sciendo

Polish Cartographical Review
Vol. 55, 2023, pp. 56–72
DOI: 10.2478/pcr-2023-0006
Received: 18.10.2022
Accepted: 30.10.2023

MATEUSZ ZAWADZKI, MARIA STADNICKA
Maria Curie-Skłodowska University
Department of Geomatics and Cartography
Lublin, Poland
orcid.org/0000000221262822; mateusz.zawadzki@mail.umcs.pl
orcid.org/0000-0002-6166-480X; maria.stadnicka3@mail.umcs.pl

MICHAŁ LUPA, KATARZYNA ADAMEK
AGH University of Krakow
Department of Geoinformatics and Applied Computer Science
Kraków, Poland
orcid.org/0000-0002-4870-0298; mlupa@agh.edu.pl
orcid.org/0000-0003-2595-9296; kadamek@agh.edu.pl

The use of web-GIS and crowdsourcing in the acquisition of historical data on the example of Polegli1939 project

Abstract. The historical Polegli1939 geoportal is probably the first thematic portal using spatial-temporal databases to present information on people killed during the World War II. In this study, we address the problem of capturing, processing and sharing historical information by supplementing it with a spatial attributes. The aim of the study is to determine the role of crowdsourcing data and the use of web-GIS applications in the process of obtaining and verifying historical information. The study was carried out as part of the implementation of the "Spatial database of soldiers killed in warfare" project, which is also known as Polegli1939. In order to achieve this goal, we acquired information using three methods: manual, automated and Volunteered Geographic Information (VGI). Making spatial data available through the Polegli1939 geoportal enables users to popularize and verify it. The method includes a geospatial web platform (GeoWeb) as well as a VGI application with an integrated process for verifying submitted information. As a result of the project, nearly 35,000 records were obtained. Making the project available to a wider audience opens up the possibility of extending the presented research or replicating it in other countries, taking into account a wider chronological range than just World War II.

Keywords: Historical GIS, GeoWeb, Volunteered Geographic Information, historical data

1. Introduction

A constant element of historians and historical geographers contemporary research is the use of IT solutions, including GIS technology. The literature on this subject describes the extensive research carried out on the construction of historical databases, geographical information systems, and the use of both in research (Burleson & Giordano, 2015; Fleming et al., 2009; Gregory, 2003; Gregory & Ell, 2007; Hu et al., 2018; Leśniak et al., 2017; Szady, 2016). Moreover, geographic-historical research has largely revolved around technology-oriented GIS projects, spatial analysis, and the latter's relevance to ongoing processes and events. Increasingly, a new approach is being used that focuses on the study of "human-land" relationships based on the four elements of history: time, place, person and event (Hu et al., 2018; Kuna & Kowalski, 2020). The study of the relationship between a person involved in an event and the environment can be conducted on geographical and historical levels. Research

on past military actions and their social consequences also fits into this framework (Juhász, 2007, 2014). The development and implementation of this research in the spatial dimension requires the development and construction of databases, which are crucial for the collection of geohistorical data. The data collected in such databases enables temporal and spatial analyses to be performed. In addition, these analyses will allow new research questions to be asked and, perhaps, e.g., many previously unknown aspects of past military actions and their socio-economic impacts to be discovered.

The acquisition of big data, especially historical data, presents challenges due to its specificity and limited availability (Manning, 2013). Participatory GIS solutions can be particularly useful in this context, as implementing an application that allows for user participation can facilitate the data gathering process (Lafreniere et al., 2019; Larrain & McCall, 2019). Moreover, concept development and the implementation of such a solution requires the use of quality-assurance systems (Ali & Schmid, 2014; Goodchild & Li, 2012; Samulowska et al., 2021).

The use of Volunteered Geographic Information (VGI) as a solution for mass data acquisition in scientific research, is not new but has recently gained increasing popularity (Finnane et al., 2018; Jiang & Thill, 2015). This is undoubtedly influenced by the growth of digital humanities, which integrates new technologies into research. Crowdsourcing allows for the engagement of a broader community in tasks that traditionally fell under the responsibilities of scientists (Dunn & Hedges, 2013; Terras, 2015). Within the humanities, these solutions have been successfully applied in areas such as archaeology (Sylaiou et al., 2013), history (Aucott et al., 2019; Ferreira et al., 2018), and art history (Dossin et al., 2017).

This project was based on information about soldiers killed during the Invasion of Poland in 1939. This period is extremely important from the perspective of historical-geographical research concerning military operations. It covers events which marked the beginning of World War II and warfare during this period. Moreover, it was characterized by great dynamism and scope, both in terms of space and the number of people affected by the hostilities. It is also worth adding that there is still no centralized system for collecting spatial information on battles, movements of military units, or the locations of soldiers' deaths. Existing solutions typically provide information about burial places, such as the location of cemeteries. Furthermore, these projects often focus on specific areas, such as the battlefields along the Bzura River (Heroes of the Battle of the Bzura, 2021). In the project carried out so far, data acquisition covered the previous territory of Poland within the borders of September 1939. Data related to the eastern half of the Second Polish Republic will be introduced in the next stage of the project.

The article discusses the process of acquiring historical information, processing it, defining its attributes, and assuring its quality. In addition, a scheme for storing spatio-temporal objects and relations between spatial and temporal objects in a data model is presented. Volunteered Geographic Information (VGI), which is an essential part of our solution, allowed the introduction of new data (through the portal provided by us) and provided the opportunity to perform mass verification of data correctness. Based on the proposed approach, the proposed procedure for obtaining and verifying data on historical figures was verified. Identifying connections between people (their military unit, participation in events [battles] and their place of death [mass graves]) allows this information to be linked to locations and dates. This makes it possible to study the movement of military units in space and to correlate the data with natural and cultural factors.

2. Materials and methods

A search of the subject literature and internet resources indicated the existence of publicly available data collections on fallen soldiers (Commonwealth War Graves, 2021; Heroes of the Battle of the Bzura, 2021; Honor the fallen, 2021; Straty osobowe i ofiary represji pod okupacją niemiecką 1939–1945, 2021). These collections are in both analog form (studies, lexicons and lists) and in the form of digital databases. The main problems diagnosed when analyzing them were:

− lack of harmonization in the recording of information,

− fragmented resources,

− information on a specific area (city, region of military operations).

