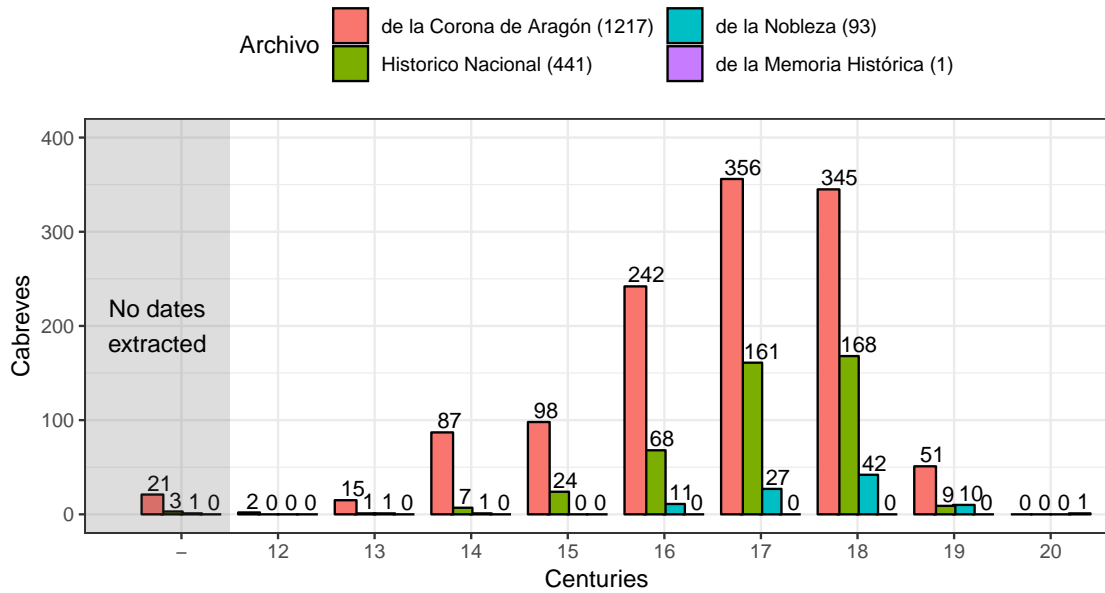


Figure 4: Repositories that contributed *cabreves* to PARES.

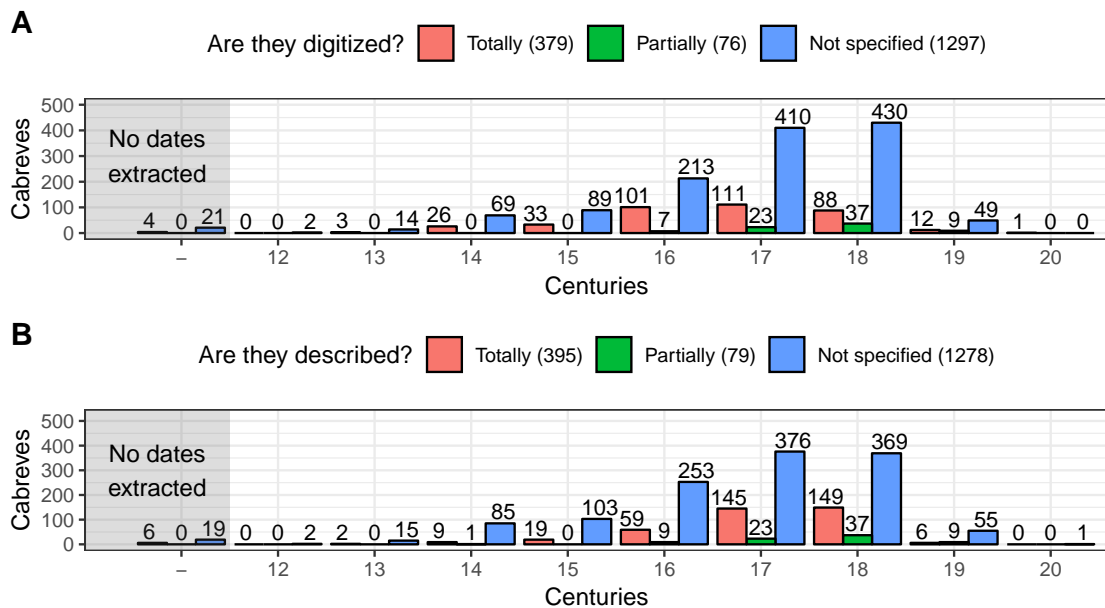


Figure 5: Degree of accessibility to information.

observations are represented as alluvia (sing. "Alluvium") that cross through the axes showing relationships. During the creation of these diagrams there are certain decisions that influence their readability. For example, in Figure 6 each alluvium represents a *cabreve*, so the way in which the different *cabreves* converge, diverge, or go parallel between axes and blocks, highlights interesting facts contained in the data. The order of the axes and blocks influences the readability of the diagram, but with a few adjustments we obtain a diagram with great visual power. These diagrams provide a great capacity to establish comparisons that cannot be easily achieved using barplots.

