

Article

Mapping Imprecision: How to Geocode Data from Inaccurate Historic Maps

Tomasz Panecki 

Tadeusz Manteuffel Institute of History of the Polish Academy of Sciences, Rynek Starego Miasta 29/31, 00-272 Warsaw, Poland; tpanecki@ihpan.edu.pl

Abstract: This paper aims to present and discuss the method of geocoding historical place names from historic maps that cannot be georeferenced in the GIS environment. This concerns especially maps drawn in the early modern period, i.e., before the common use of precise topographic surveys. Such maps are valuable sources of place names and geocoding them is an asset to historical and geographical analyses. Geocoding is a process of matching spatial data (such as place names) with reference datasets (databases, gazetteers) and therefore giving them geographic coordinates. Such referencing can be done using multiple tools (online, desktop), reference datasets (modern, historical) and methods (manual, semi-automatic, automatic), but no suitable approach to handling inaccurate historic maps has yet been proposed. In this paper, selected geocoding strategies were described, as well as the author's method of matching place names from inaccurate cartographic sources. The study was based on Charles Perthées maps of Polish palatinates (1:225,000, 1783–1804)—maps that are not mathematically precise enough to be georeferenced. The proposed semi-automatic and curated approach results in 85% accuracy. It reflects the manual workflow of historical geographers who identify place names with their modern counterparts by analysing their location and proper name.

Keywords: historic maps; toponyms; place names; historical GIS; early modern period; geocoding; georeferencing; Charles Perthées; Poland



Citation: Panecki, T. Mapping Imprecision: How to Geocode Data from Inaccurate Historic Maps. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 149. <https://doi.org/10.3390/ijgi12040149>

Academic Editors: Wolfgang Kainz and Florian Hruby

Received: 3 February 2023

Revised: 29 March 2023

Accepted: 1 April 2023

Published: 2 April 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Historic (old) maps play an important role in historical geography, environmental history, historical GIS (Geographic Information Systems) and beyond. Such maps are a valuable source of spatial and descriptive data that can be used for geographic analyses which involve a look into the past. In recent years, due to the development of digital and spatial humanities, we may witness a rapid evolution of framing historical scholarship with computer methods and tools [1–4]. Place names are one of the particularly important elements in geohumanities. They play a crucial role in the so-called ‘deep maps’ that “engage evidence within its spatiotemporal context and to provide a platform for spatially embedded arguments”. Deep mapping forms a perfect example of taking the advantage of combining humanities and information sciences [5,6]. By analysing historical toponyms, we might also, for example, rediscover lost places [7], learn how settlements developed [8], what was the geographic knowledge of past societies [9] or how politics influenced the process of naming space [10]. To conduct research which involves historical place names analysis, it is necessary to geocode them, i.e., supplement them with geographical coordinates. Since there is still a limited number of sources which are available in ready-to-use spatial databases, geocoding of place names derived from analogue sources remains a challenge [11].

Databases containing geocoded historical places (place names, toponyms) can usually be developed based on two kinds of sources: written records and maps. The approach in geocoding data fundamentally differs between these two. Based on a written source, a list of places can be made and then geocoded [12]. In certain cases, each place name derived from such a source can be supplemented with information that specifies its geographical

extent, such as administrative affiliation. However, information about a historical district or country might not be helpful for geocoding, especially automatized, as historical administrative units often differ significantly from contemporary ones. A historic map can also be a source of place names. If a map is mathematically precise enough to be georeferenced, its content (i.e., place names) can be simply vectorised [1]. Therefore, there is no further need for geocoding: geographical coordinates are already in the database. In this case, we would rather speak of reconciling or identifying place names from historical maps with reference datasets.

There is also a third category of sources which falls between written and cartographic at least in the case of approaches in geocoding: maps which cannot be georeferenced due to their geometric imprecision. Such maps were usually drawn in the early modern period (approximately from the 16th century to the end of the 18th century) when survey topographic methods and tools did not yet provide sufficient results and were not widely used. Such maps are hard or even impossible to be georeferenced as errors in geometry (measured by analysing distances between points on the map and in reality) may rise to dozens of kilometres [13] (Figure 1). This does not mean that the only possible approach to geocoding is manual. Even if they are imprecise, such maps do carry spatial information, especially in terms of relations between particular objects represented by cartographic signs. These relations might be used for semi-automatic geocoding. In this paper, an approach utilising this feature of historic maps as well as the similarity of place names between a map and reference datasets is proposed and evaluated. The approach results from developing a digital edition of the ‘special maps of palatinates of Poland’ drawn by Charles Perthees (1:225,000, 1783–1804) <https://perthees.ihpan.edu.pl/> (accessed on 5 January 2023). During the works, nearly 37,000 place names were analysed in terms of the possibility of their geocoding.

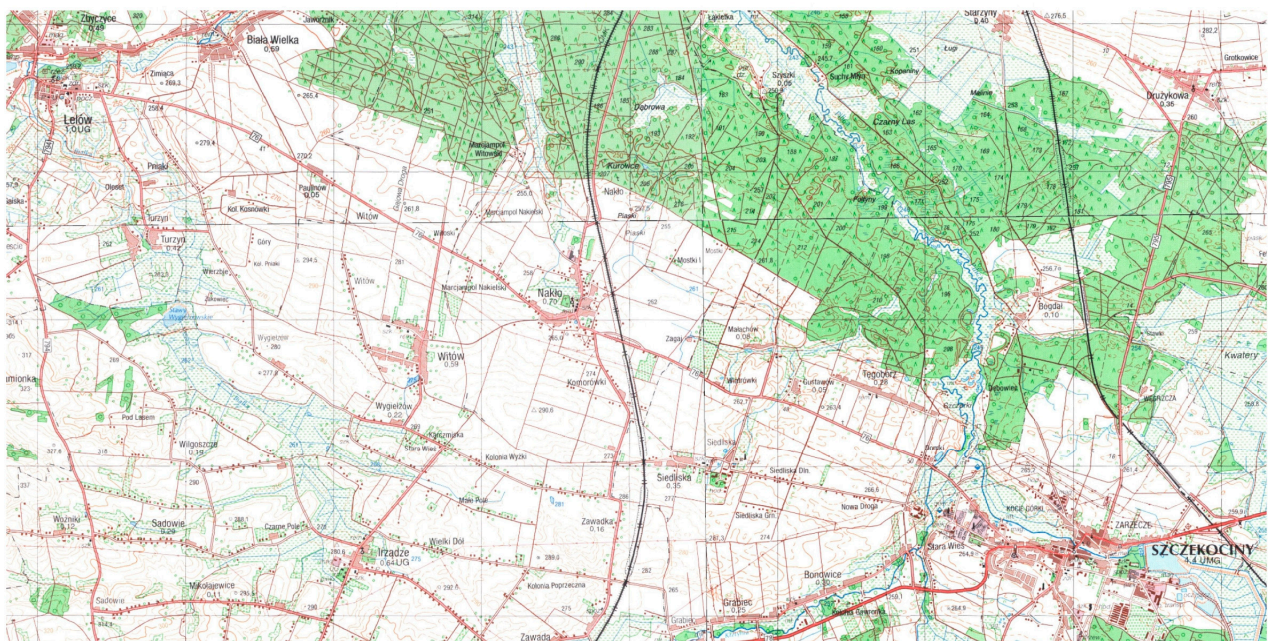
The main idea behind geocoding of place names depicted on ‘special maps’ was to georeference the vector data (place names) instead of the raster image of the map, and process (geocode) them in the subsequent stages. The geocoding was performed in a semi-automated way. A list of potential matches between the historic map and reference datasets was created based on spatial relations of georeferenced place names, and then each of the matches was assessed in terms of textual (string) similarity. Eventually, matches were checked and verified manually. Such a method, as described below, allows for accurate and fast geocoding of data derived from inaccurate historic maps.

