

## Creating large, high-quality geospatial datasets from historical maps using novice volunteers



Adela Sobotkova <sup>a,\*</sup>, Shawn A. Ross <sup>b</sup>, Christian Nassif-Haynes <sup>c</sup>, Brian Ballsun-Stanton <sup>d</sup>

<sup>a</sup> Department of History and Classical Studies, Aarhus University, Denmark

<sup>b</sup> Department of History and Archaeology, Macquarie University, Sydney, Australia

<sup>c</sup> Independent Scholar Working in the Software Industry, Canva, Sydney, NSW, Australia

<sup>d</sup> Faculty of Arts, Macquarie University, Sydney, Australia

### ARTICLE INFO

Handling Editor: E. Hunja Waithaka

**Keywords:**

Historical maps  
Crowdsourcing  
Landscape archaeology  
Machine learning  
Participatory GIS  
Mobile computing  
FAIR data  
Bulgaria

### ABSTRACT

Unlocking data from historical maps for landscape analysis is costly. Automatic extraction using Machine Learning (ML) requires extensive preparation and expertise. Crowdsourcing scales better than direct digitisation by experts, but requires an appropriate platform and the technical skills to adapt it. Existing research provides little guidance as to when investments in these approaches become worthwhile. Here we present a customisation of the Field Acquired Information Management Systems (FAIMS) Mobile platform tailored to offer a streamlined, collaborative system for crowdsourcing map digitisation by volunteers with no prior GIS experience. Deployed in Bulgaria as an ancillary activity during 2017–2018 archaeological fieldwork, FAIMS Mobile was used to digitise 10,827 mound features from Soviet military topographic maps. This digitisation required 241 person-hours (57 from staff; 184 from novice volunteers), with an error rate under 6%. The resulting dataset was consistent, well-documented, and ready for analysis with a few hours of processing. A conservative estimate based on our work suggests our crowdsourcing approach is most efficient for digitisation projects of 10,000–60,000 features, but may offer advantages for datasets as small as a few hundred records. Furthermore, it indicates that systems designed for field data collection, running on mobile devices, can be profitably customised to serve as participatory geospatial data systems accessible to novice volunteers.

### 1. Introduction

This article presents a case study of crowdsourced cultural heritage digitisation from historical maps undertaken by volunteers using a lightweight, streamlined Geographical Information System (GIS) running offline on mobile devices. The value of this approach lies in the unanticipated success of a minimally resourced digitisation effort, as well as the resulting dataset's value for archaeological research and cultural heritage management. Digitisation was undertaken as a secondary activity on a landscape archaeology project focusing on pedestrian feature survey. Undergraduates in the associated field school digitised data from maps using a system repurposed from other project activities. Compared to manual digitisation approaches based on desktop GIS, it required little training or supervision of students, used open-source software and low-cost equipment, yet produced a large, accurate, analysis-ready dataset. It complements Machine Learning

(ML) and other automated approaches in that it requires less technical expertise, time, and resourcing to undertake. Such an approach is suitable for projects working with small to mid-sized data sources (100s–10,000s of features) that do not warrant the investment needed for successful ML-based data extraction – as well as for exploratory work preceding automated analyses or the production of the training and quality assurance datasets needed for ML. The approach taken here can be replicated using other mobile GIS systems, scaled up, or applied to other types of archaeological features. As such, it shows promise as a low-cost means of extracting locations and other data from historical maps concerning endangered and poorly documented material remains.

#### 1.1. The Tundzha Regional Archaeology Project

The Tundzha Regional Archaeological Project (TRAP), is an investigation of the middle (Yambol region) and upper (Kazanlak Valley)

\* Corresponding author.

E-mail addresses: [adela@cas.au.dk](mailto:adela@cas.au.dk) (A. Sobotkova), [shawn.ross@mq.edu.au](mailto:shawn.ross@mq.edu.au) (S.A. Ross), [ch.na.ha@gmail.com](mailto:ch.na.ha@gmail.com) (C. Nassif-Haynes), [brian.ballsun-stanton@mq.edu.au](mailto:brian.ballsun-stanton@mq.edu.au) (B. Ballsun-Stanton).

Tundzha River watershed in Bulgaria. It has explored long-term cultural development in its environmental context since 2008 (Ross et al., 2010, 2018; Sobotkova, 2013). Methodologically, it combines historical research, archaeological survey, remote sensing, excavation, and ecological sampling. The aims of the project include reconstructing the ancient environment, mapping the evolution of habitation, and combining those outcomes to explain long-term, human-environment interactions (Connor et al., 2013; Sobotkova, 2018b, [a] 2018; Bishop-Taylor et al., 2018). It also seeks to produce an inventory of archaeological heritage for research and conservation purposes, including analyses of threats (Eftimoski et al., 2017; Sobotkova & Ross, 2018). Between 2008 and 2016, TRAP catalogued some 773 mounds in the Kazanlak Valley and 431 mounds in the Yambol region using pedestrian surface survey supported by manual digitisation of satellite imagery and maps.

During 2017 and 2018, TRAP recorded the location, characteristics, and condition of burial mounds in the Yambol region of Bulgaria. Three activities were undertaken: (1) visiting known burial mounds, registering their location and condition; (2) identifying changes in mound condition using satellite imagery; (3) digitising mounds from over 20,000 sq km of Soviet military 1:50,000 topographic maps covering southeast Bulgaria, followed by ground-truthing (which continued through 2022). All three approaches involved digital field data collection workflows, improving research transparency and facilitating production of Findable, Accessible, Interoperable, and Reproducible (FAIR) and analysis-ready datasets. The results of mound registration and monitoring have been presented elsewhere (Valchev & Sobotkova, 2019; Sobotkova & Weissova, 2020). This paper discusses the digitisation of mound symbols from maps using a crowdsourcing approach involving undergraduate students associated with the project.

## 1.2. Burial mounds in Bulgarian archaeology

An estimated 50,000 burial mounds were built in Bulgarian lands from the Early Bronze Age through the Middle Ages (Shkorpil & Shkorpil, 1989, p. 20; Kitov, 1993, p. 42). They were constructed of earth and rubble, and range in size from 10 to 50 m in diameter and 0.5–20 m in height (see Fig. 1). Some contain modest burials (or nothing at all), while others enclose elaborate brick or stone structures and valuable grave goods. The mounds and their contents are important archaeologically as they attest to demographics, social complexity, international connections, and other important historical questions. They are also culturally significant, as attested by several UNESCO World Heritage listings (UNESCO World Heritage Centre 1979, 2016), and drive heritage tourism in rural areas of a middle-income country.

Because they may contain intrinsically valuable grave goods,

mounds are at risk of looting (Loulanski & Loulanski, 2017). Others are destroyed in the course of development, sometimes preceded by rescue excavations (Cholakov & Chukalev, 2008). Still others are degraded by agricultural activities, particularly as pasturelands are repurposed to arable farming (Eftimoski et al., 2017).

Burial mounds are an irreplaceable - but endangered - aspect of Bulgarian cultural heritage, making their systematic recording and registration an urgent undertaking for both research and cultural heritage management. To this end, the Bulgarian Academy of Sciences maintains the Archaeological Map of Bulgaria, a legally established, national archaeological information system underpinning site protection (Kecheva, 2019). The ability to record mounds that no longer exist, but are represented in historical maps, is especially important as evidence of past landscapes and to assess ongoing damage to mounds.

## 1.3. Extracting data from historical maps

Historical maps constitute an important source of information for historical landscapes and the cultural heritage they contain. They predate aerial photography (1920s) and satellite imagery (1960s). Prior to the age of modern commercial satellite imagery (1990s), they provided the most comprehensive coverage. Historical maps help heritage specialists analyse change over time in order to monitor and mitigate the impact of urban development, climate change, biodiversity loss, erosion, agricultural extensification, and other kinds of land use and environmental changes (Rondelli, Stride, and García-Granero 2013; Petrie et al., 2018).

The value of such maps has, however, been underexploited due to the entailed process of manual digitisation needed to transform their analogue information into machine-readable data. A historian wishing to create digital spatial data from scanned and georeferenced historical maps has a range of digitisation options, from employing or training specialists to digitise using desktop GIS, through digitisation using volunteers (either with desktop GIS or a customised system), to deploying an automated approach such as ML.

Manually drawing and annotating shapes in historical maps using a desktop GIS is time-consuming and requires specialised skills (Can, Gerrits, and Kabadayi 2021; Petrie et al., 2018; Jones & Weber, 2012). Its principal limitations are difficulty of scaling the effort, particularly restricted time and availability of expert users to either undertake digitisation themselves, or to train and support novices to the extent required for efficient and accurate work (see also Jessop, 2007 for other challenges).

In recent years, the development of automated approaches, especially ML, has opened new possibilities for producing machine-readable spatial data from a variety of data sources (Groom et al., 2021; Ekim,



**Fig. 1.** Excavated mound profile at Maglizh showing a robber's trench inside. Credit: Simon Connor @ 2009

Sertel, and Kabadayı 2021; Saeedimoghaddam & Stepinski, 2020; Caspari, 2020; Lambers et al., 2019; Caspari & Crespo, 2019; Cowley, 2012), including historical maps (Groom et al., 2021; Ekim, Sertel, and Kabadayı 2021; Can & Kabadayı, 2021). ML has unmatched potential for large-scale data digitisation, but it requires specific programming expertise, and familiarity with the capabilities and limitations of ML. Naive use of ML is likely to produce biased or otherwise unreliable results (Mehrabi et al., 2021; Schwemmer et al., 2020; Besse et al., 2018; Fuchs, 2018; Haas, 2017), even if pretrained image recognition models are used (Jiang et al., 2022). Correctives have been developed to address these problems, but their use requires experience and expertise (Bahng et al., 2020; Wang et al., 2020; Sengupta et al., 2018). This technical expertise is in high demand, and may be difficult to attract to the smaller humanities and social sciences (HASS) projects that undertake much of this digitisation. Training an algorithm, moreover, requires manual creation and preparation of training data and manual quality assurance (Bennett et al., 2014; Can & Kabadayı, 2021).