Analyzing the first alluvial (Figure 6) the importance of the source archive is clear in the attributes extracted from

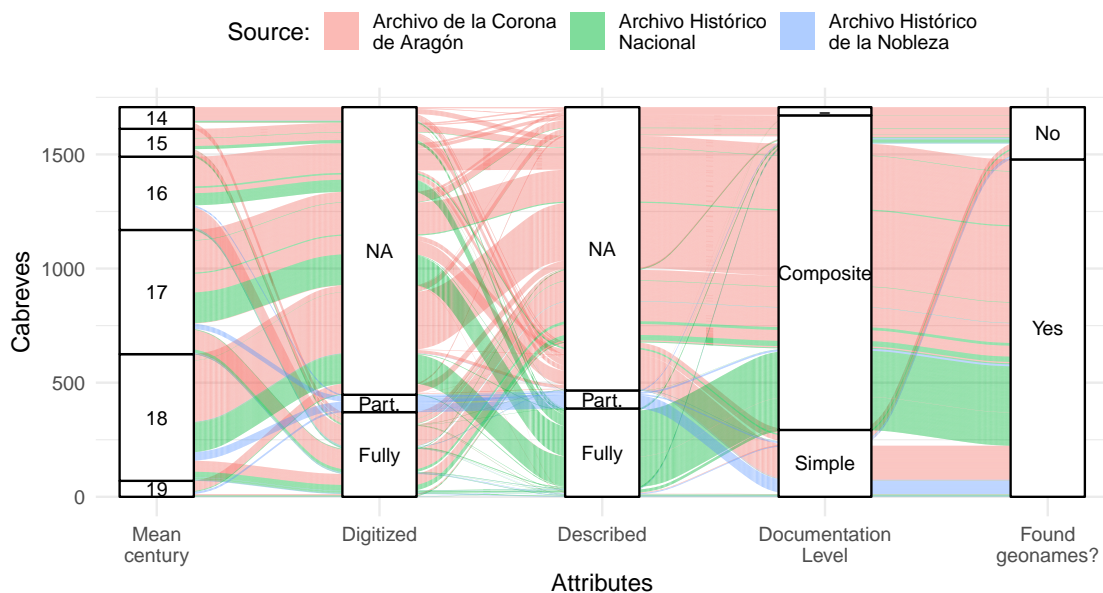


Figure 6: Alluvial diagram showing the structure of the data downloaded and the geonames extracted. The 12th and 13th centuries have been excluded to enhance readability. Documentation level categorized as simple, composite, and others (four categories) for readability.

the PARES. As for the distribution by dates, this can also be seen in Figure 4, but the difference between different archives is also evident. Thus, while ACA provides most of the fully digitized archives, Archivo Histórico Nacional (AHN) provides all of those that have been fully described. In the case of ACA, in line with the splendor of the Crown of Aragon, it is interesting to note that there is a large number of medieval documents with a very considerable historical and heritage value. The Archivo Histórico de la Nobleza (AHNob) indicates that all its documents have been partially digitized and partially described. Similarly, the vast majority of documents provided by the AHN belong to a level of composite documentation, while simple archives predominate in the AHNob. In both cases, it does not seem clear that the type of documentary unit has an influence on whether it is possible to find geographical names in the description field, but what is evident is that when one of the other four types of documentary unit is analyzed, it is not easy to extract place names. These types of documentary units refer to filing cabinets or cabinets that contain numerous documents, so their description may be somewhat less specific than in other cases. In AHNob it is significant that it is possible to find geonames in almost all of its records. However, it can also be seen that the records that have not been digitized or described include those in which geographical names cannot be found in the description.

From previous observations it is verified that PARES is an aggregator, but it does not homogenize or significantly process the data provided by the source archives. As a result, it is difficult to determine the degree of availability and accessibility of the *cabreves* without knowing how to work in each source archive. What is clear is that much remains to be digitized at ACA, but that which has been digitized has been digitized completely. ACA never describes the *cabreve* records. AHN does not digitize, but almost everything is completely described. Finally, AHNob partially describes and digitizes.

Once the downloaded information is analyzed, the geolocation of the place names is carried out according to the methodology explained above (see Figure 3). Of the 1,752 references to *cabreves*, 1,408 have been geocoded (80%) and this is a positive result, since it provides a significant sample that can contain *cabreves* distributed throughout the study area and is representative of all periods. Another issue that is quite evident is the effect caused by the antiquity of the place-names found in the geocoding process – probably due to disuse. The success rate in geocoding increases from the 12th century to the 19th century: from 60% in the 14th century to 87% in the 18th century. Therefore, this phenomenon should be given attention when investing the effort and resources to update a gazetteer.

In the alluvial diagram of Figure 8, we can analyze the success rate of geoparsing, the details of how the process was developed, information on the term used to refer to the *cabreves* in different languages, as well as the type of place