The concept of geocoding, i.e., converting descriptive geographical data, such as addresses, buildings, streets and localities, into a set of coordinates in a given datum is fundamental for spatial analysis in many disciplines and domains, not excluding historical research. Geocoding plays an important role from the perspective of Geographic Information Systems (GIS) as it enables the conversion of large volumes of data containing spatial information to machine-understood spatial datasets [14].

Geocoding can be described by its four major components: the input, output, matching algorithm and reference dataset. Input data is a list of place names to be geographically mapped. Usually, it was associated with addresses or streets, but other geographical features can be geocoded as well. The output contains the above-mentioned features with geometry (usually points), but it could also be an external identifier from a reference dataset that allows for locating. Choosing a particular matching algorithm is important from the perspective of the process of finding potential counterparts between the input data and reference dataset(s): How names are compared? Can geographical extent be limited? Which dataset should be chosen for possible matches? Finally, there are reference datasets: databases, gazetteers or spatial data collections carrying particular information on names (or also variants thereof), features classes and geographical coordinates [14]. Since different software can be used for geocoding, either online or standalone, it would be worth adding another component to the list: tools. Geocoding produces output with particular level of accuracy, which is mostly determined by the reference dataset instead of the matching algorithm [15]. In general, the accuracy of geocoding varies from approx. 50% to 100% depending on the area, data to be geocoded, reference datasets and algorithms [16–20].



Map of Cracow palatinate and Siewierz Duchy drawn by Charles Perthées (1:225,000, 1792), scaled up to approx. 1:80 000



Topographic map from <https://mapy.geoportal.gov.pl/> (1:50,000, 1995), scaled down to approx. 1:80 000

Figure 1. A perfect example of inaccurate maps is the ‘special maps of palatinates of Poland’ drawn by Charles Perthées (**upper**) juxtaposed with a modern map of the same area (**lower**). Topographic features on inaccurate maps roughly correspond with their modern counterparts. Such maps cannot be georeferenced due to large mathematical errors which would cause enormous distortions. Therefore, a special approach towards them has to be adapted in the context of data processing.

2. Tools, Datasets and Methods to Geocode Historical Place Names

Geocoding of historical place names takes advantage of tools, reference datasets and methods used for geocoding in general but boasts of some specific characteristics. They come mostly from the nature of historical data combined with digital processing. Szady underlines the difference between history (humanities) and computer science (sciences), asking how these “(. . .) two ‘worlds’ could cooperate and support one another to better

and more fully comprehend both our contemporary reality and that of the past” [21]. Data derived from historical sources might be vague, uncertain or incomplete, but this does not preclude them from being processed digitally [22,23]. Scholars must pay special attention while referring historical place names to their modern counterparts using specific tools, reference datasets and methods as outlined below.

2.1. Tools

There are three general approaches for geocoding historical place names: manual, semi-automatic and automatic, and they can be performed using either desktop or online tools [20]. Manual geocoding is usually achieved through vectorisation (in the case of georeferenced maps) or matching place names one by one to a reference dataset. A semi-automatic solution involves (1) manual input of an individual address or place and then the automatic response from software with location or (2) manual input followed by a suggestion provided by the tool, e.g., in the form of values in a drop-down menu. Automatic geocoding tools require input data to be prepared in a certain way and format to be processed (usually online) and geocoded. The results might be one-to-one relationships; therefore, there is only one potential match between the input place name and referenced place name or one-to-many where the end-user might decide which match is the most appropriate. The latter solution allows for quality control but is more “semi” than automatic.

Entirely manual geocoding can be performed using most desktop GIS software. The process depends on the source of place names. In the case of a map, it involves georeferencing it and vectorising toponyms in a (usually) point geometry. The result is a geocoded layer but without identifiers from reference datasets. When a written source is a basis for geocoding, each of the place names should be manually input as an element of a GIS point layer or identified with a corresponding place name from a dataset which already has coordinates. The former approach results in a geocoded layer of place names while the latter requires converting coordinates into geometry, but the resulting dataset is already equipped with external identifiers [1].

An example of a semi-automatic solution, especially for historical place names, is “Recogito” <https://recogito.pelagios.org/> (accessed on 5 January 2023). It is an online application designed for the semantic annotation of images, usually representing historical manuscripts, prints or maps. The main feature is the possibility of identifying and marking named entities, including places, people and events. Data from annotated images can be exported in spreadsheets and further processed in external software [24]. Geocoding in “Recogito” is done by selecting the fragment of an image manually, usually containing a proper name, then entering the name into the system, as well as selecting the type of the entity (person, place, event) and eventually choosing a corresponding place name from a list. The list is generated automatically after the name of the entity is given by the user. The matching algorithm compares the name with an array of online reference datasets. They comprise databases and gazetteers storing modern or historical data, e.g., HistoGIS, Kima Historical Gazetteer, Pleiades and GeoNames. If one is looking for a locality that is not part of any of these datasets, there is no possibility of a manual addition. It can be done only after downloading the data, in postprocessing.

Automatic geocoding solutions for toponyms are also available. One of the recently (2021) developed “Historical Geocoding Assistant” <https://dissinet.cz/apps/hga/> (accessed on 5 January 2023). It is an open-source web-based software developed explicitly for spatial historical data geocoding [17]. The tool requires data to be stored in the Google Sheets table, and the proper name is the only necessary field. Then, the algorithm processes the data against five databases (Geonames, Wikipedia, Getty TGN, Pleiades, and China Historical Gazetteer) and automatically returns the most probable match by analysing the proper names in a given geographic context. The results are not only map coordinates but also certainty levels for each of the input place names. Moreover, the user does not need to rely only on the reference dataset as there is a possibility of manually searching and adding a place to be referenced with.

World Historical Gazetteer <https://whgazetteer.org/> (accessed on 5 January 2023) is a web platform providing services to researchers who are interested in places and the geographic dimensions of their research subjects. Primarily, it is an aggregator of various gazetteers, databases and collections and provides links between reference datasets, such as Wikidata and Getty TGN, and others which were uploaded to the system and asserted to be a “close match”. Therefore, it is not a unified gazetteer per se with its data, nor a geocoding tool. Nonetheless, it allows for the reconciliation of place names from the contributor’s dataset against the “World Historical Gazetteer index” (approximately 60,000 records). For a dataset to be reconciled, it needs to be uploaded to the system compliant with Linked Places format (LPF), which adds temporal extensions to the GeoJSON standard [25,26]. In practice, it can be a spreadsheet based on the ready-to-use template. The user can upload data with place names without adding any additional information but places’ proper names. The reconciliation process will be possible, yet much more accurate results can be achieved with a dataset already supplemented with coordinates.

2.2. Reference Datasets

These tools, to properly link historical place names to reference entities, need a set of well-established reference datasets. Such a dataset needs to represent place names (contemporary, historical or both) along with their attributes (proper name, variant names, location, type and optionally external identifiers) and of course geographical coordinates. A good example of such a dataset would be a gazetteer, either with modern or historical data. Gazetteers are in general lists of places with basic information about them. The extent of the information has been developed into a “Gazetteer Content Standard” that specifies three necessary elements: feature name (including variants), classification (type of the place name) and location (usually given in the WGS-84 coordinate system) [27,28].

Currently, there are dozens of datasets (gazetteers, databases, collections) available which can serve as a reference for identifying historical place names. They can be divided into general datasets with modern places, and datasets developed specifically for historical research. In the first group, one can use the OSM-Nominatim dataset and GeoNames <https://www.geonames.org/> (accessed on 5 January 2023). The latter is a geographical database containing over 27,000,000 geographical names and consists of over 12,000,000 unique features (4,800,000 populated places) and 15,000,000 alternate names [29]. It is a part of the “Recogito” web application. Wikidata and Getty Thesaurus of Geographic Names (TGN) are reference datasets that can be used as such in “World Historical Gazetteer” and “Historical Geocoding Assistant”, both comprising more than 3,500,000 records each. Among the second group, there are “Pleiades” (for places in ancient Europe) [30], “China Historical Gazetteer” (for China) [31], and “Historical Atlas of Poland” (for Poland in the early modern period) <https://atlasfontium.pl/language/en/polish-lands-of-the-crown/> (accessed on 5 January 2023).