Crowdsourcing can expand the scope of map digitisation. In this context, crowdsourcing indicates the voluntary creation of geographic information by amateurs ('Volunteered Geographic Information'), using software that exposes a subset of GIS capabilities in a manner intuitive enough to be utilised by minimally trained actors (Goodchild, 2007, pp. 212–214; see also Owen et al., 2009, pp. 16–19). The promise and goal of this activity is to unlock analogue information contained in historical maps, improving the reliability and completeness of geospatial datasets used by cultural heritage workers and researchers in circumstances where these specialists do not have the time or resources to digitise the information themselves (analogous to the grassroots production of data discussed in Elwood, 2008a, pp. 82–84). In related applications, crowdsourcing has been used to create machine-readable data from images of texts (see Sturgeon, 2022 for a recent example). It has also been used to digitise heritage information from satellite imagery and other remotely sensed data (Duckers, 2013; Lin et al., 2014). Its application to historical maps has focused on the transcriptions and annotation of place names (Pödör, 2015; Simon et al., 2015; Vitale et al., 2021), but crowdsourcing has only rarely been used to extract the location of archaeological features from symbols on historical maps. Such efforts have been limited to projects that have the expertise to customise general-purpose GISes for this activity (e.g. Koski et al., 2021; Kendall, Hedenstroem, and Hedenstroem 2021) or the resources to construct purpose-built platforms (e.g. the Pleiades, see Simon et al., 2015; Vitale et al., 2021).

This paper argues that crowdsourcing offers advantages compared to alternative approaches under many, if not most, map digitisation scenarios. Crowdsourcing requires more upfront investment in system setup than using available GIS specialists or training and supervising volunteers using desktop GIS, but it scales better in the face of likely constraints related to expert staffing and volunteer attrition. Compared to ML, it requires less specialised and hard-to-come-by expertise, requires less time for initial setup, and produces high-quality datasets with more predictable errors. Since choosing between these approaches depends largely on the size of the dataset being digitised, we suggest payoff thresholds for each alongside qualitative assessments.

#### 1.4. Sociotechnical barriers to collaborative map digitisation

The expertise necessary to use a GIS has hindered collaboration between specialists and the broader community in the past (Jessop, 2007, p. 43). Producing and manipulating geospatial data, even if it is limited to 'simple' activities like symbol digitisation (Goodchild, 2007), challenges novice users more than other data tasks. Structured data, for example, can be manipulated using familiar desktop software (typically spreadsheets). Working with raster images from technical photography, and to a lesser extent vector graphics from drawings, builds on skills that many people acquire from everyday activities involving consumer desktop, web, or mobile applications. A novice wishing to contribute to

geospatial data creation from historical maps, however, needs all three of these general technical skills, plus specific expertise related to cartographic concepts and their application via specialised software – not to mention grounding in the research discipline itself.

Volunteers often lack the skills necessary to use GIS software (Elwood, 2008b; Jones & Weber, 2012; Owen et al., 2009). Often, their experience is limited to 'dropping pins' on web 'slippy-maps' like Google Maps. These tools have made consumption and perhaps simple annotation of geospatial data easier, but obfuscate the complexities of producing data. This 'digital divide' can be bridged in part through more expert training and guidance of novice users (Owen et al., 2009, p. 24). Such an approach scales poorly, however; the supply of expert time available to any project is restricted and often oversubscribed, while the willingness of volunteers to learn specialised skills and software has limits.

While some expert interaction is unavoidable and desirable, it can be profitably supplemented by the development of 'useful' tools that combine 'utility' with 'usability' (Nielsen, 2012). Utility indicates the ability of the system to accomplish the task at hand, in this case producing standardised, interoperable, well-documented digital data from scanned historical maps. Usability includes the degree to which a system's User Interface (UI) allows first time users to accomplish basic tasks, helps more experienced users to complete tasks quickly, eases resumption of work after a period of disuse, is pleasant for users to employ, and mitigates the rate and severity of errors ('learnability, efficiency, memorability, satisfaction, and accuracy'; Jones & Weber, 2012, 530–531; Nielsen, 2012). Heuristics based on these principles aim to limit the cognitive load on users, and therefore encourage UIs that are focused on essential tasks, expose key functionality, present controls clearly and consistently, provide adequate feedback to users, and limit the impact of errors (Jones & Weber, 2012; Kaplan, 2021; Nielsen, 2020). These and similar usability approaches have produced a range of streamlined tools with intuitive UIs for participatory or collaborative geospatial research in other settings (e.g. Koski et al., 2021; Kendall, Hedenstroem, and Hedenstroem 2021; Vitale et al., 2021; Simon et al., 2015; Pödör, 2015; see also Owen et al., 2009, pp. 16–18; Elwood, 2008a, pp. 84–85; Goodchild, 2007, p. 218).

Mobile applications may be particularly suited to the production of Volunteered Geographic Information (VGI). Most potential users are familiar with applications like Google Maps that utilise interactions via touch screen and integrated sensors like GNSS receivers and cameras. Zooniverse, a popular platform for large-scale volunteer research participation, observes that 'mobile-enabled projects receive more classifications' than others ('About: Zooniverse Mobile App' 2023). The challenges of data capture during fieldwork have, moreover, prompted a three-decade literature on Human-Computer Interaction (HCI) that includes collaborative production of geospatial data. Approaches from this community of practice complement foundational usability principles, and can inform deskbound approaches to producing VGI. UIs for mobile data collection systems must allow the user to (1) focus on observations while minimising interactions with the recording mechanism, (2), enter large amounts of data quickly and accurately, with appropriate automation and validation, and (3) aid recording of data context such as metadata or related data (Sobotkova et al., 2021; Wagtendonk & De Jeu, 2007; Pascoe, Ryan and Morse 1998, 2000; Ryan et al., 1999; Cheverst et al., 1998; Johnson et al., 1998). In short, such systems should be unobtrusive, conform to research workflows rather than forcing users to compromise, and automate and validate entered (meta)data aggressively - all attributes that help novice users as they begin to digitise historical maps.

## 2. Approach

For the 2017–2018 field seasons, TRAP staff created a simplified and streamlined data capture system built using the FAIMS Mobile platform. This system allowed any number of participants to digitise map features

using mobile devices, regardless of network connectivity, and consolidated the resulting data when a network became available. Although use was limited to students affiliated with the project, both the approach and the technologies applied can scale to accommodate large numbers of contributors. A similar approach could also be developed using other platforms.

### 2.1. Archaeological features in Soviet topographic maps

The goal of the work was to extract archaeological features from 1:50,000 scale Soviet military topographic maps dating to the 1980s. Available as georeferenced GeoTIFFs (from [http://web.uni-plovdiv.bg/vedrin/index\\_en.html](http://web.uni-plovdiv.bg/vedrin/index_en.html) in 2008), each of these maps covers ca 400 sq km. We wished to extract symbols from them that might represent burial or settlement mounds in our study area (see Fig. 2). Such symbols occurred at a high density, averaging about 200 per tile (0.5 per sq km), with counts per tile ranging from about 50 to 400. The mound symbols were moderately obtrusive; some aspects of shape or colour were shared with other map symbols. The records we sought to create were relatively simple: a point for the feature, a record number, plus ten attributes. We knew that the maps of the Yambol region contained 1,000+ of these targeted symbols.

### 2.2. Crowdsourcing digitisation with field-school participants

The task of digitising potentially thousands of mounds provided an opportunity to involve students in authentic research. Our students came from a range of academic backgrounds in Arts and Humanities. Most had no training in archaeology, cartography, or digital methods (unlike Pödör, 2015 or Can et al., 2021). The students' motivation in joining our project included curiosity about field archaeology, the desire to travel outside Australia, and the satisfaction of assisting with heritage preservation.

TRAP staff never had the time to undertake large-scale map digitisation work themselves. To overcome this problem, in 2010, project staff worked with student volunteers to digitise map features using ArcGIS. Our experience was much like that of other projects: novice volunteers found learning to configure and navigate desktop GIS challenging; many quit and those who continued required ongoing support. In the end,

volunteer attrition combined with demands on staff time during the height of fieldwork rendered this approach unsuccessful.

In 2017, faced with a short field season and little time for student training, we focused on implementing tools that would empower volunteers to digitise maps independently. Our approach sought to help novice users begin digitising quickly and then work productively for the duration of each field season with minimal support or frustration. As such, it stripped GIS functionality to its essentials, focusing on three tasks: layer selection, shape digitisation, and annotation, with validation and automation to improve data quality. Geospatial data preparation and management was relegated to staff, while a simple and intuitive UI allowed students to begin digitising after only minutes of training. This approach was refined during the 2018 field season. The outcomes of both field seasons are discussed in this paper.

### 2.3. Using a mobile application for map digitisation

Having decided to adopt a crowdsourced approach to produce VGI, we chose to customise FAIMS Mobile for map digitisation. The history of the FAIMS Project and the features of FAIMS Mobile have been described elsewhere (Sobotkova et al., 2021; Ballsun-Stanton et al., 2018; Ross et al., 2013). Briefly, FAIMS Mobile is a server-client platform that generates customised Android applications for data collection during offline field research. Customisation is accomplished via definition files that can be shared, modified, and redeployed. FAIMS Mobile interprets the definition files to generate a specific data capture application. It can collect, manage, and bind spatial, structured, multimedia, and text data as part of a single record, obviating the need to use multiple applications. Data collection works offline, and can employ as many devices as necessary. It is later synchronised opportunistically, when a network is available. Data can be validated on devices at the time of capture, or on the server after synchronisation. The server can also be used to edit data, view data history, selectively revert data to earlier states, and export data in a variety of formats. The system is designed for field research, with the goal of requiring as few compromises from researchers or fieldworkers as possible. FAIMS Mobile had previously been used with some success for a student-led ceramics recording project (VanValkenburgh et al., 2018), and improved recording efficiency in other settings (Sobotkova et al., 2016).

The decision to use mobile software, and FAIMS Mobile in particular, was based on several factors. First, FAIMS Mobile worked offline. Our digitisation took place alongside fieldwork, at field bases in rural Bulgaria. Reliable internet connectivity could not be guaranteed under these circumstances; a system that tolerated degraded network connectivity was required. In our experience, digitisation work associated with archaeological research often proceeds under such conditions. Second, this system met the functional requirements we identified for geospatial software. It supported the production of a customised map digitisation system with a simple UI and streamlined workflow, while still providing essential features including layer management, geometry creation and editing, capture and association of structured data, import and use of arbitrary rasters (scanned maps as geotiffs), automated metadata creation, and data validation. Third, it allowed us to test the idea that usability approaches from data capture during kinetic fieldwork were beneficially transferable to digitisation work. Fourth, we were already using FAIMS Mobile for in-field legacy data verification, the project's main activity. Reusing the platform for digitisation offered a consistent working environment for users, reduced administrative load on staff, leveraged our experience with the platform, and avoided any additional hardware or software costs. Fifth, student volunteers are accustomed to, and even prefer, 'slippy-map', touch-screen interfaces on mobile devices over the point-and-click, desktop UI idiom. We believed that the use of the former would make it easier for students to learn the system, and more likely to stick with digitisation. This choice also reduced competition for the limited number of computers, ESRI licences, and desk space available in the field, plus it allowed students to use their

Burial mound	
Settlement mound	
Triangulation point	
Triangulation point on a burial mound	
Bench mark	
Bench mark (on a burial mound)	

Fig. 2. Map symbols potentially indicating burial or settlement mounds in Soviet topographic maps. Credit: Véra Doležálková @ 2017

own devices (only two of 12 students brought computers, and none brought mice, but all had mobile devices). Finally, we preferred the use of open-source software customised via code for its redeployability and transparency advantages.