These observations influenced the selection

MÜLLER Józef, strz. 62 pp, † 8.10.1939 r. w Giżycach, 31-KW Giżyce.
MÜLLER Józef, szer. 40 pp 5 DP, † 30.09.1939 r. w Jarosławiu, 33-KW Jarosław.
MÜLLER Leon, strz. 1 mps, † 11.09.1939 r. w Dębogórze, 10-CW Gdynia, Redłowo.
MÜLLER Wilhelm, szer., † 25.09.1939 r. w Warszawie, 1-KW Warszawa, Żoliborz, ul. Powązkowska.
MÜLLER Zofia, sanit., † 1939 r. w Gdyni, 10-KW Gdynia, ul. Witomińska.
MULMEISTER Izaak, st. szer. 56 pp, † 3.10.1939 r. w Łowiczu, 38-KW Łowicz, ul. Listopadowa.
MUNECHEBER Henryk, strz. 30 pp, † 21.09.1939 r. w Warszawie na Placówce, 1-KW Warszawa, Żoliborz.
MURAWSKI Kazimierz, ☆ 1910 r., kpr. 1 p.lotn., † 3.09.1939 r. w rej. Kocka, 1-MI Warszawa, Żoliborz, ul. Powązkowska.
MURAWSKI Leonard, żołn., † 11.10.1939 r. w Kutnie, 31-KW Kutno.
MURAWSKI Michał, ☆ 9.09.1914 r., strz. 68 pp, † 1939 r. w Szwarocinie, 38-KW Rybno.
MURAWSKI Witold, strz., † 17.09.1939 r. w Jagodzinie, 36-MZ Domanice.
MURCZYK Wacław, strz. baon ON, † 8.09.1939 r. w Koleczkowie, 10-KW Gdynia, ul. Witomińska.
MURCZYŃSKI Wincenty, st. uł. 9 p.uł., † 20.09.1939 r. w Łomiankach, 1-KW Kiełpin.
MUREK Andrzej, ☆ 18.09.1909 r., szer. 58 pp, † 16.09.1939 r. w Złotej, 38-KW Rybno.

Figure 1. A scan of a page from *The Book of Buried Soldiers*, illustrating the form of the source data record

of one primary input resource: *The Book of Buried Polish Soldiers Fallen in World War II* (hereinafter: *The Book of Buried Soldiers*, Affek-Bujalska & Kospath-Pawłowski, 1993), which is a two-volume publication. This study is a collection (available in an analog version) of information concerning the burial places of soldiers whose identities are known. The data available in this lexicon-like publication comes from the resources of the Polish Red Cross.

2.1. Data acquisition methods

The specificity of the input data influenced the definition of the data extraction methods. *The Book of Buried Soldiers* has the character of a lexicon, and the special characters used in the text (Figure 1) define the type of the attribute information.

The lack of a tabular record hinders the process of automatic text entry into the structured database. Therefore, it was decided to process the data from analog to database form in three ways:

1) manual method,
2) automated data acquisition,
3) volunteered geographic information (VGI).

The implementation of the first and second methods involved processing the content of *The Book of Buried Soldiers* and required the preparation of a digital copy of this book. For this purpose, the two volumes of the study were scanned (826 pages), then the prepared scan files were organized and described with metadata. The scans were divided alphabetically into two groups: the first group was processed manually, while the second was processed using an automated method. The subsequent steps of methods 1 and 2 will be discussed in detail (Figure 2) in sections 2.1.1 and 2.1.2.

The VGI method was intended to check the data digitized by methods 1 and 2 and to obtain data from other sources. Following the earlier publication of preliminary project data, this method was implemented in the second phase of the project. It is described in section 2.1.3 and in section 3.

2.1.1. Manual method

The manual processing of data from analog to digital form required the involvement of a team. Each project participant had to process 40 pages (approx. 1,500 records) from *The Book of Buried Soldiers*. Each participant read the data and inserted it into a structured table in MS Access. The next stage was the initial verification of the information. Each person from the digitization team verified the content of another person's table against the scans (data source). In the

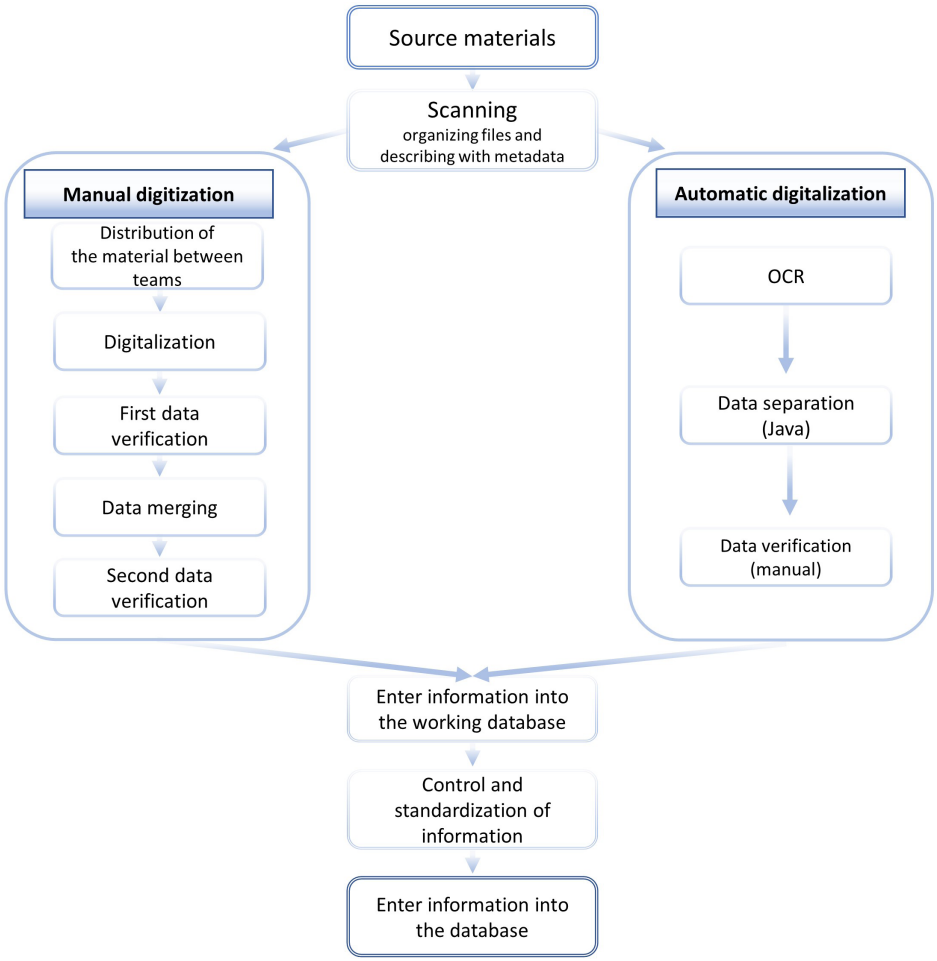

Figure 2. Schematic workflow for preparing historical material for addition to the database

next stage, all the tables were merged into one, which then also underwent data quality control. After this two-step verification, the information was added to the database.