There are of course many more datasets with place names that can be used by scholars for referencing historical data. Choosing an appropriate one should be based both on their objective features and usefulness for particular studies. Usefulness is mostly driven by the level of detail and geographical scope of the dataset. Using both modern and historical gazetteers is also a good idea. Therefore, in many projects, various datasets were used such as “OS AddressBase Premium” or “GB1900 Gazetteer” for the United Kingdom or the gazetteer of place names published by the French National Library for France for geocoding places from that country [16,18].

2.3. Methods

Tools for geocoding and reconciling historical data to run properly need not only datasets but also specific methods for linking input (historical place names) and reference data. Scholars underline that this stage, i.e., matching entities, is the most important part of geocoding [14]. In general, during geocoding a text identifying a particular geographical feature is processed by the algorithm. Input is normalised, standardised and compared

with the array of potential counterparts in the reference dataset, and then the output is produced in the form of ranked matches for input data or a single result. The best scenario is when there is one match per entity; however, this does not always mean a positive match (so-called “false positive”). Other scenarios include zero matches or multiple matches. In both cases some more identifying data is necessary. If there is no such opportunity, then the entities remain unmatched or have to be referenced manually [14].

Regardless of the technical foundations of any matching algorithm, it is important to indicate what features of place names play a major role in their identification between datasets. To perform a link between two spatial entities, especially modern and historical, three relations are taken under consideration: proper name, location and type or in other words—place name, footprint and place type [32]. These are the most important criteria which identify particular localities and differentiate one from another over time and space in diachronic and synchronic contexts. Comparing names of places is the first step of identification and thus geocoding. Nonetheless, slightly different proper names in the input and reference datasets do not necessarily mean different entities. Names might have variants, different spellings or even be written in the form of exonyms and, regardless of differences, still point at the same place name (such as “Vienna” and “Wien”). There can be, of course, cases where completely different proper names denote the same spatiotemporal entity, e.g., “Byzantium”, “Constantinople” and “Istanbul”. Secondly, two different localities might hold the same proper name then the location disambiguates them. It is therefore important for the algorithm, and, in general, for the geocoding, to pay special attention to geographical scope, and match only these places which are at a given distance one from another. The spatial extent of matching for a particular place name from input data should be as large to find a possible match, and as small as not to find a locality with the same name in different locations. Thirdly, the type of place names is also crucial in disambiguation. The historical place name, especially human settlement, might have several possible modern counterparts sharing the same proper name and holding similar locations. Moreover, what was then a separate settlement, could now be a part of another one, e.g., historical village becomes a district of a city. To conclude, the three above-mentioned variables of historical places (place name, footprint and place type) significantly influence their geocoding, especially when they change simultaneously. A locality might change its name, type and location and still maintain its identity over time [32].

To sum up, there are two prerequisites for an approach to be used in geocoding place names from imprecise maps (and perhaps for geocoding historical data at all). The first would be to make use of high-resolution and tailored reference datasets: official, state database with contemporary place names (or streets, addresses, buildings) supplemented, if possible, with a database containing historical data. The second would encompass an elastic algorithm of comparing proper names between possible matches (e.g., considering strings’ similarity degree), within the specified geographical extent.

3. Materials and Methods

3.1. Materials

As underlined in the introduction, the experimental part of the paper is based on the analysis of so-called ‘special maps’ of palatinates of Poland by Charles Perthées (1:225,000, 1783–1804), which are considered to be a highly valuable source of information for historians and geographers analysing the space of the Polish-Lithuanian Commonwealth at its dawn [33–35]. Maps are highly detailed, yet rather inaccurate in terms of mathematical precision, which is caused by the lack of topographical surveying or triangulation. The concept of detailed maps of palatinates was presented to King Stanisław August Poniatowski in 1779 by the king’s geographer—Charles (Karol) Perthées (1739–1815). The geographer’s goal was to prepare maps for each of Poland’s palatinates and most probably use them as a base to draw a general map of the whole state. Perthées’s method of work, unlike Austrian or Prussian cartographers of that period, did not take advantage of triangulation and field surveys [13]. Instead, it involved relying almost entirely on written topographic

descriptions of parishes made by the parsons of the Catholic Church. A similar method of gathering geographic information was used before in France, but there is no evidence that Perthées knew it. As a consequence, maps elaborated that way were quite reliable in terms of features' characteristics, but the maps' geometric precision was low. Scholars indicate that average errors vary from 5 to 10 km but could reach up to 20–25 km in the worst cases [34,36,37]. The first map drawn by Perthées covered the palatinate of Mazovia (1783). Other palatinates were mapped from 1784 to 1804 and covered the entire Polish lands of the Crown except for the palatinates of Gniezno and Sieradz. Out of 12 maps, 5 were printed in the Paris Tardieu printing house, 7 remained in manuscripts, 2 of which remain until today in the form of poor photographic copies that had been made before the Second World War (Figure 2, Table 1).

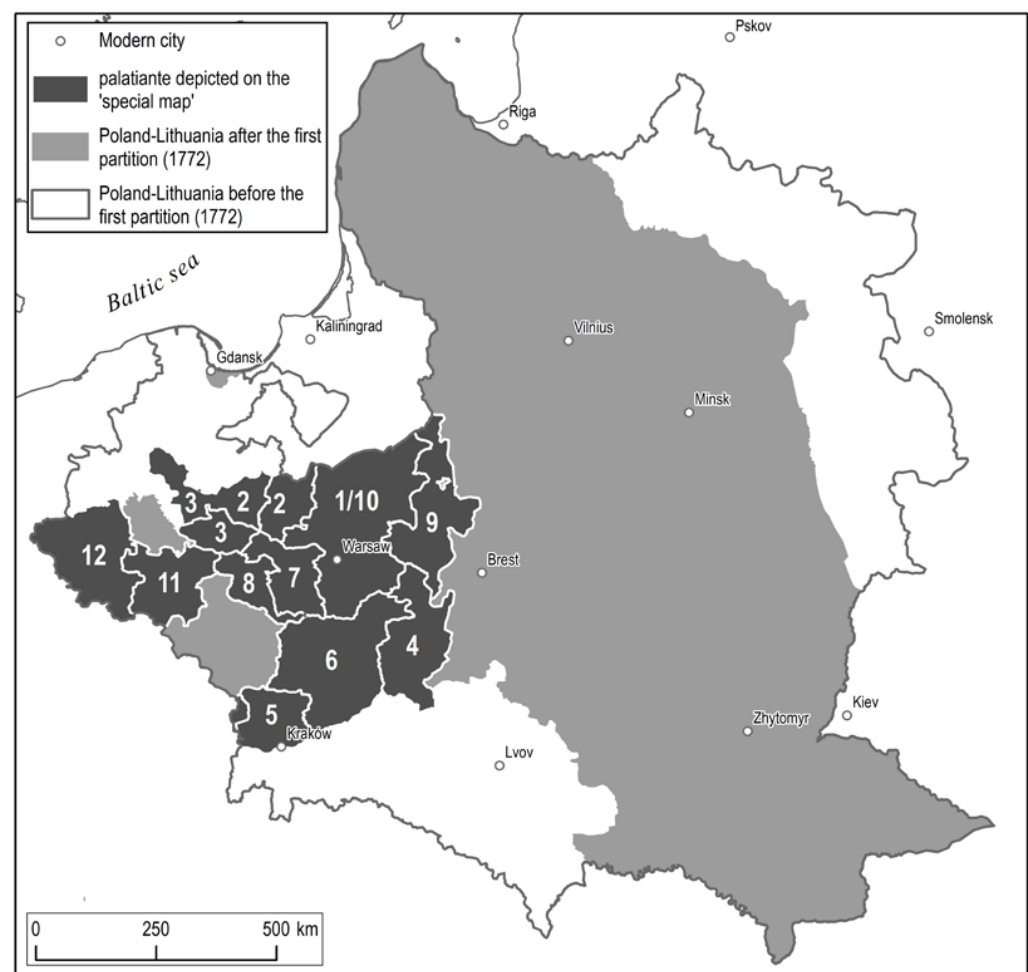


Figure 2. A map representing the geographical extent of 'special maps'. Numbers placed on palatinates correspond with numbers in Table 1.