No existing system met these requirements ‘off-the-shelf’, without significant customisation, and no competing product offered enough of an advantage to justify adopting a second geospatial recording system. The primary alternatives assessed, ESRI ArcGIS Collector (now succeeded by Field Maps) and Survey123, had a number of disadvantages, particularly the need to divide map-centric and form-centric recording between the two applications and reconcile the data later, the need for internet connectivity at crucial stages of project setup, and the proprietary nature of the system (Sobotkova et al., 2021, p. 20).

Digitisation of scanned maps deviated from the ‘core’ FAIMS Mobile use-case of field data collection, so it served both as a test of the platform’s adaptability, and as an exploration of the potential benefits offered by repurposing mobile applications for field data capture to digitisation projects.

#### 2.4. Design and implementation of the recording system

The stages of FAIMS Mobile implementation (Fig. 3) included: (1) project staff modelled the data and workflow to ensure that the final dataset met research needs, (2) a junior software developer worked with

project staff to customise the system, (3) project staff defined a spatial reference system (SRS) and imported preprocessed historical maps, (4) volunteers drew a shape (usually a point) wherever they saw a target symbol and (5) volunteers transcribed attributes from the map, (6) project staff exported data using the FAIMS Mobile server, (7) project staff undertook a targeted accuracy-checking exercise. This approach moved activities requiring technical expertise to phases where specialists could contribute, while simplifying the tasks assigned to student volunteers as much as possible.

To support this workflow and make work easier for both project staff and participants, FAIMS Mobile automated a number of tasks and provided necessary capabilities. It applied the spatial reference system, rendered maps in the workspace, provided layer management (including a data entry layer), enforced shape topology, displayed pre-defined controlled vocabularies for attribute terms, recorded creation time and author for each record, maintained a history of all changes to data, applied validation to ensure record completeness, merged data from multiple devices, and exported data in common formats. The digitisation interface itself was as streamlined as possible (see Figs. 4 and 5). Volunteers could toggle between a map view for geospatial data interactions and a form view for attribute creation and editing. In the map, they could adjust layer focus and visibility, pan, and zoom. Existing records could be searched, retrieved, inspected, and edited.

Since project staff set up the infrastructure and pre-processed and

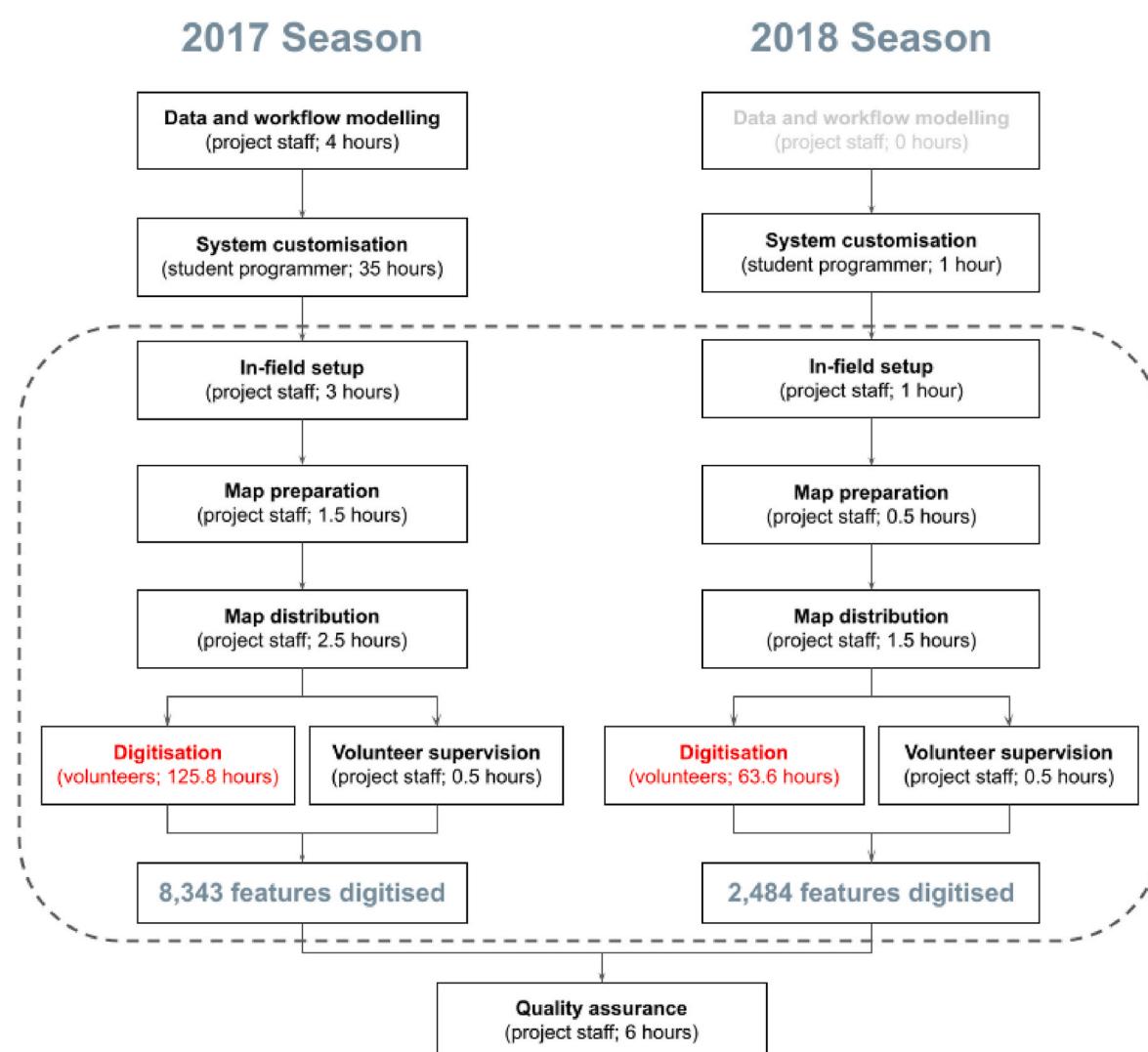
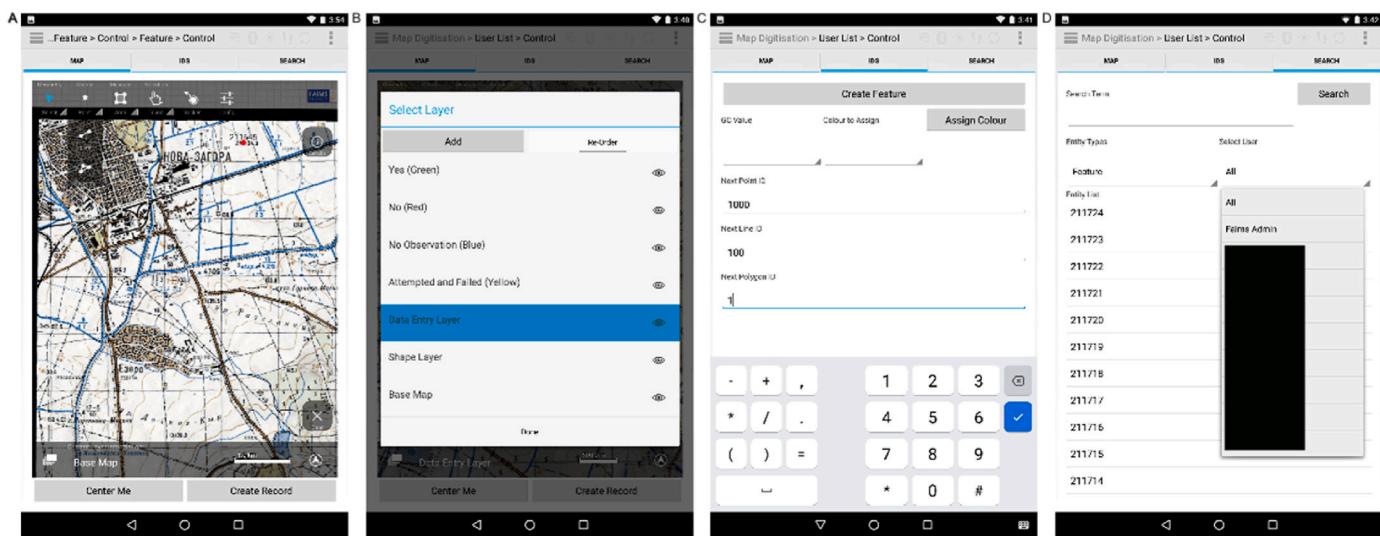
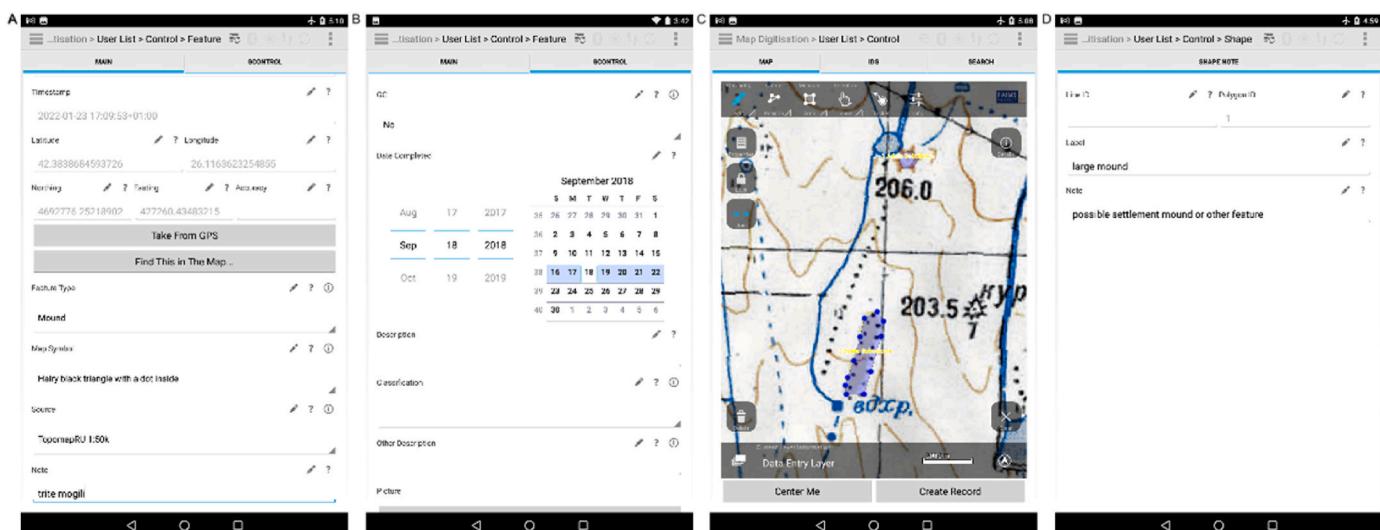


Fig. 3. Crowdsourcing approach workflow diagram. Student volunteer activities highlighted red. Dashed box indicates activities taking place in the field.