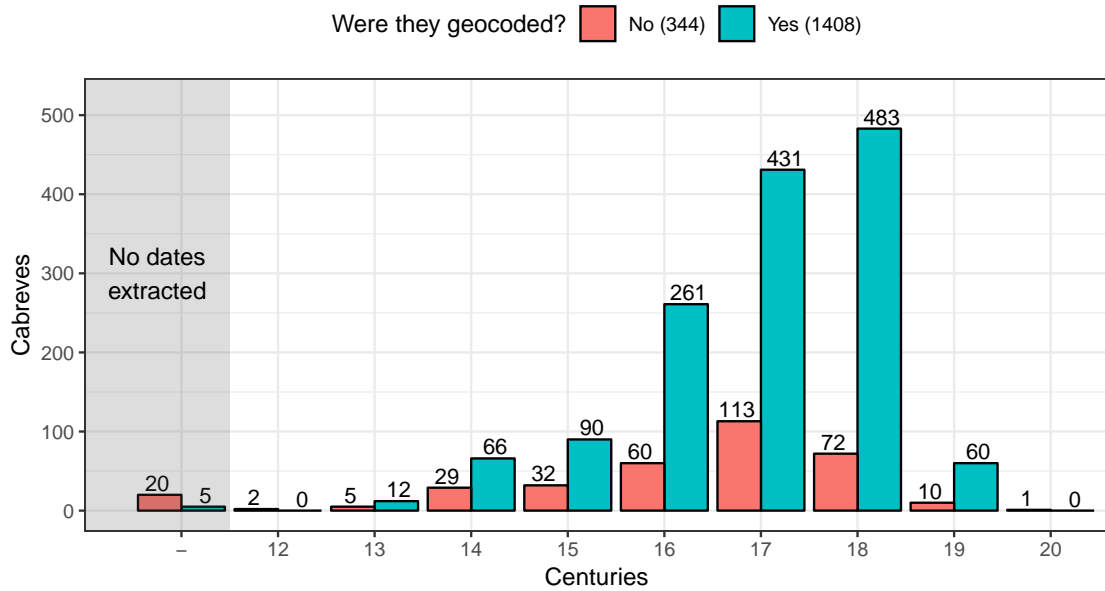


Figure 7: Success in geocoding PARES records.

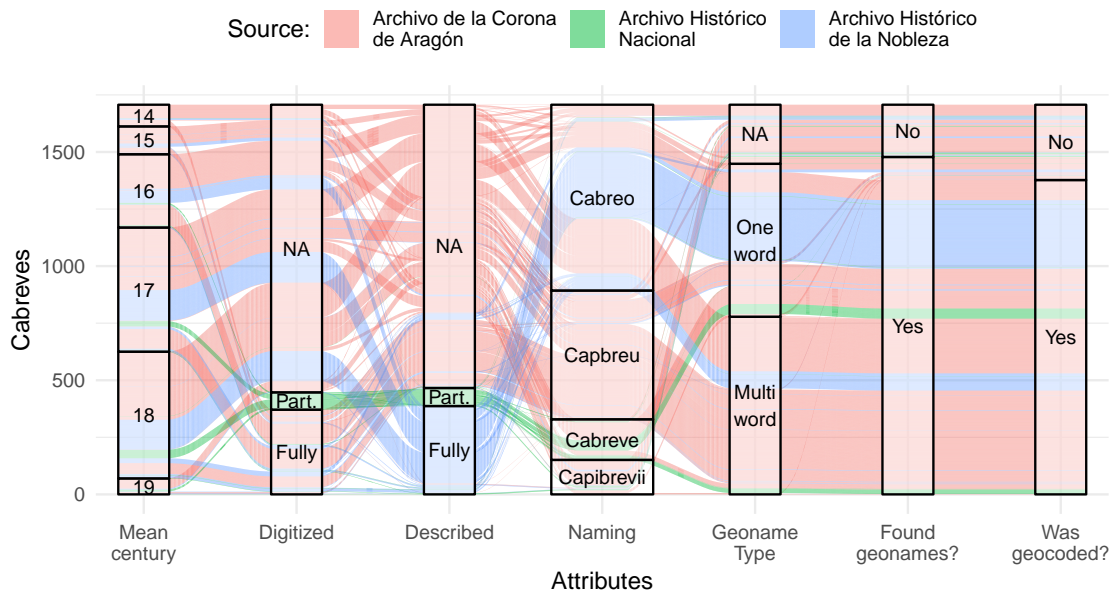


Figure 8: Alluvial diagram showing details of the geoparsing process. Excluded 12th and 13th centuries for increasing readability.

name found and whether it could finally be geocoded.

As in the previous alluvial diagram, in Figure 8 it can also be seen that the data extracted from PARES are greatly influenced by the source archive. In this case, the almost exclusive use of the Catalan term *capbreu* by the ACA, or that only the AHNob uses the term *cabreo*, is noteworthy. Obviously, the ACA uses all the terms available to the *cabreves* given its greater number of entries. Finally, the AHN opts for the use of the Spanish term *cabreve*. The Latin term *capibrevii* seems to be used by all but AHNob. Its use does not seem exclusively related to the century to which the

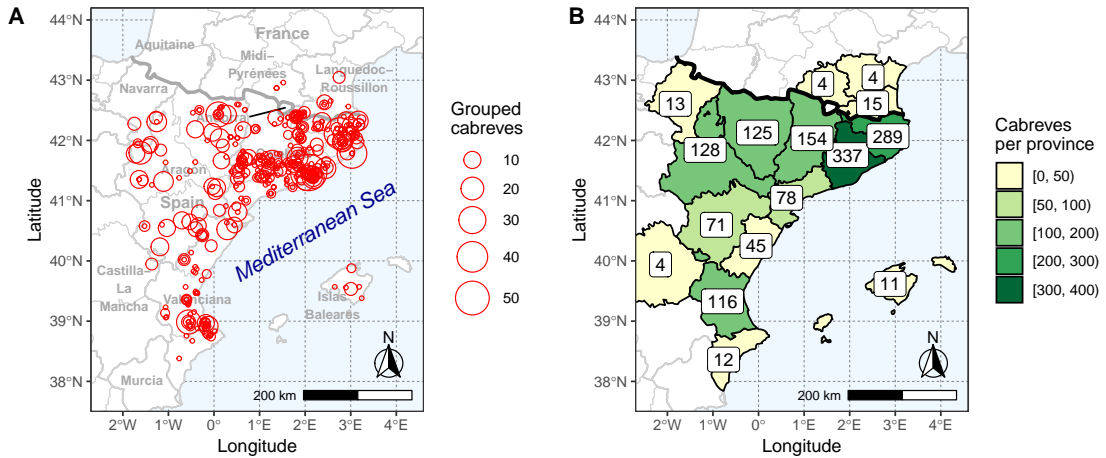


Figure 9: First distribution maps of the *cabreves*. The scale bar is for reference only.

documents belong.

