Source-driven data model for geohistorical records’ editing: a case study of the works of Karol Perthéès

Abstract
The aim of the paper is to propose a data model for editing historical records using the example of Karol Perthéès’ works from the late 18th century. These consist of cartographic sketches, elaborated on the basis of parochial questionnaires and the maps of palatinates of the Crown, which are at a scale of 1:225,000. The model links the advantages of direct source data representation with scalability and flexibility features, which provide analytical possibilities. The sketches are indexed using the INDXR application and the structure of the model reflects the structure of the source. The maps were analyzed in a GIS environment, but no georeferencing was performed. The data collected separately as symbols and annotations were treated independently and were linked via features. The model is illustrated by a preliminary study of materials covering the area of the Zemborzyce parish near Lublin.

Keywords
Karol Perthéès • historical sources’ editing • maps • sketches • INDXR • Zemborzyce

Introduction
Maps occupy a special place in the scholarly digital editing of historical sources. Without delving into a discussion on the changes brought about by the use of the “digital” in the editing process of sources, it is necessary to underline the benefits achieved by this type of elaboration (Slotr 2015, Bem 2016, Jurek 2016, Slotr 2017). The most important feature of digital editing has greater implications than just the possibility of digitally visualizing historical sources: it should involve documenting and disseminating the entire research process as a key functionality. Digital editions also allow for new interpretations, additions, and verifications (Vogeler 2019).

The digital editing of cartographic sources is strongly coupled with comparative methods in historical research, in two variations: spatial and chronological. Maps often present the same space at different scales and levels of generalization (Panecki 2020). The content of each map is conditioned by many factors: the purpose of its development, technical capabilities, cartographic knowledge, the conceptualization of space, etc. (Harley 2002, Edney 2019). For this reason, the digital editing of a cartographic source should involve the simplest possible transformation (from the image to the data) of the source image. A data model built on the basis of the source itself, without imposing modern conceptual structures and spatial categories, has the chance to capture the above-mentioned factors. This “source-driven” model can then be processed into an analytical model during the subsequent stages of the editing process. The results can be compared with other representations of the same area in written and cartographic sources.

Sources
Karol Perthéès’ materials from the late 18th century – maps of palatinates (Perthéès 1783–1804), and the parish sketches (Perthéès 1790–1796) used for their elaboration – can be used as a case study for presenting the idea behind a source-driven data model. Perthéès was a leading cartographer under King Stanislaw August Poniatowski, and was author of many maps of the Commonwealth of Poland, among which, the so-called 1:225,000 “particular maps” of the palatinates are of greatest importance. Over 21 years he managed to elaborate 12 maps for 11 palatinates of the Crown from after the First Partition (Tab. 1).

The maps of the palatinates formed the first of the Crown’s map series to cover such a large area at a large scale, with full settlement network, administrative borders, and natural environment. These maps were drawn with almost no field measurements or triangulation, but, instead were created on the basis of textual descriptions prepared by parsons from 1778 to 1785. The work on the maps had three main stages. In the first, parish surveys prepared according to a given questionnaire were elaborated. Then, on the basis of these, draft sketches were drawn. Finally, based on this material and supplemented by other maps and written sources, the “particular maps” were prepared (Perthéès 1783–1804, Buczak 1966, Buczak 2003, Wemrowa 2003, Rutkowski 2014).

1 The Crown of the Kingdom of Poland and the Polish Crown: the names for the Late Middle Ages and Early Modern territory that covered two provinces – Greater Poland and Lesser Poland. The Polish Crown, along with the Grand Duchy of Lithuania, formed part of the Polish-Lithuanian Commonwealth from 1569 to 1795.
These sketches and maps have not yet been published. The only attempt to do so was made by W. Wernerowa. Her intention was to present not only the rewritten text surveys but also their photocopies; but due to their poor legibility, she decided to redraw them on a similar scale and rewrite the text elements of the sketches (Fig. 1) (Wernerowa 1994, 1996). This was the only possibility to edit and publish this source under the limited technical conditions of the Polish scientific reality of the 1990s.