**Fig. 4.** Control Tabgroup from left to right: (A) Map Tab with shape creation tools, base map, and a captured point (red dot near top centre); (B) Map Layer management; (C) Identifier tab with seed values and symbol colour assignment; (D) Search tab.



**Fig. 5.** Feature and Shape tabgroups: (A) Main tab of the Feature form with inherited (light grey) and controlled vocabulary data entry fields, and other controls; (B) Ground Control tab with a date picker, free text and controlled vocabulary data entry, and photo capture; (C) Map tab showing polygon creation; (D) Shape Note tab for attributes of lines and polygons created in the Map tab.

loaded the required maps, volunteers were insulated from the friction of setup, layer management, data aggregation, export, and backup. GIS features not needed for digitisation were hidden or eliminated. Digitisation and metadata creation required no GIS or computing skills. Students capable of selecting files from a list, panning and zooming a map, dropping a point, and filling out a form were able to create data. Only a few important controls, including layer management, map navigation, record search and retrieval, and shape and attribute creation and editing, were present. As a result, users required almost no training and could focus on the act of digitisation without being distracted by the technology used to accomplish it (Pascoe et al., 2000).

Exported data was consistent and complete, ready for analysis with minimal cleaning. This data adhered to key elements of the FAIR data principles, especially the production of 'rich' and 'plural' metadata at the time of data creation (principles F2, R1.1–1.3; GO-FAIR, 2017).

The code defining this customisation, along with a description and user guide, can be found on GitHub (<https://github.com/FAIMS/map-digitisation/releases/tag/map-dig-2018>).

## 2.5. Evaluating the digitisation approach

The success of this approach became apparent early in the 2017 field season. At that point, we decided to catalogue inputs (time invested by staff and volunteers) versus outputs (features digitised) as part of a research program to evaluate digital approaches to fieldwork (e.g., Sobotkova et al., 2016). To measure inputs, we collated the amount of time spent by various participants in the process, including the student programmer who instantiated the customisation, the student volunteers who undertook the digitisation, and project staff who configured the system, supported volunteers, exported data, and checked for errors. Project records provided much of this data (timesheets from the programmer; record creation timestamps for students using the system), while project staff logged time-on-task for activities in journals. We took the number of features digitised as the output. Additional care was taken in the subsequent 2018 season to ensure these records were as complete as possible, and confirm 2017 time estimates for system set-up, administration, and support in the field. Finally, project staff reviewed

randomly selected digitisation work completed by volunteers to characterise errors.

### 3. Results

#### 3.1. Project staff time for setup, support, and accuracy-checking

For the first season of use (2017), creating the Map Digitisation customisation of FAIMS Mobile required 35 h from an undergraduate student programmer plus 4 h from staff (Nassif-Haynes et al., 2021). Setup of the server and configuration of the client devices in the field required 3 h from staff. Map preparation (tiling, adding pyramids) required about 1.5 h. Monitoring file compression, upload to the server, and download to devices took an additional 2.5 h. Training and supervision of students took no more than half an hour of staff time across the entire season.

For the second season, adding additional validation to ensure population of latitude and longitude from GPS (see ‘Recoverable data omissions and incomplete records’ below) took 1 h of development from the programmer. In-field set-up required an hour (reusing the same equipment and system as the previous year), map preparation 30 min, file preparation and transfer 1.5 h, and student training and supervision another 30 min.

Across both seasons, customisation, setup, and supervision took about 51 h, including 36 h from the programmer and 15 from project staff. Of this time, initial customisation and setup time before fieldwork was 44 h, while the time required during fieldwork to prepare and distribute maps, and then supervise participants, was 7 h. Finally, re-examination of four randomly selected maps after fieldwork required 6 h of staff time, including desktop GIS setup, confirmation of feature digitisation, and tabulating errors and error rates.

#### 3.2. Student-volunteer digitisation velocity and volume

Fieldwork participants used the Map Digitisation customisation in 2017 and 2018. In 2017, it was used for a total of 125.8 person-hours concentrated across five rainy days, during which time 8,343 features were digitised from 42 Soviet topographic maps (ca. 17,000 sq km). The average time to record a point feature was 54 s, based on start and end times of feature creation as recorded by the devices (representing work time excluding pauses between records).

In 2018, use was more sporadic; participants who stayed at the base for any reason sometimes undertook digitisation. The system was used for 63.6 person-hours, with 2,484 features recorded from 16 maps (ca. 6,500 sq km), an average rate of one record every 92 s.

In total, 10,827 point features, mostly burial and settlement mounds, were recorded in 189.4 student-hours (63 s per record). Fifty-eight map tiles representing about 23,500 sq km were digitised. Since a single point record could be associated with 11 attribute-value pairs, the dataset contained as many as 119,097 discrete values (many captured automatically). The concentrated digitisation in 2017 was more productive than the intermittent work of 2018, but both seasons yielded large and valuable datasets utilising time that might otherwise have been lost (e.g., to inclement weather), while requiring little supervision by project staff.

#### 3.3. Digitisation comparison with desktop GIS

As noted above, TRAP had attempted an unsatisfactory digitisation effort in 2010 by students using ArcGIS. Although we did not maintain detailed volunteer time-on-task records, we know this effort produced a dataset of 915 features and required about 5–7 h of staff training, support, and error-checking over a three-week period (based on our field journals). The need for high-touch training and support by project staff during fieldwork, combined with a dislike for the activity on the part of students, prevented us from scaling up the use of desktop GIS for

volunteer digitisation further, despite attempting to do so. Most volunteers did not persevere with the activity long enough to repay their initial training time. Indeed, one persistent student accounted for almost all the digitised features; without his perseverance, the digitisation effort would have failed entirely. These problems reflect the high cognitive load desktop GIS places on novice users.

By contrast, our customised application met fundamental usability requirements (e.g., Nielsen, 2012), both due to careful design of the customisation itself, and the underlying platform’s implementation of Google’s Material Design guidelines. It also benefited from the UI/UX approach employed for kinetic fieldwork (e.g., Pascoe et al., 2000). The principle that the technology had to conform to the workflow, rather than vice versa, translated particularly well to map digitisation. The result was a simple, familiar mobile application interface that let novices begin work with little training and helped them resume work after any hiatus. Use of controlled vocabularies, automation, and validation reduced errors. Low volunteer attrition indicated satisfaction with the experience. Volunteers could attain a high rate of digitisation quickly and maintain it, although further design refinement could improve the ability of more experienced users to enter data even more quickly (for an example of an optimised system in another domain, see Noble et al., 2020, 2018).

#### 3.4. Application performance

FAIMS Mobile supports automated performance testing (Sobotkova et al., 2021). Such testing, however, works best for structured data input rather than map-driven recording. Automated testing of other customisations suggested that performance would degrade once approximately 3,000–6,000 records had been created. In use, automated extraction of coordinates from GPS into the Latitude/Longitude and Northing/Easting fields, which took three to 5 s with an empty database, took as long as 30 s once a device exceeded about 2,500 records. Deteriorating performance was mitigated by exporting all data and instantiating a new and empty version of the application. Since data structures were identical, aggregation of multiple exports was trivial.

#### 3.5. Data quality

Reports from on-device validation, as well as quality assurance by project leaders, suggest that digitisation accuracy was good. The best digitisers, furthermore, could be both fast and accurate, while the poorest digitisers were often neither fast nor accurate (see Tables 1–3).

##### 3.5.1. Recoverable data omissions and incomplete records

Recoverable data omissions across both years totaled 223 (2.06% of records), including 205 spatial and 18 attribute omissions. Most occurred in 2017 when 192 records (2.3%) had empty latitude and longitude fields and 17 (0.2%) were missing specification of the map symbol. Spatial data omissions resulted from a failure of the software to populate the latitude and longitude fields from the application’s SpatiaLite geodatabase due to users moving through the forms too quickly (see ‘Application performance’ above). Before the 2018 season, we added validation addressing this problem, resulting in only 13 spatial errors and one attribute omission (0.52%). Since the geodatabase preserved geometries, spatial omissions were corrected by re-extracting latitude and longitude; only two data points could not be recovered. Although fixable, these omissions delayed visualisation and analysis of the data.

##### 3.5.2. Digitisation errors

Unlike some volunteer digitisation projects, overall accuracy was high, over 94% for processed maps. First, participants failed to digitise some assigned maps, leaving noticeable gaps (see Fig. 6). This omission was obvious, and could be corrected in a later digitising session. Second, a review by project staff of four randomly selected maps (7% of the total)

**Table 1**

2017 digitisation statistics, including time-on-task, digitisation rate, and omissions. Compare [Table 3](#) for digitisation error rates.

2017	Hours digitising	Features recorded	Time per feature (seconds)	Missing Lat-Long	Map symbol omissions	Cumulative omission rate
Student A	28	2302	44	0	1	0.04%
Student B	22.3	1799	45	36	4	2.2%
Student C	27.3	1615	61	2	9	0.68%
Student D	30.8	1526	73	136	2	9.04%
Student E	17.1	1090	57	18	1	1.74%
Tester (staff)	0.3	11	20	–	–	–
<b>Overall</b>	<b>125.8</b>	<b>8343</b>	<b>54</b>	<b>192</b>	<b>17</b>	<b>2.51%</b>

**Table 2**

2018 student digitisation statistics, including time-on-task, digitisation rate, and omissions. Compare [Table 3](#) for digitisation error rates.