As mentioned before, maps drawn by Perthées are characterised by a high level of detail. They include settlements, industrial facilities, roads, rivers, lakes and afforestation. Maps contain a multitude of toponyms related to dwellings and natural landscapes allowing for place name analysis. On the other hand, the scientific potential of 'special maps' is limited by their geometric precision, significantly worse than the Austrian or Prussian maps of these times. Therefore, taking full advantage of the maps' content is a challenge and requires a specific approach.

Table 1. A list of special maps with the dates of their elaboration and material form. Numbers correspond with numbers on the map in Figure 2.

Number	Palatinate Name	Date of Elaboration	From
1	Mazovia (first version)	1783	manuscript (photocopy)
2	Płock and Dobrzyń Land	1784	copper plate
3	Brześć and Inowrocław	1785	manuscript
4	Lublin	1786	copper plate
5	Cracow and Duchy of Siewierz	1787	manuscript; copper plate
6	Sandomierz	1788–1791	manuscript; copper plate
7	Rawa	1792	copper plate
8	Łeczyca	1793	manuscript
9	Podlasie	1795	manuscript
10	Mazovia (second version)	1789–1791	manuscript
11	Kalisz	after 1798	manuscript
12	Poznań	1804	manuscript (photocopy)

Such an approach is being developed in the project “Cartography at the service of political reforms in the times of Stanisław August Poniatowski—a critical elaboration of ‘Geographical-statistical description of the parishes in the Kingdom of Poland’ and the maps of the palatinates by Charles Perthées” that is currently conducted at the Institute of History, Polish Academy of Sciences in Warsaw <https://perthees.ihpan.edu.pl/> (accessed on 5 January 2023). The main focus of the project lies in the online publication of Charles Perthées’s works, including the ‘special maps’ in the form of a Scholarly Digital Edition [38]. It will contain a WebGIS application with two modules: one allowing users to explore maps supplemented by annotated layers with place names. Basic functionality will be searching, filtering, and downloading the data. Since ‘special maps’ cannot be georeferenced due to a lack of geometric precision, another module of the application will integrate modern geographical data with geocoded (georeferenced) place names from the ‘special maps’. To achieve this goal, data from maps are being collected and analysed in the so-called “Source-driven Data Model”. Its main feature is linking the advantages of direct source data representation with scalability and flexibility features, which provide analytical possibilities [39].

In the case of maps, cartographic signs and map annotations are gathered in separate layers and eventually linked together by overlapping polygons. It is done so due to highly ambiguous map drawing: very often it is hard to connect annotation with a sign or group of signs. Using polygons to overlap the most probable connections between them allows for flexibility in data modelling (Figure 3). Another feature of the adapted model is using a source geometry of maps instead of georeferencing them. In other words, the content of the maps is indexed in the GIS environment, but maps are treated as non-geographical images with only a local coordinate system. This approach follows the methodology of historical manuscripts indexing using GIS tools [40]. This raised, however, an issue of eventual geocoding collected information. For nearly 37,000 annotations collected from 11 ‘special maps’, a manual approach would be too time-consuming, so a method facilitating this process had to be developed. It is based on spatial relations between annotations collected from ‘special maps’ and corresponding entities from reference datasets. It considers the string similarity for a particular place name. The proposed method allows for semi-automatic place names geocoding from inaccurate historical maps, such as ‘special maps’ of Polish palatinates from the end of the 18th c.

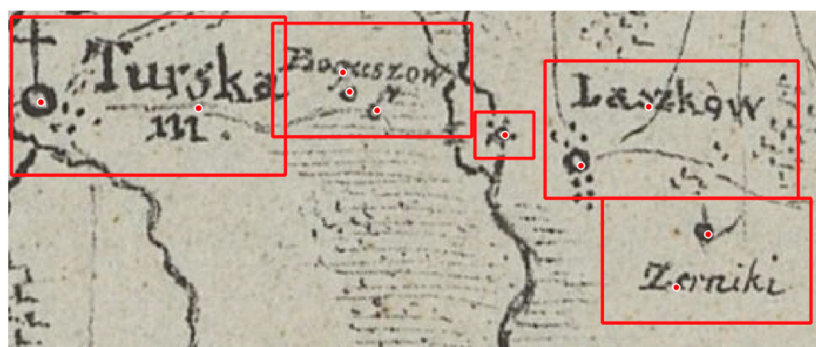


Figure 3. Fragment of the ‘special map’ of Kalisz palatinate depicting the complexity of the drawing. Going from the left, there is “Turska m.” which is a parochial village (proper name and sign are overlapped by the red polygon), then there is “Boguszow” village with an inn (proper name and two signs are overlapped), a separate mill and two other villages. The polygons are features in the database as allow for flexible linking between sometimes ambiguously represented annotations and signs.

3.2. Methods

The proposed method of semi-automatic place names geocoding from the ‘special maps’ maps involved nine major steps (Figure 4):

1. Exporting the historical place names collected from ‘special maps’ in separate layers, grouped by the maps sheet (one map sheet represents one palatinate);
2. Georeferencing historical place names to modern geographical data;
3. Spatially joining to each of the historical place names all place names from a given reference dataset within a 10 km radius;
4. Calculating the degree of similarity of proper names (0–1) for spatially joined place names (for historical and reference place names);
5. Creating a ranking list of most probable matches ordered by the similarity degree;
6. Limiting the results for further investigation by filtering matches with less than 0.5 of similarity degree;
7. Automatically appending ‘certain match’ to these historical place names where there is only one match with a similarity degree equal to 1;
8. Exporting the list of historical place names with their potential matches from reference datasets back to the source geometry of the ‘special maps’;
9. Manually iterating over the historical place names and assessing the certainty of matches with their counterparts from reference datasets.

The starting point for place names’ semi-automatic geocoding was to have them collected from a historic map in the form of a GIS layer in point geometry, where one point represented one place name with its name, type and coordinates. The geometry of place names was non-geographical XY because ‘special maps’ in the project were treated as images with only a local coordinate system. Names were collected in the form of transliterations, i.e., taken directly from the map with no corrections or additions, and then transcribed into a normalised form. Examples of transliteration are “Zb (...) adz” and “Zaluzie Szlach” and their transcription “Zbądz” and “Załuzie Szlacheckie”, respectively. In the former case, the name is illegible from the map, and in the latter, it is cropped. The type of the place name at this stage of the project was either ‘dwelling’ (e.g., town, village or hamlet) or ‘natural landscape element’ (e.g., river or lake). Coordinates as underlined before were non-geographical and represented the location of a place name on a scanned map sheet.

The first step of the geocoding was exporting the historical place names collected from ‘special maps’. All of the place names, originally, were stored in one database table, but to be georeferenced, place names from each of the map sheets (i.e., a map of palatinate) had to be exported as separate layers. The result of this stage was 11 separate layers with place names collected from 11 maps of palatinates.