Parish sketches

Preparing a source-driven data model requires a precise and complete description of the source content. The sketches were an intermediate product between the parish surveys and the maps, and their structure reflected the structure of the parish surveys. The concept of gathering geographical data in this way was formulated by Michał Jerzy Poniatowski, bishop of Plock, and later, Polish primate, who was the brother of King Stanisław August Poniatowski. The structure of the survey was probably prepared by Franciszek Czajkowski, who proposed a model survey for Tarchomin parish, where he was parson (Olszewicz 1921, Szady 2013).

The parish surveys consisted of nine points, and the parish church was at the center of each survey. The survey concerned the administrative affiliation of the parish church, all the settlements in its area, the location and distance to neighboring churches, information on industrial facilities within the parish, road networks, rivers, lakes, and other elements of topography within the parish. The location of all the investigated elements was to be determined according to the eight compass point directions (Rozporządzenia 1785, Buczek 2003).

A preliminary comparative analysis of parish surveys and sketches shows that the sketches were a kind of abstract of the surveys prepared for the development of maps. The Crown’s parish sketches are currently located in the manuscript section of the Vernadsky National Library of Ukraine and are entitled Geographical and Statistical Description of Parishes of the Kingdom of Poland (ref. no. I 5975). They form a collection of 12 volumes, in which more than 2,000 parishes are described, organized by diocese, archdeaconry, and deanery. Each volume begins with a list of archdeaconries and deaneries, giving the total number of parishes in each unit. Each deanery begins with a list of the parishes belonging to it, and a general sketch showing the parish churches, the boundaries of the deanery, and main roads. Each parish, with a few exceptions, is presented on a single page whose structure corresponds to the above-mentioned points of the parish surveys.

Each of the surveys has a header and footer in which the name of the parish and its administrative affiliation are given: state (palatinate, district) and ecclesiastical (deanery). In the central part of the page there is a draft map of the parish. A common element of each sketch is the road and settlement network. Other elements, such as borders, rivers, lakes, and afforestation, appear less regularly. In the bottom-left corner of each page, there is a list of the names of the parish settlements, with the number of houses and their distance and direction from the parish church. Affiliation to both royal and church properties is also recorded. In the lower-right
corner, there are the names of neighboring parishes and major settlements, also with their distances and directions relative to the described parish. In the part below and to the left of the sketch, there is a description of industrial facilities. Other elements, like land cover and relief descriptions, are included quite sporadically. In general, the scope of the content and its completeness may differ quite significantly between parishes (Fig. 2).

Maps
The idea of the maps of the palatinates was presented to the king in 1779. Perthées’ goal was to prepare these particular maps and use them as the main sources for a general map of the whole state and thus give a more accurate and detailed representation than the Zannoni map (1772, 1:690,000). As indicated above, Perthées’ method of work, unlike Austrian and Prussian cartographers of that period, did not encompass triangulation and field surveys but involved using parish surveys as topographic descriptions and the sketches elaborated on their basis. As a consequence, the maps are quite reliable in terms of attributes, but their geometric precision is low and errors could be by as much as 20–25 kilometers (Szady 2012). The first map drawn by Perthées covered the Mazowsze Palatinate (1783), as

Figure 2. Sketch of Zemborzyce parish (near Lublin)
Source: Perthées 1790–1796
the diocese of Płock, which made up the majority of the area, already had an existing survey. It is worth noting that there was a second, ameliorated, version of this map drawn in 1798 based on a refined survey. Other maps drawn from 1784 to 1804 covered the entire Crown except for the Gniezno and Sieradz palatinates, which, for unknown reasons, were not drawn. Out of twelve maps, five were printed in the Paris Tardieu printing house, seven remained in manuscript form, and two remain until today in the form of poor photocopies (Perthées 1783-1804, Rutkowski 2016) (Tab. 1).