2018	Hours digitising	Features recorded	Time per feature (seconds)	Missing Lat-Long	Map symbol omissions	Cumulative omission rate
Student F	29.1	1305	80	3	0	0.23%
Student G	25.8	810	115	10	1	1.36%
Student H	8	335	86	0	0	0%
Student I	0.4	13	103	0	0	0%
Tester (staff)	0.3	21	60	–	–	–
<b>Overall</b>	<b>63.6</b>	<b>2484</b>	<b>92</b>	<b>13</b>	<b>1</b>	<b>0.56%</b>

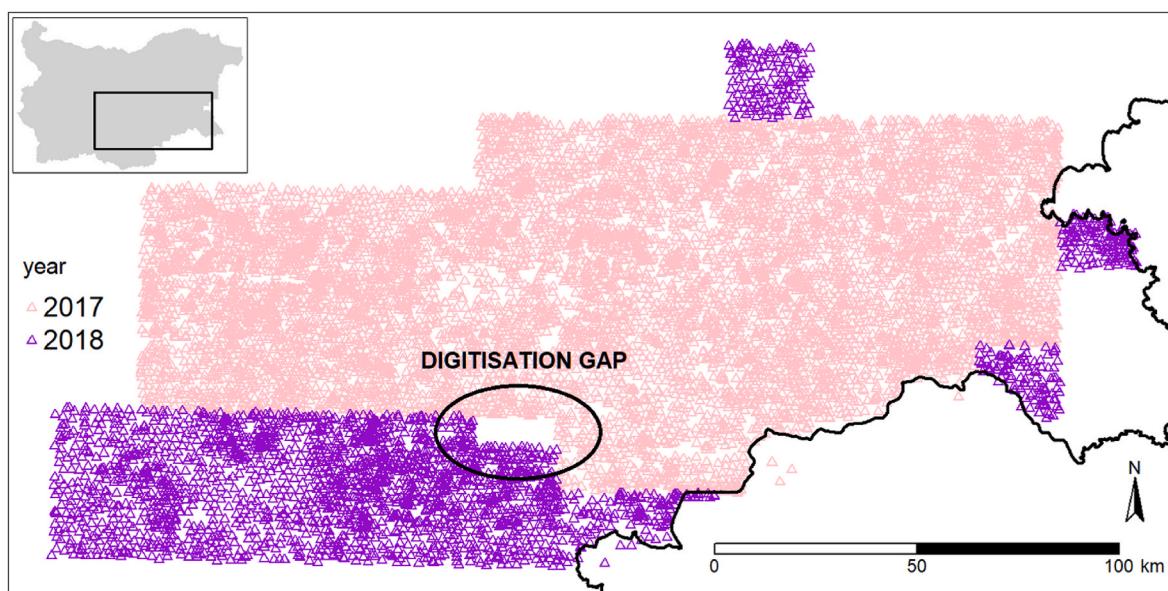
**Table 3**

Digitisation errors (A) and error rates (B) by student. These error rates are calculated from a quality assurance check completed by project staff. Student C's high false negative rate arises from three sizable, contiguous map sections the student omitted. Compare [Table 1](#): 2017 Digitisation statistics.

A						
Volunteer	Features identified	False positive	Double- marked	False negative	Classification error	Total errors
Student A	227	0	0	3	0	3
Student B	203	0	0	6	0	6
Student C	305	0	6	29	0	35
Student D	64	1	0	4	0	5
<b>Cumulative</b>	<b>799</b>	<b>1</b>	<b>6</b>	<b>42</b>	<b>0</b>	<b>49</b>

B						
Volunteer	True positives	True feature count	False positive rate	Double- marking rate	False negative rate	Total error rate
Student A	227	230	0.0%	0.0%	1.3%	1.3%
Student B	203	209	0.0%	0.0%	2.9%	2.9%
Student C	299	328	0.0%	2.0%	8.7%	10.6%
Student D	63	67	1.6%	0.0%	5.9%	7.4%
<b>Cumulative</b>	<b>792</b>	<b>834</b>	<b>0.1%</b>	<b>0.8%</b>	<b>5.0%</b>	<b>5.9%</b>



**Fig. 6.** Features digitised in 2017 (pink) and 2018 (purple). Internal white area indicates unfinished map tiles.

found 49 errors from a true count of 834 features, a 5.87% error rate (see Table 3). Forty-two of these errors were false negatives (symbols missed by students). Six were double-marked (Student C digitised a section of a map twice). Students made only one classification error (a similar symbol mistaken for a benchmark), and no outright false positives. Students' individual error rates ranged from 1.3% to 10.6%. Note that the two fastest digitisers (Students A and B; 44 and 45 s per feature respectively) also had the lowest error rates (1.3 and 2.9%), while the two slowest (Students C and D; 61 and 73 s) had the highest error rates (10.6 and 7.4%). Moreover, 35 of the 49 false negatives were the result of Student C failing to digitise three contiguous sections of an assigned map. These mistakes made his error rate of 10.6% an outlier; excluding Student C would have cut the cumulative error rate in half to 2.8%. Nevertheless, the overall rate of 5.9% exceeded our expectations.

Moreover, the pattern of errors - mostly false negatives and double-marked features, mostly from contiguous map sections - made them relatively easy to identify and correct. Simple expedients, such as assigning multiple students to digitise the same map tiles independently or assigning one student to review work by another, would likely eliminate most errors. Even using staff time, it was much faster to check volunteer work than digitise from scratch.

#### 4. Discussion

Our crowdsourced digitisation effort involving novice volunteers using an adapted mobile application for data capture proved unexpectedly successful. It was only an auxiliary activity undertaken on a time-available basis, intentionally secondary to pedestrian survey. While volunteer-based geographic digitisation projects often enjoy high productivity at low-cost (Goodchild, 2007; Simon et al., 2015), they may face data-quality problems and require continuous access to a network (Pödör, 2015; Budig et al., 2016; Lin et al., 2014). Our approach was done under field conditions, with inexpensive equipment and limited internet connectivity, yet produced a large (>10,000 features), high-quality (<6% error rate) dataset while placing reasonable demands on both volunteers and staff compared to other approaches.

##### 4.1. Choosing an approach

Digitisation projects will likely choose between one of four principal approaches to digitising historical maps.

1. Have expert staff or specialists digitise features using a desktop GIS;
2. Employ volunteers to digitise features using a desktop GIS with training and support from expert staff;
3. Customise and deploy a collaborative geospatial system that can be used by volunteers, including novices with little GIS experience, backstopped by training and support from expert staff;
4. Invest in a ML approach to automatically extract features, requiring extensive collaboration between project staff and external specialists to train and evaluate a model.

These approaches fall along a continuum from the first, which requires the least setup cost, time, and technical support, but the most ongoing expert involvement, to the last, which requires the greatest

setup cost, time, and technical input but the least ongoing expert involvement (Fig. 7). Choosing amongst the approaches involves weighing both qualitative and quantitative considerations inherent to each.

Note that in all following comparisons, we present our experience as a (perhaps idiosyncratic) example. Time per feature included locating it and completing the record, hence the rate reflects the high density and moderate obtrusiveness of archaeological features in Soviet topographic maps, combined with the relative simplicity (and, ultimately, automation) of our digital recording forms. Ranges indicate the limits of our earlier time-on-task recordkeeping. This discussion, furthermore, focuses on our most limited resource: staff time. As such, the calculations below, which propose dataset size thresholds for various approaches, prioritise staff time required for digitisation.

##### 4.1.1. Desktop GIS approaches versus crowdsourcing

After brief workspace setup, project staff with desktop GIS experience could digitise at a sustained rate of 60–75 features per staff-hour. At this rate, the 57 h of staff time devoted to set-up, support, and quality assurance for our crowdsourcing system could have resulted in some 3,420–4,275 staff-digitised features (see Table 4). Such a payoff threshold suggests that digitisation by project staff will be suitable only for smaller datasets. We could not have afforded to dedicate 3.5–4.5 weeks of staff time to digitisation.

Had specialist project staff instead trained and supervised volunteers to use desktop GIS for digitisation, based on our 2010 digitisation rate of 130–180 features per staff-hour, 57 h might have produced 7,410–10,260 features. At the highest rate, desktop GIS digitisation using novice volunteers is almost competitive with the mobile application approach we used. Scaling to this dataset size, however, assumes that enough volunteers could be retained to complete the work - something we were unable to do in 2010.

By comparison, the 57 h of staff time required for our digitisation approach using a customisation of FAIMS Mobile produced 10,827 features, or about 190 features per staff-hour. This figure, however, understates the value our project realised from volunteer digitisation.

First, customisation of systems like FAIMS Mobile can be outsourced more easily than other project activities. Only 21 of the 57 h needed to support the system came from project staff, while the other 36 h were

**Table 4**

Estimates for payoff thresholds for our crowdsourcing approach versus desktop GIS alternatives, in terms of features digitised using staff time as the criterion. The 'Low estimate' considers only in-field staff time, the 'Mid estimate' includes only internal staff time (excluding the student programmer hired to implement the customisation), while the 'High estimate' includes all staff time for all stages of customisation, deployment, support, and quality assurance.

Approach	Payoff threshold in features digitised			
	Low estimate	Mid estimate	High estimate	
Crowdsourcing versus project staff using desktop GIS	420	525	1,250	1,600
Crowdsourcing versus volunteers using desktop GIS	910	1,260	2,750	3,000
				7,400
				10,250

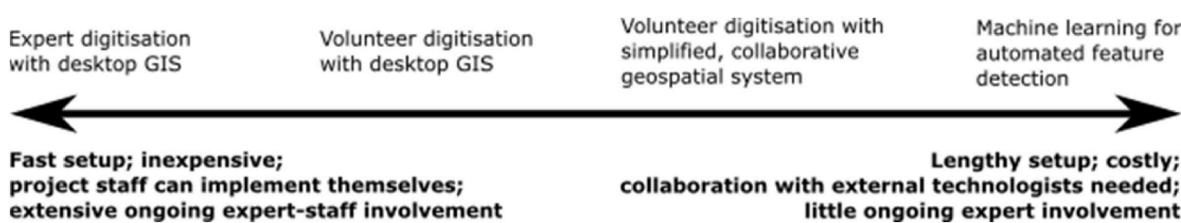


Fig. 7. Trade-off between time required for setup versus time required for ongoing expert involvement.

completed by a student programmer for a modest cost (ca. AUD \$2,000). Those 21 internal staff hours represent a digitisation rate of over 500 features per staff-hour. Twenty-one hours would have yielded just 1,260–1,575 features if staff had digitised them directly, or 2,730–3,780 had we supervised students using desktop GIS.

Second, given competing responsibilities, staff time during the field season was scarce and valuable. We could afford to invest in customisation and setup beforehand, and error checking after, if it reduced staff obligations during fieldwork. Across two seasons, in-field support for volunteers was only 7 h, representing about 1,550 features per in-field staff-hour. Seven hours would only allow staff to directly digitise 420–525 features, or supervise the digitisation of 910–1,260.