The diagram shows information that is quite obvious, such as that if it was not possible to find place names in the description field, then it is logical that geocoding could not be performed either. However, there are more interesting aspects. In the case of AHNob, it also seems that there is a greater presence of single word place names, which *a priori* should facilitate the geocoding process. Finally, finding place names of one or more words does not seem to be too important when trying to geocode them. However, as can be seen below, the type of place names – according to the number of words – does affect the accuracy of geocoding.

As a final result of this work, we have created initial maps of the places where most *cabreves* were made and where most documents have been preserved. Figure 9 shows complementary representations of this distribution. Figure 9.A shows the spatial concentration of *cabreves*. Especially noteworthy is the case of Catalonia, where the oldest manors are found (documented since the 12th century) and where the ACA has acted as a reference center for archiving much of the documentation. It is also significant the concentration of *cabreves* in large manors, as is the case of the Ducado de Gandía in Valencia whose documents are kept in the AHNob. Counting the *cabreves* according to current Spanish provinces and French departments (see Figure 9.B) shows a slightly different distribution, although the *cabreves* are still concentrated in certain provinces given that cities such as Barcelona, Girona, Zaragoza, and Valencia, were also the main population centers in the 17th and 18th centuries. In the proximities of these cities, the preservation of documents in large archives has been more common than in the case of more distant manors, where the documentation has either disappeared, is in private hands, or is kept in small provincial or local archives not indexed on PARES. The scarce proportional presence of *cabreves* in the south and north of the Kingdom of Valencia is striking given that it was a highly feudal regime. There was a significant revival of the generation of documentation after the expulsion of the Moorish in the early 17th centuries (Gómez Benedito, 2018). This is mainly due to a lower level of conservation because they were small manors, or to dispersion in local archives and archives that are not yet indexed in PARES – such as the ecclesiastical archives that retain much of the documentation from the south of Alicante in the region of Bajo Segura on the border with Murcia (Gil Olcina and Canales Martínez, 2007).

As previously mentioned, the status of the processes of digitization and description of the *cabreves* is influenced by the source archives, but additionally there is also a geographical bias to consider. In Figure 9.A it can be seen that in Valencia and Catalonia there is a greater number and a greater percentage of fully digitized *cabreves* – in Valencia there is also a significant number of partially digitized *cabreves*. The level of digitization is much lower in Aragón and in the more peripheral areas. Something slightly different happens regarding the level of description of the *cabreves* by region. This, together with the aforementioned shortage of documents in the peripheral areas of the Crown of Aragón, implies the need to study other archives such as those mentioned in the introduction. As an example, the Sella *cabreve* of 1726, studied in Zaragozaí et al. (2019), does not appear in PARES nor is it due to be incorporated.

Visual data analysis can benefit from crossing even more variables on maps. Figure 11 shows a series of miniature maps in which the relationship between the use of the different terms to refer to the *cabreves* (*cabreve*, *cabreo*, *capbreu*)

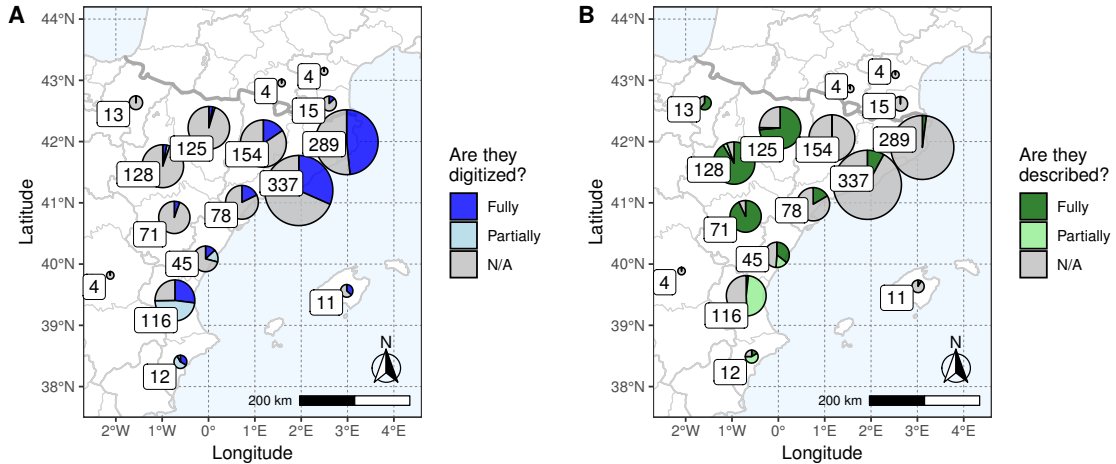


Figure 10: Per province distribution of the *cabreves* in the study area. A) Digitization status; B) Descriptions status. The scale bar is for reference only.