2.1.2. Automated data acquisition

The automation of the data digitization process required the development of a workflow using multiple functions. One by one, these functions were used to prepare the data for subsequent processing stages while ensuring their quality. The first stage involved the use of an optical character recognition system OCR (ABBYY FineReader software). The inconsistent quality of the scans necessitated some pre-processing. The correction of errors was carried out using the following functions: division of adjacent pages, detection of page orientation, correction of skewness, straightening of text lines, improvement of trapezoidal distortions, correction of image resolution. This was followed by use of the training mode to improve OCR accuracy (Support ABBYY, 2021). This function is necessary for symbols in the text that are not included in the character coding system. The use of the training option avoided a pro-

blem related to recognition of the cross sign (signifying death) as the number 1 or the letter f, which was the case when using only the standard templates. The scanned cards were converted into editable text in .txt format.

The information processed into text form reflected the content of *The Book of Buried Soldiers*. In many cases, data was missing (e.g., dates of birth and death, first name), which disturbed the fixed layout of the information record and made it difficult to automatically separate data into attributes. This necessitated the creation of a script which, based on the relevant characteristics of each data type, separated the different types of information and made records in the appropriate columns. The pre-processing stage described in the previous step did not eliminate all errors as it was not possible to avoid errors such as:

- data for one person were in two rows,
- data for two people were in one row,
- missing source information (e.g., date of birth).

To save the data correctly, an automated script was used to split the data (initially the data was saved as a string within a single column) into the prepared schema. The script recognized (based on the characteristics of the data recording) the individual parts of the text string and divided it into information corresponding to one person. The developed method allowed the identification of the following elements:

- first and last name,
- date of birth,
- military rank and military unit,
- date of death,
- place of death,
- place of the burial.

After performing recognition operations on all of the elements, the resulting values were saved on a single line and then exported to a working database.

2.1.3. Volunteered geographic information

The third method of obtaining information was the use of a crowdsourcing application. The development of this application followed the dissemination (in the form of a geoportal) of data acquired by the previously described methods (Section 3). The project and the data provided were positively received by users. The research value of the data was indicated and the novelty of the spatial data visualization was appreciated. It was also pointed out that the information collected and made available contained errors, but these were not the result of faulty data processing. Incorrect or incomplete information resulted from actual errors in *The Book of Buried Soldiers* and concerns a wide range of attributes (soldiers' first or last names, dates of birth and death, places of death).

The implementation of VGI solutions was identified as a way of adding new records to the database and verifying errors. An additional functionality of the application is the possibility to contact the administrators of the portal and enter data on the location of a deceased person's death or burial place (grave). Users enter attribute information in accordance with the structure of the database: first name or names, surname, military rank, military unit membership, date of birth, date of death, place of death, cemetery location, cemetery section number. In addition, in order to be accepted by the administrator, each entry should have the source of the information indicated and, optionally, the contact details of the author of the entry so they can be contacted to clarify the data. Additionally, it is possible to add attachments (e.g., documents) confirming the information entered.

The next stage is verification of the correctness of the data by the system administrator. Data are checked for correctness and possible duplication. Moreover, the data validation step

Figure 3. Procedure for adding data using the VGI application

also assesses the authenticity of the data. After verification, data is added to the database and made available in the public map application (Figure 3). If the data is not valid, it is removed from the database.

2.2. Evaluation of data-acquisition methods

The data-acquisition methods were evaluated on the basis of the problems generated by each method. The analysis of the advantages and disadvantages showed that each of these methods requires some degree of correction at the level of data control. The results of the analysis are presented in Table 1.

The manual method was the most imperfect (due to the human factor). It generated a large number of minor errors, such as imprecise assignment of information to the database structure, single inconsistent characters, and misspellings of place names. Records of local names required special verification. The spatial dimension of the data requires attention to the source record of place names. The source record of a place name in the locative form (e.g., "w Parzniewie") often results in a misreading of the nominative form: "Parzniewo" instead of "Parzniew". Such uncertain place names were verified by searching for both possible variants in the National Register of Geographical Names (Polish registry – PRNG, 2021) and choosing the one correct name. In the example given above, "Parzniewo" does not exist in PRNG. Another example is the existing villages of "Domanowo" and "Domanów", whose the locative is "w Domanowie". If the place of burial was the same as the place of death, then a further search for the correct record of the place of death was unnecessary because the burial was always written in the denominator in the source material. One more example is the incorrect recording of military ranks that were abbreviated in the source material, so these were also verified and harmonized (e.g., "wachmistrz" and "wachm"). Other problems encountered when verifying the database were duplicates or incorrect formatting of date-storing cells, which often changed the resulting data, e.g., the date record changed. In addition, the efficiency of this method was not sufficient.

The use of automation in the digitization process made it possible to minimize problems caused by the human factor. Diagnosis of the problems that arose during testing of the manual method made it possible to eliminate a significant number of them. The only remaining problem was the correct reading of place names in the dative form and their conversion to the nominative version.

The last method used is VGI. Involving the community in entering information eliminates the previously described errors. The information entered usually results from the user's research interests, thanks to which it is correct in the vast majority of cases. The only drawback of this method is its low efficiency and the constant need to promote the project and use various forms of incentives to co-create the database.

2.3. Data harmonization and geometrization

An important step in the project was data harmonization and geometrization. Due to the large discrepancies in the source data records

Table 1. The advantages and disadvantages of the methods used to digitize data

✘ – WRONG ✓ – CORRECT

Problem	Manual method (unsupervised)	OCR + automation	VGI
Assignment of data to the database structure	✘	✓	✓
Digitizing information (e.g. place name)	✘	✘	✓
Typo error	✘	✓	✓
Efficiency	✘	✓	✘

and their specificity (historical data), this process required the development of a detailed workflow (Figure 4). The data unification process mainly related to the attributes that define the location, i.e. place names. Geometricization consisted in assigning geographical coordinates to the places of death and burial. As a result of the spatial range of the source data, the research area adopted for the study was the contemporary borders of Poland.

2.3.1. Data harmonization

The data harmonization stage included attributes such as place names, dates of birth and death, military ranks and military units. Harmonization turned out to be especially necessary in the case of place names. This concerned several recording variants, the first of which was related to differences in the notation of the same place of death in the source data (e.g., "Płock, Radziwie", which can also be written as "Płock, na Radziwiu"). The harmonization of recording consisted in determining the correct form of the nominative form of the local name. Another case which was subject to correction was records of locations which did not clearly indicate the place of death and only suggested that the soldier died in the vicinity of a specific location, e.g., near Ciechanów, near Falenica, near Kock. In such cases, the place of death was assigned as the place specified in the record.