The maps’ contents is typical for a medium-scale topographic map of these times (Fig. 3). First of all, settlements are represented and distinguished according to three criteria: type (towns, villages, or other settlements), ecclesiastical function, and size. As for villages, there are parish villages: villages with a Latin or Orthodox church, or a monastery. In terms of size, these can be “long and large villages”, “villages”, “smaller villages”, or “even smaller villages” (the last two are depicted only on the map of Podlasie). A separate symbol for so-called Dutch settlements (pl: “Olędry”) in the Poznań palatinate, colonies (pl: “Romunki”) in the Płock & Dobrzyń palatinates, and “New settlements” in the Brześć & Inowrocław palatinates are included. Industrial facilities include, for example, mills, inns, and windmills. Separate symbols represent the settlements’ attributes, like post offices and leases. The settlement network is supplemented by roads, rivers, lakes, forests, and swamps, and a schematic relief shown using hachures. Therefore, all nine points from the parish surveys are covered by the map legend.

### Data model: primary data sources and secondary data sources

The methodology for collecting geographical information from historical sources differs from the acquisition of modern data. Gregory and Ell use the terms “primary” and “secondary” data sources. In the first case, data are taken directly from the real world using measurements, while in the second, it is taken from indirect sources, such as written sources or maps (Gregory, Ell 2007, p. 41-62; Sevara 2012, p. 75-76). For texts, classical editing methods involved transcribing them from the source (e.g. manuscript), editors preparing a preface, and adding footnotes and indexes. In the digital paradigm, the source can be annotated (image of the source) or tagged (text of the source). In both solutions, the content is structured and classified. For cartographic sources, the facsimile (scanned image) of the map is the basis for its vectorization in a
As the aim of the whole project was to prepare a map of the settlements and administrative units of the Crown during the second half of the 18th century, the basic element selected for indexing at the level of Entries is the toponym, mainly settlement names. Proper names are the most important attribute for distinguishing geographic entities from each other and an important component of the place’s identity (Czemy 2011, p. 33-36). The “toponym” is understood here as a proper name expressed as a noun in the nominative case. Adjectival forms are also recorded, but only when they indicate an economic facility. Neither institutions (e.g. Kraków bishops’) nor territorial units (e.g. Starostwo of Krasnystaw) are collected. Due to the diversity of spelling, historical names are recorded in both source (transliteration) and normalized (transcription) versions. If the source contains such information, the place type, the number of houses, the owners, and administrative affiliation are acquired for each toponym. Preliminary identification of settlements is also carried out by providing them with identifiers from the NRGM, the National Register of Geographic Names (pl: Państwowy Rejestr Nazw Geograficznych), or from detailed maps of the 16th century in HAP, the Historical Atlas of Poland (Słoń 2014). Such identification will speed up the work in the second stage of research when the data collected in the source-driven model will be transformed into an analytical model in which each toponym becomes a topographic feature with a name, type, and location. A feature’s name and type will be derived from Karol Perthées’ materials, and the location (geometry) will be based on the analysis of modern and historic maps.

Data model for maps

The process of map indexation has two objectives: (1) preparing data for a digital edition, and (2) providing a set of information for the development of a historical map presenting the settlements and boundaries of the Crown during the second half of the 18th century. The first objective required that the map and its content should be represented digitally in the source-driven model.
Figure 5. Database model for indexing sketches
Source: own elaboration.
as close to the source as possible. The second entailed that the data obtained from the maps should be georeferenced and related to the actual historical location of the features they represent. Combining these two objectives required taking a different approach to the vectorization of historic maps than the classic methodology based on map georeferencing and content vectorization.

The methodology and the data model treated maps as images with only a local reference system (XY Cartesian), and that the data model was subordinated to the map legend, which only after transformation could serve analytical purposes. First, the map sheets were entered into a GIS application and placed next to each other in the same way as the manuscripts of the sketches in the INDXR application. Second, a database structure was developed for the indexation of the maps, consisting of five tables designed to represent the structure of the maps’ content: Maps, Units, Features, Symbols, and Annotations (Fig. 6).

The Maps table contains metadata with basic information about the particular map sheet that is indexed, such as date of elaboration and issue, material form, etc. Entities in this table are actually vector frames drawn around each of the sheets.