Third, the marginal cost for each additional feature digitised is low. This figure includes in-field support and quality assurance (13 h), and translates to 4.3 s of staff support per additional feature. Thus, the larger the dataset, the more value is extracted from the setup and deployment time. Adding more volunteers does not increase setup time at all. Preparing and distributing additional maps took only 6 min per map (6 h for 58 maps). Even adding another field season only costs one additional hour of setup time (based on our 2018 redeployment). By comparison, in 2010, the demands on staff time related to volunteer support never plateaued, as attrition meant that we were constantly onboarding new volunteers, while the learning curve of desktop GIS meant that support time declined only slowly as students became more experienced. The scalability of our crowdsourcing approach makes it more attractive if a project may expand over time to include more volunteers, more redeployments, or more maps.

Finally, qualitative factors also argue for implementing a crowdsourcing approach using a mobile application. First, the need for staff to be continually available to troubleshoot problems with desktop GIS, lest digitisation stall, provided a continual source of stress and distraction. Second, when desktop GIS was used, volunteers perceived digitisation as a burden. Tethered to a computer (which was also needed for other work), doing a time-consuming and tedious task, digitisation was one of the least popular activities, reducing morale and causing friction with other project activities and participants. The switch to a lightweight GIS running on mobile devices nearly eliminated the need for staff interventions and improved volunteer satisfaction. Learning the system was less of a burden, while offline data collection using mobile devices freed volunteers to work when and where they chose, regardless of demand for project computers or the availability of internet in rural Bulgaria. Taken together, this approach better utilised the time, attention, and motivation of both staff and participants.

#### 4.1.2. Machine learning versus crowdsourcing

Automation only pays off if a task is repeated frequently enough (see <https://xkcd.com/1205/>). Unfortunately, ML papers rarely quantify time-on-task (Uhl et al., 2020; Saeedimoghaddam & Stepinski, 2020; Herrault et al., 2013; Gimmi et al., 2016, p. 9), making it difficult to assess how large a dataset needs to be before the investment in ML approaches yields time savings. The ERC-funded Urban Occupations Project (Can, Gerrits, and Kabadai 2021), however, provides one benchmark for judging when pursuing a ML approach might be worthwhile. This project reported 1,250 h of manual digitisation to create enough training data to classify roads visible in historical maps of the Ottoman Empire. Using this input, and after additional preprocessing and filtering, an ML expert spent seven days testing and fine tuning the model. The output was impressive: some 300,000 km of roads were digitised. This example, which appears to have required a minimum of about 1,300 h of preparation time alone, suggests that ML approaches are worthwhile for large-scale projects that benefit from the consistent symbology and style (as found in British and Ottoman imperial maps). Time was not divided between specialists and (perhaps untrained) staff or volunteers undertaking training set digitisation, so comparisons with our approach include volunteer time.

A minimum threshold for automation can be extrapolated from our

2017–18 fieldwork and the Urban Occupations Project. We spent 44 staff hours customising and deploying a streamlined geospatial system in FAIMS Mobile, 184 participant-hours digitising features, seven staff-hours directly supporting that digitisation, and six staff hours checking for errors. These 241 h produced a dataset of 10,827 features, a rate of 44.9 features/person-hour. At that rate, the 1,300 h it took to deploy the ML approach taken by Can, Gerrits, and Kabadai would yield about 58,400 records, assuming that all features discovered by the ML model take zero additional personnel time, that our target symbols are no more difficult to extract than road segments, and discounting time the Urban Occupations Project spent on quality assurance, which they did not report.

To summarise in round numbers, a crowdsourcing approach like ours is most suitable for datasets numbering perhaps 10,000–60,000 records, assuming similar feature characteristics and data collection requirements (see Table 5). Below 10,000 records, approaches using desktop GIS should be considered. Under the most conservative scenario that considers all invested time, the use of our system pays off between about 3,500–4,500 features versus direct digitisation by staff using desktop GIS, and about 7,500–10,000 features versus support for volunteers using desktop GIS. Projects where staff time is at a premium, or that operate alongside fieldwork where staff have many competing demands, may find it valuable for smaller datasets (even those below 1,000 records). Above 60,000 records, ML approaches should be contemplated, but only if a project has access to the requisite expertise.

#### 4.2. Combining crowdsourcing and ML approaches

Finally, since training a model requires a manually produced dataset and a degree of manual error-checking, a combination of ML and crowdsourcing approaches might serve even large-scale projects. A dataset big enough to justify ML will likely need a training dataset big enough to warrant crowdsourcing, especially if the features or background are variable. Once the crowdsourcing platform has been built, moreover, it can be used to produce additional datasets for error-checking to confirm the accuracy of the ML results. The approaches are not exclusive, therefore, but complementary.

#### 4.3. Overall feasibility

Today, a typical project in history or archaeology – often small, under-resourced, and pursuing several research activities – may not be able to dedicate the personnel, infrastructure, or attention needed to incorporate ML successfully, but could deploy a collaborative geospatial system for crowdsourcing map digitisation. Deploying such a system requires a higher up-front investment in time and expertise than use of desktop GIS approaches, but it is feasible. A project with a digital humanist or similar technologist with skills at the level of core Software Carpentry lessons ([TheCarpentries, 2023](#)) can customise and operate a generalised platform such as FAIMS Mobile to implement an effective crowdsourcing system. An entire class of software for customisable geospatial data collection on mobile devices exists (e.g., ArcGIS Field Maps, QField, Mergin Maps, or GeoODK). Many of these systems attempt to make customisation as easy as possible, a goal at the heart of recent FAIMS redevelopment ([ARDC, 2022](#)); in future the technical barriers to

**Table 5**

Conservative estimates of payoff thresholds in terms of features digitised for our crowdsourcing approach versus desktop GIS and ML alternatives.

Approach	Payoff threshold (number of features)
Crowdsourcing versus project staff using desktop GIS	4,500
Crowdsourcing versus volunteers using desktop GIS	10,000
Crowdsourcing versus machine learning	60,000

deploying such systems will likely decline.

## 5. Conclusion

Toolkits for historical map digitisation have expanded dramatically in the last decade. Projects need to weigh the trade-offs between different approaches. The complexity of full-featured, desktop GIS makes it difficult for novices to use, limiting the scale of digitisation. ML approaches require significant resources to implement and expertise to avoid failures arising from training bias or other pitfalls. Purpose-built, lightweight, collaborative geospatial data recording systems fill the gap between. Our use of FAIMS Mobile facilitated offline, multi-user map-feature digitisation, minimised post-processing, supported multiple data types, and produced high-quality data aligning with our methods and aims - with few compromises, work-arounds, or other technological distractions.

The deployment of the Map Digitisation FAIMS Mobile customisation facilitated the rapid digitisation (57 staff-hours; 184 volunteer-hours; 241 total) of 10,827 features found in Soviet topographic maps, including the collection of geospatial data, structured data, text, and metadata. It required only modest hardware and minimal supervision, but supported offline operation, including in-field setup, data collection, synchronisation across multiple devices, and data backup and export, so it could be deployed alongside other project activities during fieldwork. All collected data was available daily for review, and a comprehensive, FAIR-compliant dataset was ready for analysis with less than 2 h of processing after collection. Overall quality of this dataset was high. Some 2% of records had recoverable data omissions which were corrected during post-processing and subsequently minimised through the addition of validation. An accuracy check by staff covering 7% of digitised features indicated an error rate of under 6%; errors were predictable and would be easily mitigated by redundant digitisation by volunteers or volunteer peer review. If staff time is the primary limiting resource, our approach becomes worthwhile for datasets no larger than about 4,500 features versus direct digitisation of features by expert staff, and no larger than about 10,000 features versus volunteer digitisation using desktop GIS. It may pay off for fewer than 1,000 if project staff time is at a premium. It remains the most efficient approach for datasets up to at least 60,000 features, above which automated approaches like ML should be considered. Our approach can also be used to produce the training datasets needed for ML, and as a tool for quality assurance for ML outputs.

This approach is readily transferable to other mobile GIS systems and map corpora, but our experience provides only a single data point for assessing the applicability of various digitisation approaches to historical maps. We next plan to take our dataset and use it to train and error-check a ML model, to more systematically compare the results of crowdsourcing versus machine learning. More projects - whether they use manual or automated approaches - need to track and publish the expert and volunteer time required for setup, training, support, and quality assurance related to map digitisation, as well as digitisation speed, error rates and types, the characteristics of the features being digitised, and the complexity of information extracted. More such data could refine and generalise the recommendations proposed here.

## Funding

This work was supported by the Endeavour Short-Term Mobility Programme from the Australian Department of Education, grant 19686; National eResearch Collaboration Tools and Resources (NeCTAR) under eResearch Tools grant RT043; the Australian Research Council under Linkage, Infrastructure, Equipment and Facilities (LIEF) grant LE140100151, the New South Wales Department of Industry under a Research Attraction & Acceleration Program (RAAP) award; the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) under the ON technology innovation program;

Macquarie University and UNSW Australia under internal infrastructure grant schemes.

## Author statement

Sobotkova, A.: Conceptualisation, Methodology, Investigation, Resources, Software, Writing –original draft preparation, Writing – Review & Editing, Project Administration. Ross, Shawn A.: Writing- original draft preparation, Validation, Writing – Review & Editing. Nassif-Haynes, Ch.: Software, Writing – original draft preparation. Ballsun-Stanton, B.: Software, Writing – original draft preparation.

## Declaration of competing interest

Sobotkova, Ross, Nassif-Haynes, and Ballsun-Stanton report that they have managed or been employed by the Field Acquired Information Management Systems (FAIMS) project, a university-based research infrastructure project. FAIMS was primarily grant-funded, but also offered customisation and support on a fee-for-service basis through a research consultancy arrangement at Macquarie University (no fees arose from this project).