2.3.2. Geometrization

The geometrization process consisted in indicating the location of each soldier's death. As a spatial reference, the place of death of a particular soldier was taken and was accurate to the center of the location (centroid). For geocoding, data from The National Register of Geographical Names (PRNG, 2021) was used in the form of a point vector layer that stores information on all currently existing localities in Poland. In the first stage, the database records were assigned ID_PRNG numbers by performing a join operation (column: name_ place_ of_death) on all matching place of death names from the table and the names of locations from the reference layer. The next step verified that the correct ID numbers had been assigned properly. It was not possible to automatically join all the locations, therefore several paths were taken to verify the information.

The verification of places of death using the PRNG register was divided into several interrelated stages and is presented in Figure 4. Green color in the diagram refers to the correctness of the place name and its presence in the register. If the place name was not present in the register, its correct recording in the database was checked, the name was corrected as necessary, and the process was restarted. However, if the record was correct, it was possible that a settlement's name had been changed since 1939. In such cases, the existence of the historical location was verified using archival documents or literature. If the name was identified, the location was assigned a contemporary name ID from the PRNG and the process was restarted. Otherwise, its existence was verified using old maps from 1930–1940 (WIG Maps) and the current name was indicated on a modern map (TOPO, OSM). If this stage was completed successfully, the process was restarted; however, if the location was not correctly identified, the status "no spatial reference" was entered in the database.

Another example of verification was the establishment of the correct form of place names that appear in the database in singular form (e.g., Gostków). This concerned settlements which could not be verified at the harmonization stage. The verification consisted in determining whether both possible Polish forms of the nominative form of a name (e.g., Gostków and Gostkowo, which refer to two different localities) existed in the register. Next, the verification process was restarted as in the case of other single or multiple locations, i.e., death locations were verified by burial place using the Select By Attributes tool (ArcMap software). The locations selected through the query were evaluated in terms of distance: the closer the places of death and burial were to each other, the higher the probability of identifying the place of death. Once the place of death had been identified, it was assigned an ID number from PRNG and given the status of "spatially referenced".

Results indicating a lack of correspondence in the distance between the places of death and burial also required verification. In such

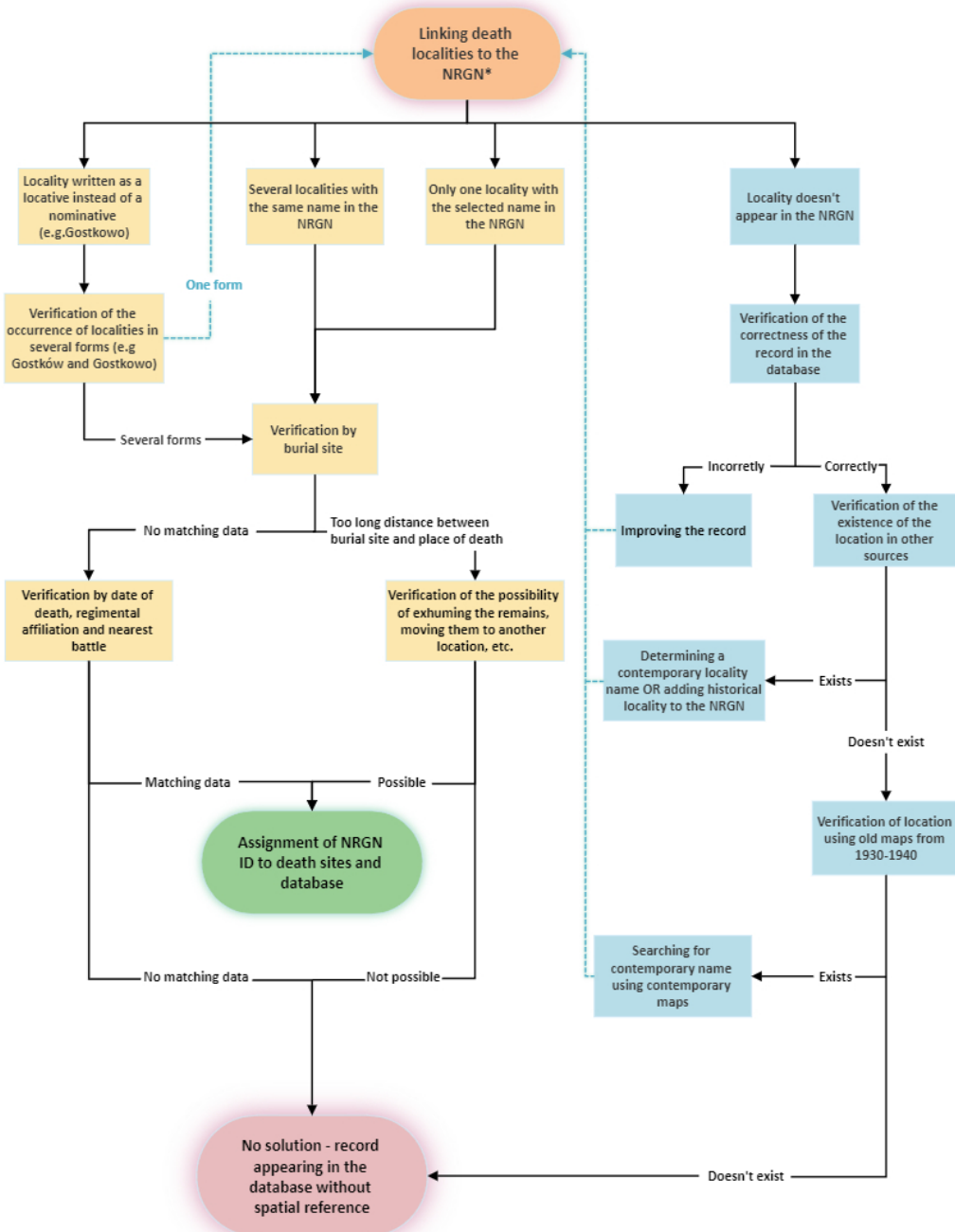

Figure 4. Geometrization scheme – determining the correct location of the place of death of a fallen soldier (* – National Register of Geographical Names)

cases, another verification was carried out in order to determine whether an exhumation had been carried out at a later date, as a result of which the corpse was transferred to another location (usually much further away from the place of death). If no exhumation had taken place, the location was given the status of "no spatial reference". If exhumation had taken place, the location of the death was assigned ID_PRNG and the status "with spatial reference".

If verification and identification of the place of death according to the place of burial did not allow the identification of the correct location, cross-checking was carried out with reference to information on the date of death, regimental affiliation and any battles fought in the vicinity. In the case of unclear records, information common to a number of individuals was checked (e.g., time of death and participation in the same battle); this allowed the probable place of death to be determined for individuals for whom the information was incomplete. The outcome of this stage was also decisive for assigning a specific status to the location.