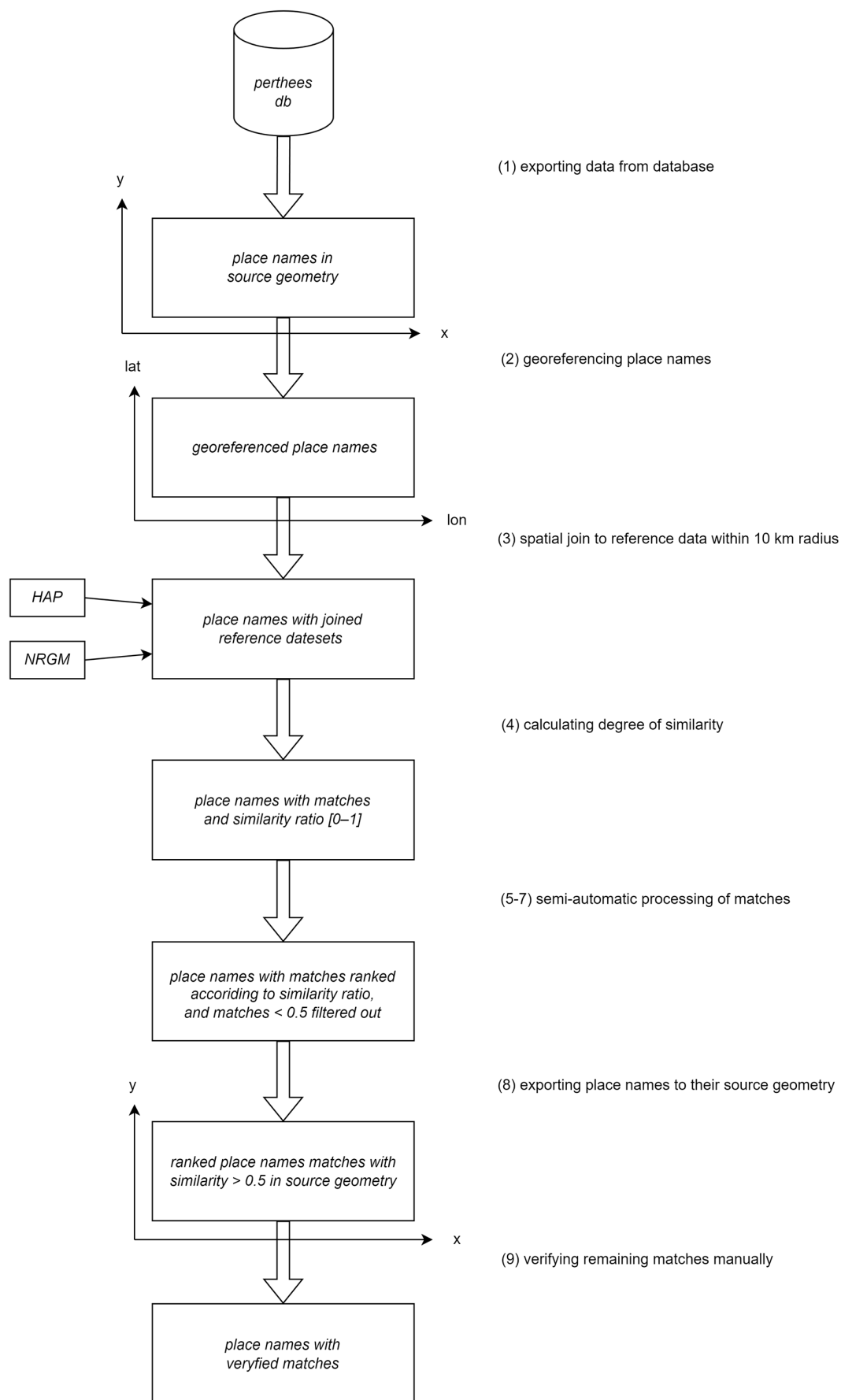


Figure 4. The outline of the proposed approach of geocoding place names from inaccurate historic maps.

Vector layers prepared in such a way were then georeferenced to modern geographical data. The expected geometric error of georeferencing was not higher than 10 km. Such a value arises from maximum geometric errors calculated for ‘special maps’ in previous works [36]. The idea behind this stage was to limit the geographical scope of the search algorithm—in other words, to set an approximate location of a place name, and then find its potential counterparts (matches) from reference datasets. The result of this step was 11 layers georeferenced (spatially aligned) to modern geodata.

The next stage involved finding potential matches for particular place names in reference datasets in a given radius. Georeferenced layers were merged into one, and then two reference datasets to be used for geolocating place names from ‘special maps’ were chosen: modern and historical to ensure the furthest possible data coverage. The “National Register of Geographical Names” (NRGM) [41] was chosen as a dataset with modern data and the “Historical Atlas of Poland (HAP) for historical <https://atlasfontium.pl/language/en/polish-lands-of-the-crown/> (accessed on 5 January 2023). The former is a GIS gazetteer covering Poland with data on localities and natural environment features, and the latter is a spatial database of historical localities in Poland for the end of the 16th century. Georeferenced layers of place names from ‘special maps’ were then joined spatially to both of the reference datasets in a 10 km radius. Such value approximates the potential error of ‘special maps’ geometric precision [34,36,37]. The result of this stage was a layer with potential matches between place names from ‘special maps’ and reference datasets based on spatial relations. Each of the historical place names was given its potential counterpart from either NRGM or HAP within a 10 km radius.

The following stage assumed refining the matching process by calculating the degree of names’ (strings) similarity of spatially joined place names. In other words, the goal was to find the best potential match for each of the historical place names by considering its name similarity to their counterparts in reference datasets. To achieve this, a Python library was used: the “SequenceMatcher” class from the “difflib” module. The idea of the “SequenceMatcher” based on the Ratcliff/Obershelp algorithm is to find the longest contiguous matching subsequence that contains no “junk” elements [42]. The “.ratio()” function measures the sequences’ similarity as a float in the range [0, 1]. The examples of comparing ‘special maps’ and NRGM/HAP place names were collected in Table 2, and they show how the similarity of strings differs for particular ratio levels. In general, we can say that, if the ratio is higher than approx. 0.8 the names are most likely to be the same, yet written in a different manner, and if the ratio is lower than approx. 0.5 names are different and therefore there is no semantic match between them. Therefore, after some manual tests, a similarity ratio of 0.5 was set as a threshold for further analysis. The result of this step was a layer with potential matches between place names from ‘special maps’ and reference datasets based on spatial relations and the similarity of names.

In the next step, all of the place names in the layer were ranked according to the similarity degree (the most probable matches were on the top) and grouped by place names from ‘special maps’ for further investigation. To facilitate manual checks of the matches, the dataset was then reduced. The matches with a similarity lower than 0.5 were filtered out as it was unlikely that a match with such a low similarity would prove to be correct. The next stage was rather technical and involved re-exporting the layer back to its original (source) geometry so that the place names with their potential matches could be seen as an overlay of ‘special maps’.

Such an automatic names’ similarity detection facilitated further steps which required manual (curated) verification of the algorithm. In the last stage of the process, a manual iteration of the matches was performed. Every single place name from ‘special maps’ along with its potential match with NRGM/HAP was evaluated. The newly created field “Identification” in the layer was manually filled with values reflecting the certainty of identification: if the identification was certain, the value of “certain” was given, and if the match was probable, yet uncertain the value was “uncertain” (Table 2). This stage was performed by experts in historical settlement studies. To facilitate the work not only

least possible matches were filtered out, but also the most probable ones were highlighted. Where there was only one match with a similarity degree equal to 1, a value of “certain” was then given automatically to the “Identification” field. The remaining matches had to be verified manually. The verification was based on analysing the layer with matches (overlaid on the ‘special maps’) and juxtaposing it with visualised spatial data from NRGH/HAP (see Figure 1 as an example of the same area shown on a ‘special map’ and a modern map). For each of the place names, if there was a potential counterpart from one of the reference datasets, the value in the “Identification” field was set to “certain” or “uncertain”. The criteria used for verification were the similarity of names (based on the computed degree of similarity) and similarity of location (based on visual analysis of the ‘special maps’ and modern geodata). These criteria were not arbitrarily set to particular values of similarity of names and locations due to the nature of historical data and required an individual approach. The result of this stage was a layer of historical place names along with their counterparts from reference datasets and assumed certainty of matches. Filtering this layer by the “Identification” field (“certain”, “uncertain”, “no identification”) results in a most probable list of matches, and therefore enables placing historical place names from ‘special maps’ on the geographical map through geocoding.