## References

- About. (2023). *Zooniverse mobile App*. Zooniverse. 2023 <https://www.zooniverse.org/about/mobile-app>.
- ARDC. (2022). 'FAIMS 3.0 | ARDC'. Australian research data commons, 13 May 2022 <https://ardc.edu.au/project/faims-3-0/>.
- Bahng, H., Chun, S., Yun, S., Choo, J., & Joon Oh, S. (2020). Learning de-biased representations with biased representations. In H. Daumé III, & A. Singh (Eds.), *Proceedings of the 37th international conference on machine learning* (Vol. 119). Proceedings of Machine Learning Research. PMLR, 528–39 <http://proceedings.mlr.press/v119/bahng20a.html>.
- Ballsun-Stanton, B., Ross, S. A., Sobotkova, A., & Crook, P. (2018). FAIMS mobile: Flexible, open-source software for field research. *SoftwareX*, 7, 47–52. <https://doi.org/10.1016/j.softx.2017.12.006>
- Bennett, R., Cowley, D., & De Laet, V. (2014). The data explosion: Tackling the taboo of automatic feature recognition in airborne survey data. *Antiquity*, 88(341), 896–905. <https://doi.org/10.1017/S0003598X00050766>
- Besse, P., Castets-Renard, C., Garivier, A., & Loubes, J.-M. (2018). *L'IA Du Quotidien Peut-Elle être éthique. Loyauté Des Algorithmes D'apprentissage Automatique*. <https://doi.org/10.48550/arXiv.1810.01729>. Arxiv.org.
- Bishop-Taylor, R., Judd, K. L., Clear, L., & Martin, L. (2018). Surface soil survey in an archaeological context:the Kazanlak geoscience project. In S. Adrian Ross, A. Sobotkova, J. Tzvetkova, N. Georgi, & C. Simon (Eds.), *The Tundzha regional archaeological project: Surface survey, palaeoecology, and associated studies in central and southeast Bulgaria, 2009–2015 final report* (pp. 51–65). Oxford: Oxbow Books, Limited.
- Budig, B., Thomas van Dijk, C., Feitsch, F., & Giraldo Arteaga, M. (2016). Polygon consensus: Smart crowdsourcing for extracting building footprints from historical maps. In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 1–4). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2996913.2996951>. SIGSPATIAL '16 66.
- Can, Y. S., Gerrits, P. J., & Erdem Kabadayi, M. (2021). Automatic detection of road types from the third military mapping survey of Austria-Hungary historical map series with deep convolutional neural networks. *IEEE Access*, 9, 62847–62856. <https://doi.org/10.1109/ACCESS.2021.3074897>
- Can, Y. S., & Kabadayi, M. E. (2021). Text detection and recognition by using CNNs in the austro-Hungarian historical military mapping survey. In *The 6th international workshop on historical document imaging and processing* (pp. 25–30). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3476887.3476904>.
- Caspari, G. (2020). Mapping and damage assessment of “royal” burial mounds in the siberian valley of the kings. *Remote Sensing*, 12(5), 773. <https://doi.org/10.3390/rs12050773>
- Caspari, G., & Crespo, P. (2019). Convolutional neural networks for archaeological site detection – finding “princely” tombs. *Journal of Archaeological Science*. <https://doi.org/10.1016/j.jas.2019.104998>
- Cheverst, K., Davies, N., & Friday, A. (1998). Developing interfaces for collaborative mobile systems. In C. Johnson (Ed.), *Proceedings of the first workshop on human computer interaction with mobile devices*. Scotland: Department of Computing Science, University of Glasgow. GIST Technical Report G98-1 [https://www.researchgate.net/profile/Adrian-Friday/publication/2251335\\_Developing\\_Interfaces\\_For\\_Collaborative\\_Mobile\\_Systems/links/0c9605253210d60c57000000/Developing-Interfaces-For-Collaborative-Mobile-Systems.pdf](https://www.researchgate.net/profile/Adrian-Friday/publication/2251335_Developing_Interfaces_For_Collaborative_Mobile_Systems/links/0c9605253210d60c57000000/Developing-Interfaces-For-Collaborative-Mobile-Systems.pdf).
- Cholakov, I. D., & Chukalev, K. (2008). Statistical data on the archaeological field activities in Bulgaria, season 2007. *Archaeol. Bulg.*, XII(3), 89–100.

- Connor, S. E., Ross, S. A., Sobotkova, A., Herries, A. I. R., Mooney, S. D., Longford, C., & Iliev, I. (2013). Environmental conditions in the SE Balkans since the last glacial maximum and their influence on the spread of agriculture into Europe. *Quaternary Science Reviews*, 68, 200–215.
- Cowley, D. C. (2012). In with the new, out with the old? Auto-extraction for remote sensing archaeology. In *Remote sensing of the ocean, sea ice, coastal waters, and large water regions 2012* (Vol. 8532) International Society for Optics and Photonics, Article 853206. <https://doi.org/10.1117/12.981758>.
- Duckers, G. L. (2013). Bridging the “geospatial divide” in archaeology: Community based interpretation of LiDAR data. *Internet Archaeology*, 35. [https://intarch.ac.uk/journal/issue35/duckers\\_index.html](https://intarch.ac.uk/journal/issue35/duckers_index.html).
- Eftimovski, M., Ross, S. A., & Sobotkova, A. (2017). The impact of land use and depopulation on burial mounds in the Kazanlak Valley, Bulgaria: An ordered logit predictive model. *Journal of Cultural Heritage*, 23, 1–10. <https://doi.org/10.1016/j.culher.2016.10.002>
- Ekim, B., Sertel, E., & Erdem Kabadayı, M. (2021). Automatic road extraction from historical maps using deep learning techniques: A regional case study of Turkey in a German World War II map. *ISPRS International Journal of Geo-Information*, 10(8), 492. <https://doi.org/10.3390/ijgi10080492>
- Elwood, S. (2008a). Grassroots groups as stakeholders in spatial data infrastructures: Challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science: IJGIS*, 22(1), 71–90. <https://doi.org/10.1080/13658810701348971>
- Elwood, S. (2008b). Grassroots groups as stakeholders in spatial data infrastructures: Challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science: IJGIS*, 22(1), 71–90. <https://doi.org/10.1080/13658810701348971>
- Fuchs, D. J. (2018). The dangers of human-like bias in machine-learning algorithms. *Missouri S&T's Peer to Peer*, 2(1), 1. <https://scholarsmine.mst.edu/peer2peer/vol2/iss1/1>.
- Gimmi, U., Ginzler, C., Müller, M., & Psomas, A. (2016). Assessing accuracy of forest cover information on historical maps (Vol. 146, pp. 7–18). Prace Geograficzne. <https://www.ceeol.com/search/article-detail?id=510015>
- GO-FAIR. (2017). FAIR principles - GO FAIR. GO FAIR, 2017 <https://www.go-fair.org/fair-principles/>
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69(4), 211–221. <https://doi.org/10.1007/s10708-007-9111-y>
- Groom, G., Levin, G., Svensenning, S. R., & Perner, M. L. (2021). Dune sand-object based image analysis for vectorization of a dotted signature in Danish late 1800s maps. *e-Perimetron*, 16(4), 156–165. [https://pure.kb.dk/ws/files/23405989/Groom\\_et\\_al\\_1.pdf](https://pure.kb.dk/ws/files/23405989/Groom_et_al_1.pdf)
- Haas, P. (2017). The real reason to be afraid of artificial intelligence | TEDxDirigo. YouTube. YouTube [https://www.youtube.com/watch?v=TRzBk\\_KulaM](https://www.youtube.com/watch?v=TRzBk_KulaM).
- Herrault, P.-A., Sheeren, D., Fauvel, M., & Paegelow, M. (2013). Automatic extraction of forests from historical maps based on unsupervised classification in the CIELab color space. In D. Vandebroucke, B. Bucher, & J. Crompvoets (Eds.), *Geographic information science at the heart of Europe* (pp. 95–112). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-00615-4\\_6](https://doi.org/10.1007/978-3-319-00615-4_6)
- Jessop, M. (2007). The inhibition of geographical information in digital humanities scholarship. *Literary and Linguistic Computing*, 23(1), 39–50. <https://doi.org/10.1093/lit/fqm041>
- Jiang, S., Zhu, Y., Liu, C., Song, X., Li, X., & Min, W. (2022). Dataset bias in few-shot image recognition. In *IEEE transactions on pattern analysis and machine intelligence*. <https://doi.org/10.1109/TPAMI.2022.3153611>. February.
- Johnson, C., Johnson, P., Tom, R., Davies, N., Goldstein, M., Alsio, G., Brewster, S., et al. (1998). In C. Johnson (Ed.), *Proceedings of the first workshop on human computer interaction with mobile devices*, G9Vols. 8–1. GIST Technical Reports. Scotland: University of Glasgow. <https://www.dcs.gla.ac.uk/~johnson/papers/mobile/HCIMD1.html>.
- Jones, C. E., & Weber, P. (2012). Towards usability engineering for online editors of volunteered geographic information: A perspective on learnability. *Transactions in GIS*, 16(4), 523–544. <https://doi.org/10.1111/j.1467-9671.2012.01319.x>
- Kaplan, K. (2021). 10 usability heuristics applied to complex applications. Nielsen Norman Group, 15 August 2021 <https://www.nngroup.com/articles/usability-heuristics-complex-applications/>.
- Kecheva, N. (2019). Archaeological map of Bulgaria – transport and pipeline infrastructure projects. *Internet Archaeology*. <https://doi.org/10.11141/ia.51.2>. February.
- Kendall, D., Martin, H., & Hedenstroem, S. (2021). ‘Zikarta’. Zikarta, a map-based real-time collaborative virtual environment, 19 July 2021 <https://zikarta.wordpress.com/haw-it-works/>.
- Kitov, G. (1993). Trakiyskite mogili. *Thracia*, 10, 39–80.
- Koski, C., Rönneberg, M., Kettunen, P., Eliassen, S., Hansen, H. S., & Oksanen, J. (2021). Utility of collaborative GIS for maritime spatial planning: Design and evaluation of baltic explorer. *Transactions in GIS*, 25(3), 1347–1365. <https://doi.org/10.1111/tgis.12732>
- Lambers, K., Verschoof-van der Vaart, W. B., & Bourgeois, Q. P. J. (2019). Integrating remote sensing, machine learning, and citizen science in Dutch archaeological prospection. *Remote Sensing*, 11(7), 794. <https://doi.org/10.3390/rs11070794>
- Lin, A. Y. M., Huynh, A., Lanckriet, G., & Barrington, L. (2014). Crowdsourcing the unknown: The satellite search for genghis khan. *PLoS One*, 9(12), Article e114046. <https://doi.org/10.1371/journal.pone.0114046>
- Loulanski, T., & Loulanski, V. (2017). ‘Thracian mounds in Bulgaria: Heritage at risk’. The historic environment: Policy & practice. <https://doi.org/10.1080/17567505.2017.1359918>. August.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 53(6), 1–35. <https://doi.org/10.1145/3457607>, 54.
- Nassif-Haynes, C., Bui, T., Sobotkova, A., & Hermánková, P. (2021). Map-Digitisation Module (version fc0b21c). GitHub. <https://github.com/FAIMS/map-digitisation/releases/tag/map-dig-2018>.
- Nielsen, J. (2012). ‘Usability 101: Introduction to usability’. Nielsen Norman Group, 3 January 2012 <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>.
- Nielsen, J. (2020). ‘10 usability heuristics for user interface design’. Nielsen Norman Group, 15 November 2020 <https://www.nngroup.com/articles/ten-usability-heuristics/>.
- Noble, R., González-Alvarez, I., Reid, N., Krapf, C., Pinchand, T., Cole, D., Lau, I., et al. (2018). *Regional geochemistry of the coompan area*. South Australia: Department for Energy and Mining. Report Book 2018/00036 <https://publications.csiro.au/rpr/download?pid=csiro:EP187470&dsid=DS1>.
- Noble, R., Reid, N., Klump, J., Robertson, J., Cole, D., Fox, D., Pinchand, T., González-Alvarez, I., Krapf, C., & Lau, I. (2020). ‘Testing a rapid sampling and analysis workflow in the remote Nullarbor plain, Australia’. Explore. Newsletter for the Association of Applied Geochemists. March 2020 <https://www.appliedgeochemists.org/sites/default/files/documents/Explore%20issues/Explore186%20March%202020.pdf>.
- Owen, D., Green, A., & Elias, P. (2009). Review of spatial data needs and resources. German Council for Social and Economic Data (RatSWD). <https://doi.org/10.2139/ssrn.1452634>
- Pascoe, J., Ryan, N. S., & Morse, D. R. (1998). Human computer giraffe interaction: HCI in the field. In C. Johnson (Ed.), *Proceedings of the first workshop on human computer interaction with mobile devices*. Scotland: GIST Technical Report G98-1. Department of Computing Science, University of Glasgow. <http://www.dcs.gla.ac.uk/~johnson/papers/mobile/HCIMD1.html#Toc420818965>
- Pascoe, J., Ryan, N., & Morse, D. (2000). Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction*, 7(3), 417–437. <https://doi.org/10.1145/355324.355329>
- Petrie, C. A., Orengo, H. A., Green, A. S., Walker, J. R., Garcia, A., Conesa, F., Robert Knox, J., & Singh, R. N. (2018). Mapping archaeology while mapping an empire: Using historical maps to reconstruct ancient settlement landscapes in modern India and Pakistan. *Geosciences Journal*, 9(1), 11. <https://doi.org/10.3390/geosciences9010011>
- Pödör, A. (2015). Experiment in involving students in preserving geographical names appearing on historical maps. *e-Perimetron*, 10(1), 42–48. [http://www.e-perimetron.org/Vol\\_10\\_1/Podor.pdf](http://www.e-perimetron.org/Vol_10_1/Podor.pdf).
- Rondelli, B., Stride, S., & Juan José García-Granero. (2013). Soviet military maps and archaeological survey in the Samarkand region. *Journal of Cultural Heritage*, 14(3), 270–276. <https://doi.org/10.1016/j.culher.2012.06.006>
- Ross, Adrian, S., Sobotkova, A., Ballsun-Stanton, B., & Crook, P. (2013). Creating eresearch tools for archaeologists: The federated archaeological information management systems project. *Australian Archaeology*, 77(1), 107–119. <https://doi.org/10.1080/03122417.2013.11681983>
- Ross, Adrian, S., Sobotkova, A., Connor, S., & Ilija Iliev. (2010). An interdisciplinary pilot project in the environs of Kabyle, Bulgaria. *Archaeologica Bulgarica*, 14(2), 69–85.
- Ross, Adrian, S., Sobotkova, A., Julia, T., Georgi, N., & Simon, C. (Eds.). (2018). *The Tundza regional archaeological project: Surface survey, palaeoecology, and associated studies in central and southeast Bulgaria, 2009–2015 final report*. Oxford: Oxbow Books, Limited.
- Ryan, N., Pascoe, J., & Morse, D. (1999). *FieldNote: Extending a GIS into the field: New techniques for old times CAA98 computer applications and quantitative methods in archaeology*, 127–32.
- Saeedimoghadam, M., & Stepiniski, T. F. (2020). Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *International Journal of Geographical Information Science*. <https://doi.org/10.1080/13658816.2019.1696968>
- Schwemmer, C., Knight, C., Bello-Pardo, E. D., Oklobdzija, S., Schoonvelde, M., & Lockhart, J. W. (2020). Diagnosing gender bias in image recognition systems. *Socius*, 6(January), Article 2378023120967171. <https://doi.org/10.1177/2378023120967171>
- Sengupta, E., Garg, D., Choudhury, T., & Aggarwal, A. (2018). Techniques to eliminate human bias in machine learning. In *2018 international conference on system modeling & advancement in research trends. SMART*. SMART. <https://doi.org/10.1109/SYSMART.2018.8746946>, 226–30.
- Shkorpil, H., & Shkorpil, K. (1989). *Mogili. Plovdiv: Pchela*.
- Simon, R., Barker, E., Isaksen, L., & de Soto Cañamares, P. (2015). Linking early geospatial documents, one place at a time: Annotation of geographic documents with recognito. *e-Perimetron*, 10(2), 49–59. <http://oro.open.ac.uk/43613/>
- Sobotkova, A. (2013). Resisting rule in ancient Thrace. In M. Manoleakis (Ed.), *Exploring the hospitable sea: proceedings of the international workshop on the black sea in antiquity held in Thessaloniki, 21–23 September 2012*. Oxford: Archaeopress, BAR, 133–46.
- Sobotkova, A. (2018a). Spatial variability in surface artefact distributions in the Kazanlak Valley. In S. A. Ross, A. Sobotkova, J. Tzvetkova, N. Georgi, & C. Simon (Eds.), *The Tundza regional archaeological project: Surface survey, palaeoecology, and associated studies in central and southeast Bulgaria, 2009–2015 final report*. Oxford: Oxbow Books, Limited, 93–116.
- Sobotkova, A. (2018b). Spatial variability in surface artefact distributions in the Yambol study areas. In S. A. Ross, A. Sobotkova, J. Tzvetkova, N. Georgi, & C. Simon (Eds.), *The Tundza regional archaeological project: Surface survey, palaeoecology, and associated studies in central and southeast Bulgaria, 2009–2015 final report*. Oxford: Oxbow Books, Limited, 164–79.
- Sobotkova, A., & Ross, S. A. (2018). Kazanlak survey results. In S. A. Ross, A. Sobotkova, J. Tzvetkova, N. Georgi, & C. Simon (Eds.), *The Tundza regional archaeological*