3. Results

The proper acquisition, collection and sharing of data is crucial for conducting historical-military research, so the results focus mainly on the data quality assurance solutions implemented to map data to soldier death sites. We first focused on the data extraction process from the available studies, which provided us with the input data for the analysis. Then, as a result of information we received about instances of erroneous and incomplete data, we implemented a crowdsourcing solution to allow user interaction. As a result, we created and proposed a portal based on the ArcGIS platform, which was configured and adapted to the project requirements.

3.1. Description of the Web-GIS technologies utilized

The technological basis of the project is the GeoWeb platform available at https://polegli1939.pl/mapa. It consists of a map application, including dashboards (web mapping applications based on ArcGIS Online (AGOL) components; see Figure 5) and a data collection application based on customizable Survey123 for ArcGIS (Survey123; Esri Inc., Redlands, CA, USA). The whole system is available to users for free on a smartphone or through a web browser in two language versions: Polish and English.

The application framework was developed based on the developer version of ArcGIS Web AppBuilder (WAB) software, which is dedicated to geoportals and is based on the JavaScript programming language and Esri API. The Application Programming Interface (API) to Esri extends the capabilities of the JavaScript language with geospatial functionality and enables faster implementation and communication with these components in the application being built. For the needs of https://polegli1939.pl, a design template for the dashboard was selected and functionalities used in the application were adjusted: appropriate widgets were chosen from the WAB library, most of which are configurable from the application level. Some of them required development work, such as improving data reception via REST (Representational State Transfer) or ensuring that they display properly (reading of received objects). The next step was to export the data in the file geodatabase to ArcGIS Online and then create Feature layers from them and import them into the map. Here the use of the Esri cloud solution proved to be effective because it did not require a powerful dedicated server to process thousands of records. The map prepared in such a way was embedded in the application framework; then, after the development work was finished, it was placed in a dedicated server space. WAB allows the service to be launched quickly without additional compilation of files or building of applications. The whole system was secured by appropriate certificates and scripts for tracking traffic and user behavior (without collecting personal data).

The VGI application includes a form for user input. The collected data is sent to the vector service via REST services. The quality of the raw data is then checked in the database using ArcGIS Desktop integrated with REST. Finally, the administrator-checked data is returned to the database and presented as results in the map application.

The questions proposed in the form (section 2.1.3.) were included in the VGI application, which includes a short scrolling form to

Figure 5. The main view of the application presenting the results of the project. The page consists of the following elements: (a) a map window displaying data acquired in the project – point locations of soldiers' places of death; (b) a form of predefined SQL queries to the database; (c) a link to the VGI application for entering locations of cemeteries, graves and burial sites; (d) a link to the VGI application for entering information on soldiers killed in armed actions; (e) a legend to the main map (a); (f) links to the project's social media; (g) pop-up window with details of a fallen soldier

facilitate user input. The optimization of the content scope made it possible to avoid discouraging users, as is often the case with complex forms, and to focus on the right part of the form, namely indicating the location (place of death or burial of the soldier). In order to improve the quality of the collected data, questions about contact details were implemented in order to prevent submissions that do not comply with the rules of the project.

The interaction between the Survey123 application and the web map module is based on a typical WebGIS architecture. The mobile application is based on Survey123 for ArcGIS components and is accessible from the main project website https://polegli1939.pl/ as well as from the application https://polegli1939.pl/mapa. The form consists of three parts: a section for text input, a map and the option to add attachments. It is available in two language versions, Polish and English. The application contains ten questions, which are supplemented by the geolocation of the place of death or the location of the soldier's grave. Additionally, users can add attachments, such as digital copies of documents (sources of information), thereby authenticating the database entry (see Figure 6). The user can also add additional information relevant to the verification of the submitted content and an email contact for additional consultation. After completing the survey, a "save" button is active at the bottom and the submission is ready to be uploaded to the geodatabase in the cloud.

3.2. Results of the project

During the two-year duration of the project, 34,898 records were collected in the data-base (as of 10 September 2021). Using manual and automated methods (first stage of the project), 29,713 records were entered into the database; as a result of the geometric process, 29,113 locations of places of death were identified. Using the VGI application (second stage of the project), users submitted 5,185 new entries to the database, and 2,379 requests concerned the verification of existing information in the database. The data sent by the form were verified and the database was updated on a continuous basis. The highest user activity was in the week in which the project was made public; this high level of activity lasted for another week and then decreased. Another increase in data was related to the inclusion of permanent project collaborators who verify data and add new content.

Professional historians, genealogists, amateur historians and family members of fallen soldiers were identified among the recipients and active participants of the project. These individuals submitted new data and verified the provided data as individuals or as representatives of research and social institutions.

The verified data are made available in the map application available at https://polegli1939.pl/mapa/ (Figure 7). Users can display available data in spatial and descriptive form. The platform makes it possible to search for the fallen through a form: by name, military affiliation of the place of death, date of death or through a spatial query. The results of the query are displayed in attribute form in the left side panel and highlighted locations on the main map. The location of the place of death is generalized to the location level, hence clicking on a selected pin returns a value indicating the number of people who died within that location (Figure 7c).

4. Discussion

The project introduced a social innovation in the collection and verification of historical data, where users (professional historians, local historians, genealogists) work to improve the quality of publicly available historical data. Public perception of the data was enhanced by its spatial context and the use of an interactive map as a kind of layout for the database. It was noted that users identify places associated with local history, and an interactive map is easier to perceive than, for example, a data table record. For the most part, data verification related to the area near the respondent's residence or origin.

Providing data free of charge via an intuitive and responsive map application expanded the audience and users – especially mobile device users. The growing popularity of the project was probably also influenced by media interest (information available on the project website at: https://polegli1939.pl). Each presentation of the project in the media resulted in increased user activity. The activity was not limited only to page views but especially concerned data