Table 2. Fragment of the table containing place names from ‘special maps’ and matched place names from NRGH/HAP.

Transcription	External Id	External Name	Similarity	External Resource Type	Identification
Wolika	Wolika_bus_wsl_snd	Wolika	0.83	hap	certain
Wolika	150489	Wolika	0.83	nrgh	certain
Wolika	180948	Wola	0.8	nrgh	no identification
Wolika	56124	Kolonia	0.61	nrgh	no identification
Wolika	73504	Łowiska	0.61	nrgh	no identification
Wolika	56133	Kolonia	0.61	nrgh	no identification
Wolika	149176	Włodarka	0.57	nrgh	no identification
Wolika	180924	Ulica	0.54	nrgh	no identification
Wolika	180926	Ulica	0.54	nrgh	no identification
Wolika	180943	Niwka	0.54	nrgh	no identification
Wolika	Kobylniki_gor_wsl_snd	Kobylniki	0.53	hap	no identification

4. Results

The method outlined above was adopted to geocode nearly 37,000 place names collected in the project and significantly facilitated the work of identifying features from ‘special maps’ against reference datasets, as otherwise, it would have to be done entirely manually. To exemplify the outcomes of this study, 100 place names were randomly selected as case studies. Within the sample, there are place names from 11 maps of palatinates and of different categories: towns, villages, hamlets, and industrial facilities.

Place names were geocoded using the method described above. In the first stages, they were exported from the project’s database to be georeferenced and then merged into one layer. Subsequently, a spatial join between the sample and NRGH and HAP was performed resulting in 19,226 matches: 14,739 for NRGH and 4527 for HAP. The difference stems from the level of granularity of the datasets: in the whole NRGH dataset there are 383,583 features, and 23,771 in HAP. Therefore, there was an average of 192 matches per single place name (147 for NRGH and 45 for HAP). Afterwards, a similarity degree between matches was calculated using the Ratcliff/Obershelp algorithm. Place names in the layer were ranked according to the similarity degree and grouped by place names for further analysis. In the next stage, the layer was limited by filtering matches with a similarity degree lower than 0.5. This resulted in cutting almost 97% of matches that were unlikely to be probable, and only 551 matches remained: 376 for NRGH and 175 for HAP.

There were also 76 cases (44 for NRGm and 32 for HAP) when there was only one match per a single place name with a similarity degree equal to 1. It means that 44% and 32% of historical place names were identified entirely automatically with reference datasets (NRGM and HAP respectively). The remaining cases had to be manually verified. Finally, out of 100 place names, only 15 were not identified with either NRGm or HAP. It means that most likely they do not have counterparts either in modern (2023) and historical (approx. 1600) datasets and their geocoding needs to be done entirely manually with the help of other sources. Out of 85 identified place names, two were identified with HAP only, 14 with NRGm only, and 69 with both datasets. Nine place names were identified with a lower level of certainty (see: supplementary material) It was because a place name from 'special maps' represented, what is in a reference dataset, a part of the locality, e.g., "Grochów" in Perthées map and "Grochów Szlachecki" in NRGm or "Kęsocha" in the 'special map' and "Tańsk-Kęsochy" in HAP/NRGm. Mereological (part-whole) changes occurring over time influence the level of certainty and often make matching impossible [32]. Overall, the above-presented method allowed for the curated, semi-automatic identification of 85% of historical place names using two datasets (modern and historical).

5. Discussion

How is the proposed approach of place names geocoding from inaccurate historic maps related to previously developed tools, reference datasets and methods? It uses desktop GIS applications for the whole process of geocoding: QGIS and ArcMap. The former is used as PostgreSQL/PostGIS database environment is used within the project, and QGIS handles these databases well [43]. The latter, on the other hand, has a "spatial adjustment" tool allowing for vector georeferencing that worked in the case of analysed data. A Python "SequenceMatcher" library is used to calculate the similarity degree between place names. Unfortunately, none of the described semi-automatic and automatic tools works well in the case of imprecise historic maps. MMQGIS <https://plugins.qgis.org/plugins/mmqgis/> (accessed on 5 January 2023) [20] might work well for contemporary data, especially addresses, but when it comes to historical localities, it fails. This is because place names cannot be geocoded without any spatial context: administrative affiliation or limited geographical extent. MMQGIS supports only "country" and "state", which have little in common with historical units. There can be (and is!) many distinct localities sharing the same proper name within one country or state. This was the same reason why "World Historical Gazetteer" and its reconciling tool were not used after some preliminary tests which resulted in a little number of geocoded place names. Reconciling works fine when a place name already has coordinates. "Recogito" with its semi-automatic approach would be a better solution; however, it does not allow the introduction of more tailored reference datasets than defaults. In this study a national Polish digital gazetteer (NRGM) and a historical database (HAP) were used, both presenting a high level of detail, a focus on Poland and two temporal snapshots (1600 and 2022). Through them it is possible to link place names from 'special maps' to other datasets as well: "Wikidata" has links to NRGm, and HAP is indexed in "World Historical Gazetteer". That makes the resulting dataset compliant with the Linked Open Data approach and would allow for SPARQL queries in the future [25].

Methods used for geocoding considered two features of historical place names: their approximate geographical location obtained through data georeferencing and proper names in the form of transcription (i.e., with corrections). Therefore, it was possible to perform matching of place names within a given geographical extent, and the matching itself considered the names' similarity. Names did not have to be identical to be matched. The "uncertain" matches (Table 3) occurred mostly because of the development of settlements over time: a settlement can split into two parts or vice versa, and two or more can merge into one [32]. Such cases also impact geocoding of historical place names and often have to be manually verified. The proposed method involved georeferencing data (place names) instead of the image (raster map). It was done so as such inaccurate maps, such as the

‘special maps’, are not accurate enough to be warped against modern coordinates. This would result in huge distortions and such maps could become illegible and lose a lot of their aesthetic and semantic richness. The proposed methods allow for processing the data without distorting the map. Given the map without georeferencing, we are, therefore, unable to capture linear and surface elements such as rivers or forests.

Table 3. Place names from ‘special maps’ with corresponding ones from NRGm/HAP are identified with a lower level of certainty, mostly because historical places do not have exact modern counterparts. If a place name occurs twice, it means it was identified with both datasets.

Transcription	External Id	External Name	Similarity	External Resource Type	Identification
Bukowiec	13334	Bukowiec Wielki	0.7	nrgm	uncertain
Grochów	38704	Grochów Szlachecki	0.56	nrgm	uncertain
Kęsocha	Tansk_Kesochy_prz_maz	Tańsk-Kęsochy	0.6	hap	uncertain
Kęsocha	138755	Tańsk-Kęsocha	0.7	nrgm	uncertain
Kuźnica Kaniów	50086	Osiedle Kaniów	0.57	nrgm	uncertain
Lepki	88932	Nowe Lepki	0.53	nrgm	uncertain
Majdan	75537	Majdan Królewski	0.55	nrgm	uncertain
Marcisze	Zebry_Marcisze_prz_maz	Żebry-Marcisze	0.73	hap	uncertain
Marcisze	92651	Olszewo-Marcisze	0.67	nrgm	uncertain
Mąchocice	78688	Mąchocice Górne	0.75	nrgm	uncertain
Żabokliki Wielkie	163087	Żaboklik	0.64	nrgm	uncertain

6. Conclusions

Place names are parts of multitude kinds of historical records: written (court books, tax registers and censuses) and cartographic (maps, plans and atlases). Usually, they are associated with settlements or localities, but they can be extended to any names that appear in the sources, such as names related to environmental features or various human-made facilities. Analysing their geographical characteristics is possible only when they are extracted from the source, identified with reference datasets and geocoded. The proposed method reflects the manual workflow of historical geographers who identify place names encountered in historical sources with their modern counterparts by considering their location and proper name [44]. They do so to map historical phenomena and take advantage of GIS and cartographic analysis of historical sources contributing to the rapidly developing field of digital and spatial humanities. The spatial and temporal distribution of place names is the main focus of ‘historical GIS’, ‘spatial history’ and deep mapping’ [1,4–6]. Enhancement of their semi-automatic geocoding techniques, such as the one presented in the paper, leads to the possibility of creating more georeferenced datasets containing historical geodata, and therefore facilitating research in the field. One of the datasets that will be created on the basis of the presented method will contain approximately 37,000 places names collected from the so-called ‘special maps’ drawn by Charles Perthées at the end of the 18th century. This set of maps is important for Polish cultural heritage and allows for geohistorical studies of Poland before it ceased to exist as a sovereign state in 1795.