and *capibrevi*), the time they were collected, and the geographical distribution can be better analyzed. In addition, we consider it interesting to maintain the distinction between the archives of origin, given the bias they introduce in the PARES. In Figure 11 it can be seen that, as discussed in analyzing the alluvial diagrams, the ACA uses all the terms, while AHN tends to use “*cabreo*” and AHNob opts for “*cabreve*”. However, from the 18th century onwards, the ACA shows an increase in the use of the *cabreve* form of *cabreu*. This reflects, possibly, a change in the language in which they are written, since between 1707 and 1716 the New Plant Decrees derived from the end of the War of Succession meant the abolition of Catalan as the language of public and private administration (García Trobat, 2001). These types of representations are far from perfect (i.e. the points overlap), but they are useful for finding new patterns and generating new questions about the information available in archives such as PARES.

5. Conclusions and future work

PARES offers a high level of geographical and historical representation that facilitates various studies. In a single repository there is a large number of *cabreves* that could be used in the reconstruction and study of the historical landscape of the old Crown of Aragon. This degree of representation in turn demonstrates the good work achieved by the archives in preserving the most important documents on a digital platform.

The workflow developed in this study has been almost entirely automated (e.g. data download, geoparsing, creation of visualizations, and containerized environment). Even so, there are still phases that will require more work if the results are to be updated or the methodology is extrapolated to other types of documents or digital archives. It will be necessary to pay more attention to the following stages: (1) the creation and updating of lists of multilingual stop words – these tasks require supervision but the effort required to update these lists will be very limited given the specific nature of this type of file; (2) the validation of the results – it is necessary to assess the results by hand to determine why there are records that were not geolocated, or why the locations obtained are unreliable; and (3) the verification of the structure of digital archives and their web pages – this methodology relies on web pages or on the archive databases maintaining their structure. However, since the published data is certainly standardized, this should not be a major problem when updating the code.

The potential of visual analysis for this type of historical archive has been made evident in this work. The visualization of the data extracted with alluvial diagrams and maps help us better understand the nature of the information available in the PARES, as well as in checking some previously known historical landmarks. However, as mentioned above, although the geoparsing process and geocoding offer satisfactory results, we must not forget that the results can be improved. Even though it has been possible to geocode most of the references of *cabreves* available in PARES, there are still improvements to be added both in this methodology and perhaps in the digital archives themselves. Therefore, we consider that the difficulties described in this work should be considered when proposing a new methodology – or in the case of modifying the PARES – to better isolate the data that may be part of a historical geographic information

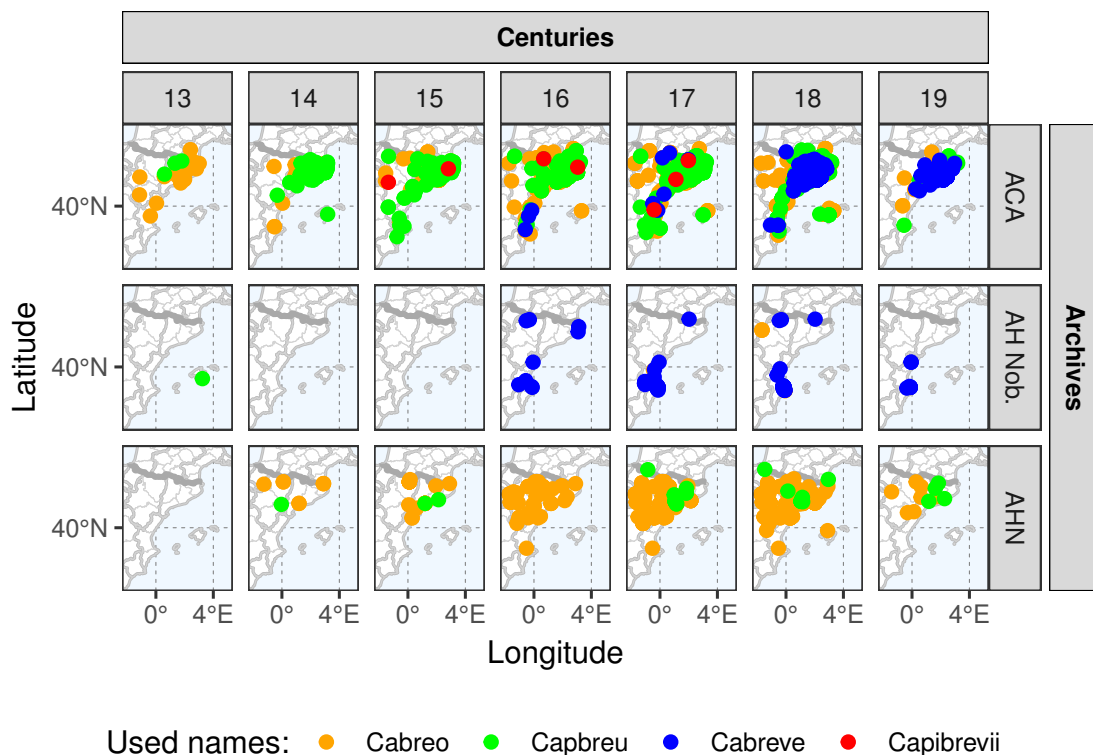


Figure 11: Spatiotemporal distribution of terms used for *cabreves* in the main archives integrated in PARES: Archivo de la Corona de Aragón (ACA), Archivo Histórico de la Nobleza (AHNob) and Archivo Histórico Nacional (AHN).

system (HGIS).