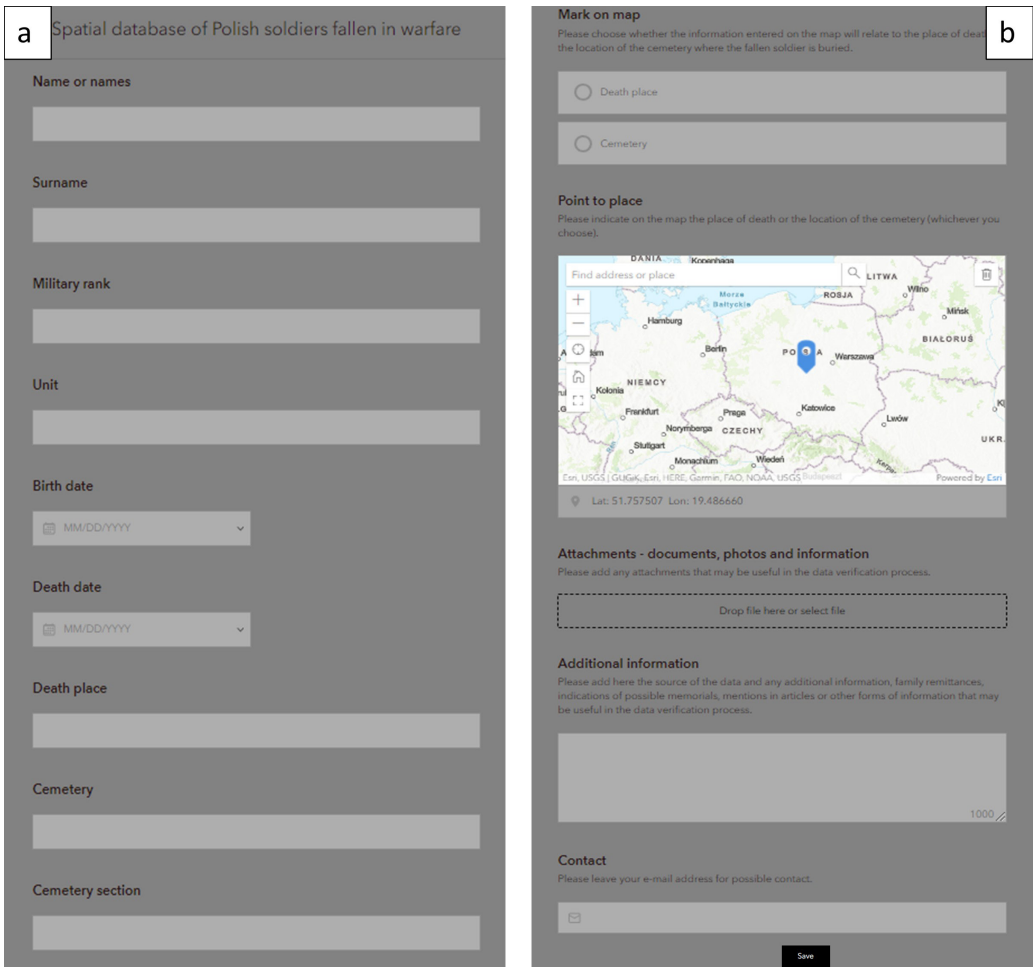

Figure 6. Layout of Survey123: (a) the part of the form relating to the attributes describing the information entered about the fallen soldier; (b) the part of the form specifying the type of location (place of death or burial) of the soldier, and a map on which the user indicates the location (pointing on the map, entering the address or indicating the location using the GPS receiver in the mobile device), the form for adding attachments (e.g., documents), the form for entering additional information and contacting the respondent

verification, namely entering corrections to existing data in the database.

The implementation of the project indicated several problems in terms of obtaining information and making it available. When working on geohistorical data, special attention should be paid to its specificity (Couper et al., 2001; Gregory & Ell, 2007). The main problem concerned the location of death and burial sites. This especially manifested itself in the identification of locations that appeared in the data source but whose names have now changed or which no longer exist. Reading and identifying place names written in a form other than the denominator was also problematic, making it difficult to automatically link source data with reference data (e.g., PRNG). As part of the project development, a mechanism will be im-

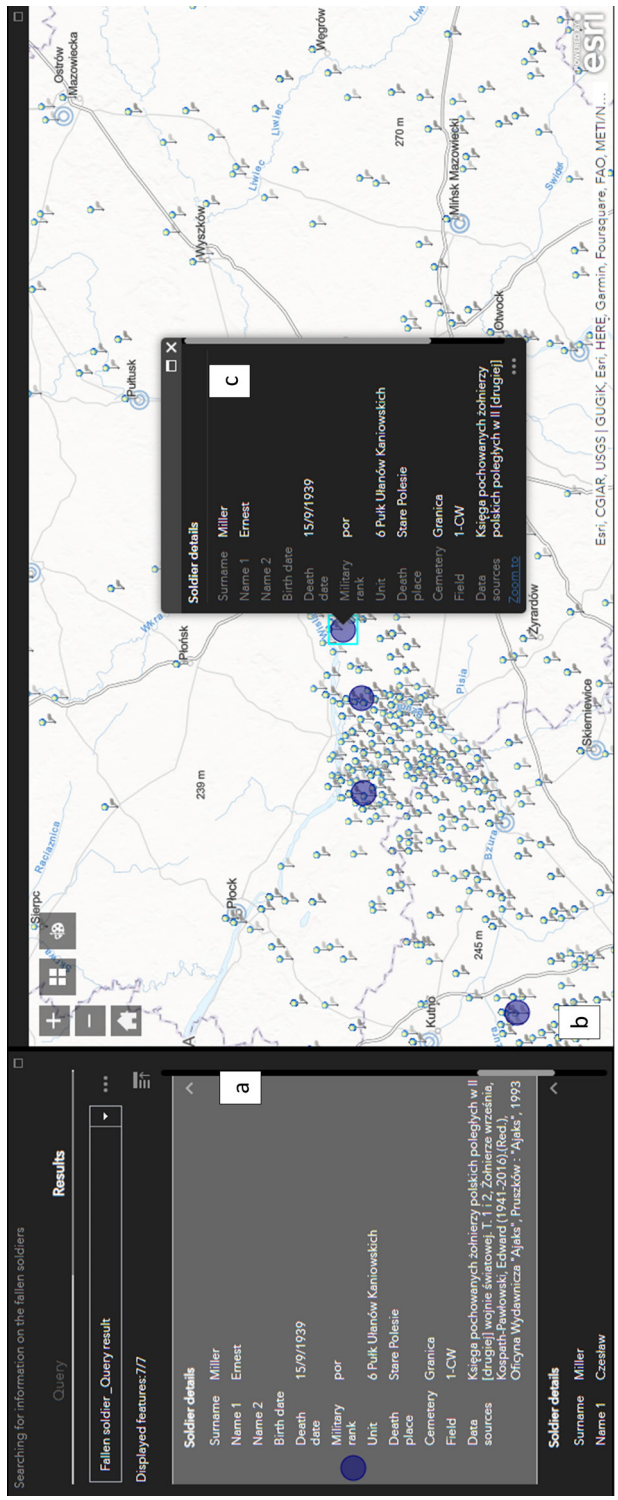

Figure 7. View of the map application window: (a) visualization of search results in descriptive form; (b) indicated locations on the map; (c) pop-up window with displayed attributes of the identified object and number of occurrences in the database referring to the indicated location

plemented to automatically recognize and identify place names written in a different form.

Another problem was the use of generalization in indicating places of death and burial (in relation to locations outside cemeteries). There were two reasons for adopting the centroid of locations. Firstly, the identification and indication of even the approximate place of death is a difficult and time-consuming process that is sometimes impossible. The intensity of combat during World War II and the limited precision of information sources sometimes made it impossible to pinpoint the exact location. For this reason, the applied unification (to towns) seems to be the most optimal solution.