- project: Surface survey, palaeoecology, and associated studies in central and southeast Bulgaria, 2009–2015 final report (pp. 66–81). Oxford: Oxbow Books, Limited.
- Sobotkova, A., Ross, S. A., Ballsun-Stanton, B., Fairbairn, A., Thompson, J., & Parker, V. V. (2016). Measure twice, cut once: Cooperative deployment of a generalised, archaeology-specific field data collection system. In E. W. Averett, J. M. Gordon, & D. B. Counts (Eds.), *Mobilizing the past: Recent approaches to archaeological fieldwork in the digital age*. University of North Dakota Digital Press in conjunction with Mukurtu 2.0 (Center for Digital Archaeology), 334–70.
- Sobotkova, A., Ross, S. A., Hermankova, P., Lupak, S., Nassif-Haynes, C., Ballsun-Stanton, B., & Kasimi, P. (2021). Deploying an offline, multi-user mobile system for digital recording of landscape archaeology in the perachora peninsula, Greece. *Journal of Field Archaeology*, 46(8), 571–594. <https://doi.org/10.1080/00934690.2021.1969837>
- Sobotkova, A., & Weissova, B. (2020). Soviet topographic maps and burial mounds of the Yambol province: Digital workflow for mortuary landscape verification. *Archaeological Prospection*, 33(February), 233. <https://doi.org/10.1002/arp.1769>
- Sturgeon, D. (2022). Crowdsourcing the historical record: Creating linked open data for Chinese history at scale. *International Journal of Humanities and Arts Computing*, 16 (1), 50–63. <https://doi.org/10.3366/ijhac.2022.0276>
- The Carpentries. (2023). 'Our lessons'. Software Carpentry. 2023. <https://software-carpentry.org/lessons/>.
- Uhl, J. H., Leyk, S., Chiang, Y.-Y., Duan, W., & Knoblock, C. A. (2020). Automated extraction of human settlement patterns from historical topographic map series using weakly supervised convolutional neural networks. *IEEE Access: Practical Innovations, Open Solutions*, 8, 6978–6996. <https://doi.org/10.1109/access.2019.2963213>
- UNESCO World Heritage Centre. (1979). *Thracian tomb of Kazanlak*. UNESCO World Heritage Centre, 1979 <https://whc.unesco.org/en/list/44>.
- UNESCO World Heritage Centre. (2016). *The royal necropolis of the thracian city of seuthopolis – a serial site, extension of the Kazanlak thracian tomb*. UNESCO World Heritage Centre, 2016 <https://whc.unesco.org/en/tentativelists/6085/>.
- Valchev, T., & Sobotkova, A. (2019). Monitoring burial mounds in the Yambol province: Deploying mobile technology to improve cultural heritage protection. In P. Kyriakidis, A. Agapiou, & V. Lysandrou (Eds.), *Spreading excellence in computer applications for archaeology and cultural heritage. Proceedings of the 3rd conference on computer applications and quantitative methods in archaeology Greek chapter (CAA-GR) limassol, 18–20 june 2018* (pp. 19–23). Limassol-Cyprus: Cyprus University of Technology. <http://www.academia.edu/download/59234548/CAA-GR-2018-proceedings20190513-87130-1c22q1m.pdf#page=31>.
- VanValkenburgh, P., Luiza, O. G. S., Repetti-Ludlow, C., Gardner, J., Jackson, C., & Ballsun-Stanton, B. (2018). Mobilization as mediation: Implementing a tablet-based recording system for ceramic classification. *Advances in Archaeological Practice*, 1–15. <https://doi.org/10.1017/aap.2018.12>
- Vitale, V., de Soto, P., Simon, R., Barker, E., Isaksen, L., & Kahn, R. (2021). Pelagios – connecting histories of place. Part I: Methods and tools. *International Journal of Humanities and Arts Computing*, 15(1–2), 5–32. <https://doi.org/10.3366/ijhac.2021.0260>
- Wagendonk, A. J., & De Jeu, R. A. M. (2007). Sensible field computing. *Photogrammetric Engineering & Remote Sensing*, 73(6), 651–662. <https://www.ingentaconnect.com/content/asprs/pers/2007/00000073/00000006/art00002>.
- Wang, Z., Qinami, K., Karakozis, I. C., Genova, K., Nair, P., Hata, K., & Russakovsky, O. (2020). Towards fairness in visual recognition: Effective strategies for bias mitigation. In 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr42600.2020.00894>.