Although we consider that PARES is a very representative source, it is clear that there are other sources that can help to better define the scope or extent covered by the *cabreves*. Therefore, the proposed methodology – or other methodologies that overcome the difficulties we have described – should be applied in other digital archives that we have commented on in the introduction. Among these archives, it seems that the archives of the Generalitat of Catalonia – given the large number of *cabreves* listed – would be a good choice to continue the work presented here.

PARES enables quite specific searches – and even to some extent considers the geographical component – but we believe that an archive of these characteristics would benefit from a geographical index for the *cabreves* and other types of documents. Given the vertical integration of these archives, these characteristics for historical and geographical indexing should appear at all levels so that they can be used in popular platforms (such as Europeana) and be accessed from other tools such as internet search engines.

Acknowledgements

We are grateful to the editor and reviewers for their help in improving this paper.

Funding

This work was supported by grants from the Spanish Ministry of Economy and Competitiveness, project SIOSE-INNOVA (CSO2016-79420-R AEI/FEDER UE).

References

Alex, B., Grover, C., Tobin, R., Oberlander, J., 2019. Geoparsing historical and contemporary literary text set in the City of Edinburgh. *Language Resources and Evaluation* 53, 651–675. URL: <https://doi.org/10.1007/s10579-019-09443-x>, doi:10.1007/s10579-019-09443-x.