A new value in the research was the use of VGI solutions. These resulted in more precise localization of transmitted data and eliminated the process of geometrization, which is necessary when digitizing the content of historical sources and studies. Although accurate location data was obtained, a generalization process is applied to these before they are made available in a publicly accessible portal. In many cases, indicating the exact locations of deaths, especially in the case of new hitherto unknown data, may contribute to attempts at illegal exploration and searching for material remains of armed activities.

The portal and data resource built as a result of the project represent a contribution to in-depth research in the field of military history. Until now, there has been no spatial database that collects data on fallen soldiers in Poland. A change in the current approach is the presentation of locations of places of death. The existing databases collect information resulting from cemetery inventories. The new spatial aspect makes it possible to ask new research questions. It is already possible to ask whether the collection of the fullest possible data on specific persons in space and time, connected, among other things, with military actions, will make it possible to discover hitherto unknown facts, such as marching routes, locations of potential (hitherto unknown) places of death or burial of soldiers. Moreover, the development of a spatial database will allow for analyses aimed at juxtaposing tactical issues and natural conditions with the course of battles and their demographic effects.

To this end, the GeoWeb application will be extended with additional functionalities such as spatial analysis, which will allow research at the interface of history and geography to be conducted in a more accessible form. Creating a participatory approach to HGIS research can increase its effectiveness. Taking research beyond the walls of universities is already being done in the form of new community initiatives (Lafreniere et al., 2019). Thus, involving the public in data collection and verification and providing immediate public access to spatial-temporal data can contribute to this. Creating a tool for making family histories publicly available by contributing one's own historical spatial knowledge to the general public can ensure the sustainability and success of projects using VGI (Sui et al., 2013).

5. Conclusions

In the field of geographical-historical research, there has not yet been a project in Poland that presents information about soldiers killed in warfare. Previous solutions focused on the presentation of graves (burial places). Information on persons was assigned to a location (a grave or cemetery site). The technological and methodical solution developed within the project enables a structured set of data to be collected, verified and made available in one place. The application of a crowdsourcing approach to data acquisition and verification introduced a new quality to the process of building geohistorical data resources. The resulting spatial database enabled the realization of several objectives. The first of these is that a new dimension is given to the study of military actions. Another is the use of a new way of presenting data in spatial terms, which contributes to an increase in public interest in the project. The spatial dimension of the research means that users are able to better identify with a place that is expressed in the form of a map (Hillier & Knowles, 2008; Stedman, 2002; Stuart, 2012).

The implementation of WebGIS and VGI solutions for historical data acquisition has yielded benefits in three dimensions: application, education, and science (Figure 8). The dissemination of applications using data generated from *The Book of Buried Soldiers* raised awareness of the need for VGI applications. This led to the activation of the online community, which has much greater capabilities for introducing new data and verifying it more effi-

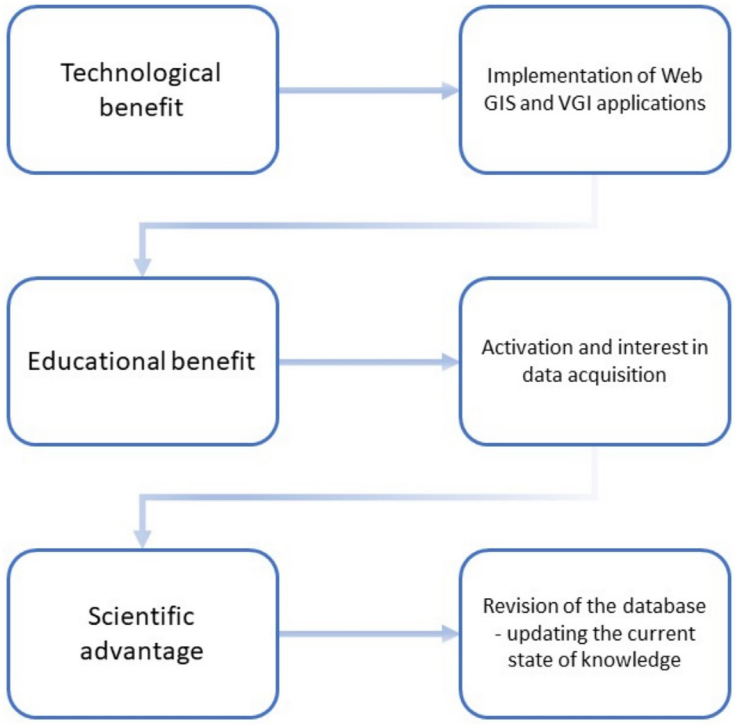

Figure 8. The scheme illustrating the benefits of sharing data about fallen soldiers using Web GIS and VGI solutions

ciently than smaller project teams or groups of scientists. By combining VGI solutions with society's enthusiasm for exploring the past, there is an opportunity to acquire large datasets, as seen in other fields, such as the exact sciences (Finnane et al., 2018). The engaged group of individuals thus verified a significant portion of the data, making corrections to those records that contained errors. In this context, we can speak of a scientific benefit, namely the updating of the existing state of knowledge.

Experience to date with the promotion of applied solutions has shown the potential of the project and proved its importance in the following areas: acquisition and verification of geohistorical data, presentation of data in the form of historical geoportal, and data analysis.

Acknowledgements. The authors would like to thank the members of the GeoIT Student Research Group of Geoinformatics for their participation in the project implementation. Special thanks go to: Sandra Chrustowska, Alicja Nieć, Wiktor Mróz, Daria Soszyńska, and Krystian Święcicki. Furthermore, special thanks to Rafał Perzyna for his assistance and support in the project's implementation.

References

Affek-Bujalska, B., & Kospath-Pawłowski, E. (Eds.). (1993). *Księga pochowanych żołnierzy polskich poległych w II wojnie światowej*, vol. 1. Żołnierze września, A–M; vol. 2. Żołnierze września, N–Z. Pruszków.

Ali, A. L., & Schmid, F. (2014). Data Quality Assurance for Volunteered Geographic. In M. Duckham, E. Pebesma, K. Stewart & A. U. Frank (Eds.), *Geographic Information Science. GIScience 2014. Lecture Notes in Computer Science*, vol. 8728. Springer.

https://doi.org/10.1007/978-3-319-11593-1_9

Aucott, P. J., Southall, H., & Ekinsmyth, C. (2019). Citizen science through old maps: Volunteer motivations in the GB1900 gazetteer-building project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *52*(3), 150–163. http://dx.doi.org/10.1080/01615440.2018.1559779

Burleson, S., & Giordano, A. (2015). Extending Metadata Standards for Historical GIS Research: A Case Study of the Holocaust in Budapest and the Armenian Genocide in Turkey. *International Journal of Applied Geospatial Research (IJAGR)*, *6*(4), 88–109. 10.4018/ijagr.2015100105.