There are some clear outcomes and “lessons learned” that can be posed. The proposed method used for gathering and geocoding cartographic data can be applied towards other similar historic maps: ones that are too inaccurate to be georeferenced, but accurate enough to show spatial relations. Virtually any map, apart from figurative representation such as medieval OT maps, could be processed this way, regardless of its scale and precision. Of course, in the case of maps at larger/smaller scales, the approach has to be tailored (search radius, similarity degree, etc.). The proposed approach is valid only for point geometries, but most of the place names depicted on maps can be abstracted into points. Therefore, political or administrative units, and natural landscape elements, such as river names, can be processed as well.

The drawback of the approach is it is time-consuming in the first, preprocessing stage. Georeferencing data, performing spatial joins and assessing similarity takes time.

However, the essential stage of identifying places only by assessing the proposed matches is quite fast. It means that the approach should rather be addressed for larger projects. It took significantly less time to process 37,000 records in the proposed, semi-automatic way, instead of analyzing place names one by one.

The proposed solution can be treated as a starting point for the future as many aspects can be improved or developed. For instance, instead of manually transliterating place names, an OCR (Optical Character Recognition) can be used, such as “Transkribus” [45]. Consequently, a fully developed system could run the matching algorithm on-the-fly and propose possible matches based on a dynamically assigned buffer radius. Future contributions to the field might also involve testing different string-matching algorithms. As this study involved the Ratcliff/Obershelp algorithm which searches for the longest common substring between two strings, there are also others that can be used, e.g., Levenshtein, Damerau–Levenshtein, Jaro or Jaro–Winkler [42].

Supplementary Materials: Spreadsheet in *.csv format with the sample of 100 place names from ‘special maps’ and possible matches from NRGH/HAP is available here: <https://doi.org/10.5281/zenodo.7594902> (accessed on 5 February 2023).

Funding: This research was funded by National Programme for the Development of the Humanities, grant number 11H 18 0122 87.

Data Availability Statement: Spreadsheet in *.csv format with the sample of 100 place names from ‘special maps’ and possible matches from NRGH/HAP is available here: <https://doi.org/10.5281/zenodo.7594902> (accessed on 5 February 2023). Whole dataset is currently not publicly available.

Acknowledgments: I would like to thank Bogumił Szady (PI of the project) and Tomasz Królik, who indexed half of the 37,000 toponyms.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gregory, I.; Ell, P. *Historical GIS: Technologies, Methodologies, and Scholarship*; Cambridge University Press: Cambridge, UK, 2007.
2. Knowles, A. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*; ESRI: Redlands, CA, USA, 2008.
3. Roberts, L.; Thevenin, T.; Hallam, J.; Beveridge, A.; Mostern, R.; Southall, H.; Cunningham, N.A.; Schwartz, R.M.; Meeks, E. *Toward Spatial Humanities: Historical GIS and Spatial History*; Geddes, A., Gregory, I.N., Eds.; Indiana University Press: Bloomington, IN, USA, 2014; ISBN 9781306546218.
4. Roberts, L.; Thevenin, T.; Hallam, J.; Beveridge, A.; Mostern, R.; Southall, H.; Cunningham, N.A.; Schwartz, R.M.; Meeks, E. *The Routledge Companion to Spatial History*, 1st ed.; Gregory, I., DeBats, D., Lafreniere, D., Eds.; Taylor and Francis: Washington, DC, USA, 2018; ISBN 9781351584135.
5. Aitken, S.; Cooper, D.; DeLozier, G.; Ethington, P.; Gregory, I.N.; Hardie, A.; Yuan, M. *Deep Maps and Spatial Narratives*; Bodenhamer, D.J., Corrigan, J., Harris, T.M., Eds.; Indiana University Press: Bloomington, IN, USA, 2015; ISBN 9780253015600.
6. Aitken, S.; Cooper, D.; DeLozier, G.; Ethington, P.; Gregory, I.N.; Hardie, A.; Martin, W.M.; McIntosh, J.; Rayson, P.; Toyosawa, N. *Making Deep Maps: Foundations, Approaches, and Methods: Foundations, Approaches, and Methods*; Bodenhamer, D.J., Corrigan, J., Harris, T.M., Eds.; Routledge: Abingdon, UK; New York, NY, USA, 2022; ISBN 9781000453287.
7. Dostál, I.; Havlíček, M.; Svoboda, J. There Used to Be a River Ferry: Identifying and Analyzing Localities by Means of Old Topographic Maps. *Water* **2021**, *13*, 2689. [\[CrossRef\]](#)
8. Zou, H.; Chen, C.; Xiao, W.; Shi, L. Spatial-Temporal Evolution Relationship between Water Systems and Historical Settlement Sites Based on Quantitative Analysis: A Case Study of Hankou in Wuhan, China (1635–1949). *Sustainability* **2022**, *14*, 14614. [\[CrossRef\]](#)
9. Driver, F. Geography’s Empire: Histories of Geographical Knowledge. *Environ. Plan. D* **1992**, *10*, 23–40. [\[CrossRef\]](#)
10. Rusu, M.S. Shifting urban namescapes: Street name politics and toponymic change in a Romanian(ised) city. *J. Hist. Geogr.* **2019**, *65*, 48–58. [\[CrossRef\]](#)
11. Yuan, M. Spatializing text for deep mapping. In *Making Deep Maps: Foundations, Approaches, and Methods: Foundations, Approaches, and Methods*; Bodenhamer, D.J., Corrigan, J., Harris, T.M., Eds.; Routledge: Abingdon, UK; New York, NY, USA, 2022; pp. 50–64, ISBN 9781000453287.
12. Schneider, P.; Jones, J.; Hiltmann, T.; Kaupinnen, T. Challenge-derived design practices for a semantic gazetteer for medieval and early modern places. *Semant. Web* **2021**, *12*, 493–515. [\[CrossRef\]](#)
13. Edney, M.H.; Pedley, M.S. (Eds.) *Cartography in the European Enlightenment*; University of Chicago Press: Chicago, IL, USA, 2020; ISBN 9780226184753.