- Anderson, B., Eaton, F., Schwartz, S.W., 2015. Archival appraisal and the digital record: Applying past tradition for future practice. *New Review of Information Networking* 20, 3–15. doi:10.1080/13614576.2015.1114823.
- Antonson, H., 2009. The extent of farm desertion in central Sweden during the late medieval agrarian crisis: landscape as a source. *Journal of Historical Geography* 35, 619–641. URL: <http://dx.doi.org/10.1016/j.jhg.2009.01.021>, doi:10.1016/j.jhg.2009.01.021.
- Bainbridge, W.S., 2018. Family History Digital Libraries. volume 24 of *Human–Computer Interaction Series*. Springer International Publishing, Cham. URL: <http://link.springer.com/10.1007/978-3-030-01063-8>, doi:10.1007/978-3-030-01063-8.
- Béaur, G., Congost, R., Luna, P.F., 2018. Emphyteusis: a practical question?, in: *Agrarian Change and Imperfect Property: Emphyteusis in Europe (16th to 19th centuries)*. Brepols Publishers, Turnhout, Belgium. chapter 1, pp. 11–38. URL: <https://www.brepolsonline.net/doi/10.1484/M.RURHE-EB.5.116116>, doi:10.1484/M.RURHE-EB.5.116116.
- Benítez, R., 1984. Formas de transmisión de la propiedad en el País Valenciano (siglos XVII y XVIII). Aproximación metodológica a un nuevo tratamiento de una fuente notarial clásica: los cabreves., in: *La documentación notarial y la historia*. Universidad de Santiago de Compostela, pp. 353–370. URL: <https://dialnet.unirioja.es/servlet/articulo?codigo=4409655>.
- Bezerra, A., Silva, I., Guedes, L.A., Silva, D., Leitão, G., Saito, K., 2019. Extracting value from industrial alarms and events: A data-driven approach based on exploratory data analysis. *Sensors (Switzerland)* 19. doi:10.3390/s19122772.
- Brunson, J.C., 2017. ggalluvial: alluvial diagrams in 'ggplot2'. R package version 0.5. 0.
- Camarero Bullón, C., García Juan, L., 2018. Geografía histórica de los espacios reales: Alóndiga, Aceca y Barciles, despoblados del rey en la vega del Tajo. *Estudios Geográficos* 79, 209. doi:10.3989/estgeogr.201809.
- Carrion, D., Migliaccio, F., Minini, G., Zambrano, C., 2016. From historical documents to GIS: A spatial database for medieval fiscal data in Southern Italy. *Historical Methods* 49, 1–10. URL: <http://dx.doi.org/10.1080/01615440.2015.1023877>, doi:10.1080/01615440.2015.1023877.
- Chae, B.K., 2019. A General framework for studying the evolution of the digital innovation ecosystem: The case of big data. *International Journal of Information Management* 45, 83–94. URL: <https://doi.org/10.1016/j.ijinfomgt.2018.10.023>, doi:10.1016/j.ijinfomgt.2018.10.023.
- Clifford, J., Alex, B., Coates, C.M., Klein, E., Watson, A., 2016. Geoparsing history: Locating commodities in ten million pages of nineteenth-century sources. *Historical Methods* 49, 115–131. URL: <http://dx.doi.org/10.1080/01615440.2015.1116419>, doi:10.1080/01615440.2015.1116419.
- Conde Villaverde, M.L., 2006. La investigación en los archivos. Evolución de su contexto y contenido. *Arbor*, 33–37doi:10.3989/arbor.2006.i1717.5.
- Conedera, M., Vassere, S., Neff, C., Meurer, M., Krebs, P., 2007. Using toponymy to reconstruct past land use: a case study of 'brüsáda' (burn) in southern Switzerland. *Journal of Historical Geography* 33, 729–748. doi:10.1016/j.jhg.2006.11.002.
- Congost, R., 2003. Property Rights and Historical Analysis: What Rights? What History? *Past & Present* 181, 73–106. URL: <https://academic.oup.com/past/article-lookup/doi/10.1093/past/181.1.73>, doi:10.1093/past/181.1.73.
- Congost, R., 2009. Dels capbreus al registre de la propietat. Drets, títols i usos socials de la informació a Catalunya (segles XIV-XX). *Documenta Universitaria*, Girona.
- Congost, R. (Ed.), 2015. Dels capbreus al registre de la propietat. Drets, títols i usos socials de la informació a Catalunya (segles XIV-XX). *Documenta Universitaria*, Girona. URL: <https://books.google.es/books?id=vEcACgAAQBAJ>.
- Corredor Plaja, A.M., 2001. Les possessions de la Universitat de Pals el 1672. *Estudis del Baix Empordà* 20, 73–80. URL: <https://dugi-doc.udg.edu/handle/10256/12164>.
- Couture, C., 2005. Archival appraisal: A status report. *Archivaria* 59, 83–107.
- Dascher, O., 2015. Data Bases and Statistical Systems: Archives and Historical Databases. volume 1. Second edi ed., Elsevier. URL: <http://dx.doi.org/10.1016/B978-0-08-097086-8.62131-1>, doi:10.1016/B978-0-08-097086-8.62131-1.
- Frajer, J., Fiedor, D., 2021. A historical curiosity or a source of accurate spatial information on historical land use? the issue of accuracy of old cadastres in the example of josephian cadastre from the habsburg empire. *Land Use Policy* 100, 104937. URL: <http://www.sciencedirect.com/science/article/pii/S0264837719322331>, doi:https://doi.org/10.1016/j.landusepol.2020.104937.
- García Juan, L., Álvarez Miguel, A.J., Camarero Bullón, C., Escalona Monge, J., 2012. Generación de una metodología para la gestión y recreación cartográfica a partir de información del Catastro de Ensenada. *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 268–302URL: <http://www.geofocus.org/index.php/geofocus/article/view/246>.
- García Trobat, P., 2001. A Forgotten Result of the Spanish War of Succession: the Cadastre and its Fiscal Effects on the Crown of Aragón, in: *BADEN-BADEN, N.V. (Ed.), Cadastre and Modern State in Italy, Spain and France (18th c.)*, pp. 193–216.
- Gatta, G., Ariotti, E., Bitelli, G., 2017. Geomatics science applied to cartographic heritage and archive sources: A new way to explore the XIXth century Gregorian Cadastre of Bologna (Italy), an ante-litteram 3D GIS. *Journal of Cultural Heritage* 23, 68–76. URL: <http://dx.doi.org/10.1016/j.culher.2016.06.009>, doi:10.1016/j.culher.2016.06.009.
- Gil Olcina, A., 1998. Control del dominio útil y salvaguarda del directo en la enfiteusis señorial valenciana: Los Cabreves. *Estudios Geográficos* 59, 201. URL: <http://estudiosgeograficos.revistas.csic.es/index.php/estudiosgeograficos/article/view/601>, doi:10.3989/egogr.1998.i231.601.
- Gil Olcina, A., 2012. Singularidades del Régimen Señorial Valenciano. Expansión, declive y extinción de la Señoría Directa. *Publicaciones Universidad de Alicante*, Alicante.
- Gil Olcina, A., Canales Martínez, G., 2007. Residuos de propiedad señorial en España. Perduración y ocaso en el Bajo Segura. *Universidad de Alicante*, Alicante. URL: <http://hdl.handle.net/10045/13583>.
- Gómez Benedito, V., 2018. Las reformas administrativas de la Casa de Medinaceli en sus dominios valencianos durante la segunda mitad del siglo XVIII. *Hispania* 77, 763. URL: <http://hispania.revistas.csic.es/index.php/hispania/article/view/531>, doi:10.3989/hispania.2017.021.
- Gregory, I., Donaldson, C., Murrieta-Flores, P., Rayson, P., 2015. Geoparsing, GIS, and Textual Analysis: Current Developments in Spatial