Commonwealth War Graves. (2021, September 10). https://www.cwgc.org/find-records/find-war-dead/

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web Survey Design and Administration. *Public Opinion Quarterly*, *65*(2), 230–253. https://doi.org/10.1086/322199

Dossin, C., Kong, N. N., & Joyeux-Prunel, B. (2017). Applying VGI to collaborative research in the humanities: the case of ARTL@S. *Cartography and Geographic Information Science*, *44*(6), 521–538. https://doi.org/10.1080/15230406.2016.1216804

Dunn, S., & Hedges, M. (2013). Crowd-sourcing as a Component of Humanities Research Infrastructures. *International Journal of Humanities and Arts Computing, 7*(1-2), 147–169. https://doi.org/10.3366/IJHAC.2013.0086

Ferreira, K. R., Ferla, L., de Queiroz, G. R., Vijaykumar, N. L., Noronha, C. A., Mariano, R. M., Taveira, D., Sansigolo, G., Guarnieri, O., Rogers, T., Lesser, J., Page, M., Atique, F., Musa, D., Santos, J. Y., Morais, D. S., Miyasaka, C. R., Almeida, C. R., do Nascimento, L. G. M., Diniz, J. A., & dos Santos, M. C. (2018). A Platform for Collaborative Historical Research based on Volunteered Geographical Information. *Journal of Information and Data Management*, *9*(3), 291. https://doi.org/10.5753/jidm.2018.2046

Finnane, M., Kaladelfos, A., & Piper, A. (2018). Sharing the archive: Using web technologies for accessing, storing and re-using historical data. *Methodological Innovations*, *11*(2). https://doi.org/10.1177/2059799118787749

Fleming, S., Jordan, T., Madden, M., Usery, E. L., & Welch, R. (2009). GIS applications for military operations in coastal zones. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(2), 213–222. https://doi.org/10.1016/j.isprsjprs.2008.10.004

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120. https://doi.org/10.1016/j.spasta.2012.03.002

Gregory, I. N. (2003). *A Place in History: A guide to using GIS in historical research* (1st ed.). Oxbow.

Gregory, I. N., & Ell, P. S. (2007). *Historical GIS: Techniques, methodologies and scholarship* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511493645

Heroes of the Battle of the Bzura. (2021, September 10). http://www.bohaterowie1939.pl/

Hillier, A., & Knowles, A. K. (Eds.). (2008). *Placing History: How Maps, Spatial Data, and GIS is changing historical scholarship*. ESRI Press.

Honor the fallen. (2021, September 10). https://thefallen.militarytimes.com/

Hu, D., Guonian L., Jiang, N., Cao, W., Liu, L., & Li, Y. (2018). Historical GIS Data Model under Geographic and Historical Perspectives. *Journal of Geoinformation Science*, *20*(6), 713–720. DOI:10.1208/dqxxkx.2018.180025

Jiang, B., & Thill, J-C. (2015). Volunteered Geographic Information: Towards the establishment of a new paradigm. *Computers, Environment and Urban Systems*, *53*, 1–3. https://doi.org/10.1016/J.COMPENVURBSYS.2015.09.011

Juhász, A. (2007). A special GIS application – military historical reconstruction. *Periodica Polytechnica Civil Engineering*, *51*(1), 25–31. https://doi.org/10.3311/pp.ci.2007-1.04

Juhász, A. (2014). New achievements in WWII military historical reconstruction with GIS. *AARMS – Academic and Applied Research in Military and Public Management Science, 13*(3), 413–424. https://doi.org/10.32565/aarms.2014.3.4

Kuna, J., & Kowalski, Ł. (2020). Exploring a non-existent city via historical GIS systems by the example of the Jewish district 'Podzamcze' in Lublin (Poland). *Journal of Cultural Heritage*, *46*, 328–334. https://doi.org/10.1016/j.culher.2020.07.010

Lafreniere, D., Weidner, L., Trepal, D., Scarlett, S. F., Arnold, J., Pastel, R., & Williams, R. (2019). Public participatory historical GIS. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 52*(3), 132–149. https://doi.org/10.1080/01615440.2019.1567418

Larrain, A. A., & McCall, M. K. (2019). Participatory Mapping and Participatory GIS for Historical and Archaeological Landscape Studies: a Critical Review, *Journal of Archaeological Method and Theory, 26*, 643–678. https://doi.org/10.1007/S10816-018-9385-Z

Leśniak, A., Lupa, M., & Polończyk, A. (2017). Analysis of the Polish Home Army drop zones during World War II, using Geographic Information Systems. *Geomatics, Landmanagement and Landscape*, (3), 89–105. http://dx.doi.org/10.15576/GLL/2017.3.89

Manning, P. (2013). *Big Data in History*. Palgrave Macmillan.

National Register of Geographical Names (PRNG). (2021, September 10). https://www.geoportal.gov.pl/dane/panstwowy-rejestr-nazw-geograficznych

Samulowska, M., Chmielewski, S., Raczko, E., Lupa, M., Myszkowska, D., & Zagajewski, B. (2021).

Crowdsourcing without Data Bias: Building a Quality Assurance System for Air Pollution Symptom Mapping. *ISPRS International Journal of Geo-Information*, *10*(2), 46. https://doi.org/10.3390/ijgi10020046

Stedman, R. C. (2002). Toward a social psychology of place: Predicting behavior from place-based cognitions, attitude, and identity. *Environment and Behavior*, *34*(5), 561–581. https://doi.org/10.1177/0013916502034005001

Straty osobowe i ofiary represji pod okupacją niemiecką 1939–1945. (2021, September 10). https://straty.pl/

Stuart, E. D. (2012). The Spatial Humanities: GIS and the future of humanities scholarship. In D. J. Bodenhamer, J. Corrigan & T. M. Harris (Eds.), *Literary and Linguistic Computing* (pp. 471–472). Bloomington: Indiana University Press. https://doi.org/10.1093/llc/fqs016

Sui, D., Elwood, S., & Goodchild, M. (2013). *Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice*. Springer Netherlands.

Support ABBYY. (2021, November 13). https://help.abbyy.com/en-us/finereader/14/user_guide/ornatetype/

Sylaiou, S., Basiouka, S., Patias, P., & Stylianidis, E. (2013). The volunteered geographic information in archaeology. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-5/W1, 301–306. https://doi.org/10.5194/isprsannals-II-5-W1-301-2013

Szady, B. (2016). Spatio-temporal databases as research tool in historical geography. *Geographia Polonica, 89*(3), 359–370. http://dx.doi.org/10.7163/GPol.0059

Terras, M. (2015). Crowdsourcing in the digital humanities. In S. Schreibman, R. Siemens & J. Unsworth (Eds.), *A New Companion to Digital Humanities*. Wiley-Blackwell. https://doi.org/10.1002/9781118680605.ch29