In the Units table, data are collected together on the homogeneous areas that can be distinguished on the map. These areas can be of either geographical (district, palatinate) or graphic character (scale bar, legend, title, etc.) (Fig. 7). Districts, which were the smallest territorial units drawn on the map, were given attributes of parent administrative units, such as land and palatinate. It was thus possible to assign administrative affiliation to spatially subordinate entities indexed at successive levels: Features, Symbols, and Annotations. The relationship between Units and Maps also has a spatial character.

One of the key decisions regarding the indexation was to separate map symbols from map annotations and to collect them in two different tables: Symbols and Annotations respectively. Preliminary studies have shown that it is often difficult to clearly link a symbol (e.g. village or mill) to its description (e.g. village name) (Fig. 8). The method of indexing topographic features by treating them as a composite of the location, type, and name, would therefore fail to apply in our work. The entities collected in the Symbols table are a vector representation (point) of cartographic symbols forming the map content. This list was derived from the map legend, but categories that were not included in the legend are also included. In the Annotations table, toponyms are collected together along with transliterations and transcriptions. Their types are classified according to the typeface
Symbols and annotations, when indexed, have no relationship to each other. The polygons in the Features table, which spatially overlap the specific map symbols and annotations, allow them to be functionally and technically related (Fig. 9). The symbol and annotation for each feature are given the same identifier ("feature_id"). Typically, we have a 1:1 relationship, i.e., one symbol corresponds to one annotation, but others are not uncommon. It happened many times that we had an unnamed symbol, a name without a symbol, or – in the worst-case scenario – great difficulty in unambiguously assigning names to symbols. Note that relating symbols to descriptions at this stage is a work in progress, and is verified in further works that involve sketches and other historic maps. Currently, the features are preliminarily identified using the NRGN and HAP resources, in a similar way to the toponyms from the sketches.

**Discussion and summary**

Thanks to the use of GIS technology and the OGC standards in the INDXR application, it was possible to introduce the source
models of sketches and maps into the database structure using PostgreSQL/PostGIS. The two models are interrelated not only conceptually, but also physically through external (NRGN, HAP) and internal IDs. These are source models designed to store data from sketches and maps. An analytical data model is based on the transformation and combination of selected layers of information from the source models. Transforming the data from the source model into the analytical model is achieved primarily by linking Entries from sketches with Annotations from maps. In this way, the records in the Features table aggregate the source data and allow them to be refined and verified. Based on the collected attributes, it is possible to partially automate the process of identifying the toponyms and features from the two source models. Importantly, even when records are linked between the source and the analytical models, the initial relationship between all of the source records is maintained, and, thus, the connection can be modified as a result of a change in interpretation or test conclusions.

An example of such a procedure is illustrated by the diagram below concerning the Zemborzyce parish, located about 10 km south of Lublin (Fig. 10). A set of source data from maps and sketches is shown in the figures (Fig. 3 and 4). A preliminary analysis of the information stored in the source model consists of two steps. Both involve using the SQL query language and the spatial functions of the PostGIS database. In the first step, names and symbols collected from maps (“Maps.Annotations” and “Maps.Symbols”) are combined into topographic features (“Maps.Features”). In the second, toponyms acquired from sketches (“Sketches.Entries”) are related to names from the maps (“Maps.Annotations”) and, thus, indirectly also with topographic features (“Maps.Features”). The attributes related to the administrative affiliation and the type of the topographic feature are also taken.
into account in linking of the data, and enable them to be verified and to help identify any inconsistencies. The whole procedure of linking data from these sources is aimed at discovering and characterizing the relationship between the parish sketches and the special maps.

Acknowledgments
The research was supported by the «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 Karol Perthées» (no. 11H 18 0122 87) grant funded by National Programme for the Development of Humanities

**ORCID**
Bogumił Szady [https://orcid.org/0000-0003-0059-5596](https://orcid.org/0000-0003-0059-5596)
Tomasz Panecki [https://orcid.org/0000-0003-3483-2035](https://orcid.org/0000-0003-3483-2035)
References