14. Goldberg, D.W.; Wilson, J.P.; Knoblock, C.A. From Text to Geographic Coordinates: The Current State of Geocoding. *Urisa J.* **2007**, *19*, 33.
15. Karimi, H.A.; Durcik, M.; Rasdorf, W. Evaluation of Uncertainties Associated with Geocoding Techniques. *Comput. Aided Civ. Infrastruct. Eng.* **2004**, *19*, 170–185. [\[CrossRef\]](#)
16. Lan, T.; Longley, P. Geo-Referencing and Mapping 1901 Census Addresses for England and Wales. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 320. [\[CrossRef\]](#)
17. Mertel, A.; Zbiral, D.; Stachoň, Z.; Hořínková, H. Historical geocoding assistant. *SoftwareX* **2021**, *14*, 100682. [\[CrossRef\]](#)
18. Cura, R.; Dumenieu, B.; Abadie, N.; Costes, B.; Perret, J.; Gribaudi, M. Historical Collaborative Geocoding. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 262. [\[CrossRef\]](#)
19. Roongpiboonsopit, D.; Karimi, H.A. Comparative evaluation and analysis of online geocoding services. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 1081–1100. [\[CrossRef\]](#)
20. Cetl, V.; Kliment, T.; Jogun, T. A comparison of address geocoding techniques—Case study of the city of Zagreb, Croatia. *Surv. Rev.* **2018**, *50*, 97–106. [\[CrossRef\]](#)
21. Szady, B. Spatio-temporal databases as research tool in historical geography. *Geogr. Pol.* **2016**, *89*, 359–370. [\[CrossRef\]](#)
22. Myrda, G.; Szady, B.; Ławrynowicz, A. Modeling and presenting incomplete and uncertain data on historical settlement units. *Trans. GIS* **2020**, *24*, 355–370. [\[CrossRef\]](#)
23. Grossner, K.E. *Representing Historical Knowledge in Geographic Information Systems*; University of California: Santa Barbara, CA, USA, 2010.
24. Simon, R.; Barker, E.; Isaksen, L.; de Soto Cañamares, P. Linking Early Geospatial Documents, One Place at a Time: Annotation of Geographic Documents with Recogito. *e-Perimetretron* **2015**, *10*, 49–59.
25. Grossner, K.E.; Janowicz, K.; Kessler, C. Place, Period, and Setting for Linked Data Gazetteers. In *Placing Names: Enriching and Integrating Gazetteers*; Berman, M.L., Mostern, R., Southall, H., Eds.; Indiana University Press: Bloomington, IN, USA, 2016; pp. 80–96, ISBN 9780253022561.
26. Grossner, K.; Mostern, R. Linked Places in World Historical Gazetteer. In Proceedings of the 5th ACM SIGSPATIAL International Workshop on Geospatial Humanities. SIGSPATIAL 21: 29th International Conference on Advances in Geographic Information Systems, Beijing, China, 2–5 November 2021; Moncla, L., Brando, C., McDonough, K., Eds.; ACM: New York, NY, USA, 2021; pp. 40–43, ISBN 9781450391023.
27. Wilson, J.P.; Lam, C.S.; Holmes-Wong, D.A. A New Method for the Specification of Geographic Footprints in Digital Gazetteers. *Cartogr. Geogr. Inf. Sci.* **2004**, *31*, 195–207. [\[CrossRef\]](#)
28. Berman, M.; Mostern, R.; Southall, H. Introduction. In *Placing Names: Enriching and Integrating Gazetteers*; Berman, M.L., Mostern, R., Southall, H., Eds.; Indiana University Press: Bloomington, IN, USA, 2016; 265p, ISBN 9780253022561.
29. Ahlers, D. Assessment of the accuracy of GeoNames gazetteer data. In Proceedings of the 7th Workshop on Geographic Information Retrieval. SIGSPATIAL'13: 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, FL, USA, 5 November 2013; Jones, C., Purves, R., Eds.; ACM: New York, NY, USA; pp. 74–81, ISBN 9781450322416.
30. Barker, E.; Simon, R.; Isaksen, L.; de Soto, P. The Pleiades Gazetteer and the Pelagios Project. In *Placing Names: Enriching and Integrating Gazetteers*; Berman, M.L., Mostern, R., Southall, H., Eds.; Indiana University Press: Bloomington, IN, USA, 2016; pp. 97–109, ISBN 9780253022561.
31. Berman, M.; Åhlfeldt, J.; Wick, M. Historical Gazetteer System Integration: CHGIS, Regnum Francorum, and Geonames. In *Placing Names: Enriching and Integrating Gazetteers*; Berman, M.L., Mostern, R., Southall, H., Eds.; Indiana University Press: Bloomington, IN, USA, 2016; pp. 110–128, ISBN 9780253022561.
32. Garbacz, P.; Ławrynowicz, A.; Szady, B. Identity Criteria for Localities. In *Formal Ontology in Information Systems: Proceedings of the 10th International Conference (FOIS 2018)*; Borgo, S., Hitzler, P., Kutz, O., Eds.; IOS Press: Amsterdam, Berlin, Germany; Washington, DC, USA, 2018; pp. 47–54, ISBN 9781614999102.
33. Buczek, K. *The History of Polish Cartography from the 15th to the 18th Century*; Zakład Narodowy im. Ossolinski: Wrocław, Poland, 1966.
34. Buczek, K. Kartograf króla Stanisława Augusta. Życie i dzieła. In *Karol Perthées (1739–1815). Kartograf Pierwszej Rzeczypospolitej I Entomolog*; Pawłowski, J., Ed.; Wydawnictwo Retro-Art: Warszawa, Poland, 2003; pp. 21–134.
35. Rutkowski, H. Mapy Perthéesa. In *Fundamenta Historiae*; Słoiński, M., Zbieranowski, M., Eds.; Instytut Historii PAN: Warszawa, Poland, 2016; pp. 269–282.
36. Szady, B. Mapa województwa lubelskiego Karola Perthéesa z 1786 roku jako źródło kartograficzne i historyczne. In *Biblioteka Polskiego Przeglądu Kartograficznego*; Konopska, B., Ostrowski, J., Paślowski, J., Wespiański, P.E., Eds.; Polskie Towarzystwo Geograficzne: Warszawa, Poland, 2012; pp. 26–35.
37. Pietkiewicz, S. Analyse de l'exactitude de quelques cartes du XVIIe, XVIIIe et XIXe siècle couvrant les territoires de l'ancienne Pologne. *Pol. Przegląd Kartogr.* **1960**, *32*, 21–27.
38. Sahle, P. What is a Scholarly Digital Edition? In *Digital Scholarly Editing: Theories and Practices*; Driscoll, M.J., Pierazzo, E., Eds.; Open Book Publishers: Cambridge, UK, 2016; pp. 19–41, ISBN 978-1-78374-238-7.
39. Szady, B.; Panecki, T. Source-driven data model for geohistorical records' editing: A case study of the works of Karol Perthées. *Misc. Geogr.* **2022**, *26*, 52–62. [\[CrossRef\]](#)
40. Borek, A.; Związek, T.; Słomski, M.; Gochna, M.; Myrda, G.; Słoiński, M. Technical and methodological foundations of digital indexing of medieval and early modern court books. *Digit. Scholarsh. Humanit.* **2019**, *35*, 233–253. [\[CrossRef\]](#)

41. Dz.U. 2021 poz. 309; Rozporządzenie Ministra Administracji i Cyfryzacji w Sprawie Państwowego Rejestru Nazw Geograficznych. PRNG: Warszawa, Poland, 2012. Available online: <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20120000309/O/D20120309.pdf> (accessed on 5 January 2023).
42. Kalbaliyev, E.; Rustamov, S. Text Similarity Detection Using Machine Learning Algorithms with Character-Based Similarity Measures. In *Digital Interaction and Machine Intelligence*; Biele, C., Kacprzyk, J., Owsinski, J.W., Romanowski, A., Sikorski, M., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 11–19, ISBN 978-3-030-74727-5.
43. Hsu, L.S.; Obe, R.O. *PostGIS in Action, 3ed.*; Manning Publications: Shelter Island, NY, USA, 2021; ISBN 9781617296697.
44. Borek, A.; Słomski, M. Greater Poland. In *Polish Lands of the Crown in the Second Half of the Sixteenth Century*; Słomski, M., Słomska-Przech, K., Eds.; Institute of History PAS: Warszawa, Poland, 2021; pp. 649–675. Available online: <https://rcin.org.pl/ihpan/publication/269261> (accessed on 5 January 2023).
45. Kahle, P.; Colutto, S.; Hackl, G.; Muhlberger, G. Transkribus—A Service Platform for Transcription, Recognition and Retrieval of Historical Documents. In *Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, Kyoto, Japan, 9–15 November 2017; pp. 19–24, ISBN 978-1-5386-3586-5.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.