- Humanities Research. *International Journal of Humanities and Arts Computing* 9, 1–14. URL: <http://www.eupublishing.com/doi/10.3366/ijhac.2015.0135>, doi:10.3366/ijhac.2015.0135.
- Gutiérrez, A., Domènech, A., Zaragoza, B., Miravet, D., 2020. Profiling tourists' use of public transport through smart travel card data. *Journal of Transport Geography* 88, 13. URL: <https://doi.org/10.1016/j.jtrangeo.2020.102820><https://linkinghub.elsevier.com/retrieve/pii/S0966692320302283>, doi:10.1016/j.jtrangeo.2020.102820.
- Harvey, F., Kaim, D., Gajda, A., 2014. Analysis of historical change using cadastral materials in the carpathian foothills. *European Journal of Geography* 5, 6–21.
- Hodder, J., 2017. On absence and abundance: biography as method in archival research. *Area* 49, 452–459. doi:10.1111/area.12329.
- Kelly, E.J., 2018. Use of Louisiana's Digital Cultural Heritage by Wikipedians. *Journal of Web Librarianship* 12, 85–106. URL: <https://doi.org/10.1080/19322909.2017.1391733>, doi:10.1080/19322909.2017.1391733.
- Khalil, S., Fakir, M., 2017. RCrawler: An R package for parallel web crawling and scraping. *SoftwareX* 6, 98–106. URL: <http://dx.doi.org/10.1016/j.softx.2017.04.004>, doi:10.1016/j.softx.2017.04.004.
- Mostern, R., 2017. World-Historical Gazetteer Research Report. *Journal of World-Historical Information* 4, 2016–2018. doi:10.5195/jwhi.2017.43.
- Nothman, J., Qin, H., Yurchak, R., 2018. Stop word lists in free open-source software packages, in: *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, Association for Computational Linguistics, Melbourne, Australia. pp. 7–12. URL: <https://www.aclweb.org/anthology/W18-2502>, doi:10.18653/v1/W18-2502.
- Pacios, A.R., La Torre Merino, J.L., 2018. Spanish historic archives' use of websites as a management transparency vehicle. *Archival Science* 18, 185–205. URL: <https://doi.org/10.1007/s10502-018-9291-x>, doi:10.1007/s10502-018-9291-x.
- Paterson, L.L., Gregory, I.N., 2019. Representations of Poverty and Place. *Representations of Poverty and Place*, 41–60. doi:10.1007/978-3-319-93503-4.
- Peris Albentosa, T., 1995. La Evolución De La Agricultura Valenciana Entre Los Siglos XV Y XIX: Rasgos Cualitativos Y Problemas De Cuantificación. *Revista De Historia Económica / Journal of Iberian and Latin American Economic History* 13, 473–508. doi:10.1017/S0212610900005280.
- Petras, V., Stiller, J., 2017. A decade of evaluating europeana - constructs, contexts, methods & criteria, in: Kamps, J., Tsakonas, G., Manolopoulos, Y., Iliadis, L., Karydis, I. (Eds.), *Research and Advanced Technology for Digital Libraries*, Springer International Publishing, Cham. pp. 233–245.
- Puig Costa, J., 2013. El castell i la presó de Siurana d'Empordà. *Annals de l'Institut d'Estudis Empordanesos* 44, 575–601. doi:10.2436/20.8010.01.145.
- R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.r-project.org/>.
- Rosvall, M., Bergstrom, C.T., 2010. Mapping change in large networks. *PLoS ONE* 5. doi:10.1371/journal.pone.0008694, arXiv:0812.1242.
- Sánchez Mairena, A., 2014. Acercando Los Archivos De Archivos Españoles (Pares) del Ministerio de Educación, Cultura y deporte. *Arbor Ciencia, Pensamiento y Cultura* 190.
- Sinn, D., 2012. Impact of digital archival collections on historical research. *Journal of the American Society for Information Science and Technology* 63, 1521–1537. URL: <http://onlinelibrary.wiley.com/doi/10.1002/asi.22883><http://doi.wiley.com/10.1002/asi.22650>, doi:10.1002/asi.22650.
- Szabó, P., Suchánková, S., Křížová, L., Kotačka, M., Kvardová, M., Macek, M., Müllerová, J., Brázdil, R., 2018. More than trees: The challenges of creating a geodatabase to capture the complexity of forest history. *Historical Methods* 51, 175–189. doi:10.1080/01615440.2018.1444523.
- Thelwall, M., Stuart, D., 2006. Web crawling ethics revisited: Cost, privacy, and denial of service. *Journal of the American Society for Information Science and Technology* 57, 1771–1779. URL: <http://doi.wiley.com/10.1002/asi.20388>, doi:10.1002/asi.20388, arXiv:0803.1716.
- Tobiáš, P., Cajthaml, J., Krejčí, J., 2018. Rapid reconstruction of historical urban landscape: The surroundings of Czech chateaux and castles. *Journal of Cultural Heritage* 30, 1–9. doi:10.1016/j.culher.2017.09.020.
- Torres Gros, J., 1996. El capbreu de la vila de Bellpuig del 1696. *Quaderns de El Pregoner d'Urgell*, 37–50.
- Turner, S., Bolòs, J., Kinnaird, T., 2017. Changes and continuities in a Mediterranean landscape: a new interdisciplinary approach to understanding historic character in western Catalonia. *Landscape Research* 6397, 1–17. URL: <http://doi.org/10.1080/01426397.2017.1386778>, doi:10.1080/01426397.2017.1386778.
- Yeung, A.W.K., 2018. Data visualization by alluvial diagrams for bibliometric reports, systematic reviews and meta-analyses. *Current Science* 115, 1942–1947. doi:10.18520/cs/v115/i10/1942-1947.
- Zaragoza, B., Giménez-Font, P., Belda-Antolí, A., Ramón-Morte, A., 2019. A graph-based analysis for generating geographical context from a historical cadastre in Spain (17th and 18th centuries). *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, 228–243. URL: <https://doi.org/10.1080/01615440.2019.1590269><https://www.tandfonline.com/doi/full/10.1080/01615440.2019.1590269>, doi:10.1080/01615440.2019.1590269.
- Zaragoza, B.M., Trilles, S., Navarro-Carrión, J.T., 2020. Leveraging container technologies in a giscience project: A perspective from open reproducible research. *ISPRS International Journal of Geo-Information* 9, 138.
- Zeng, M.L., 2017. Smart Data for Digital Humanities. *Journal of Data and Information Science* 2, 1–12. doi:10.1515/jdis-2017-0001.