Archeologia e Calcolatori
35.2, 2024, 355-362
doi 10.19282/ac.35.2.2024.37

LABELS AND SYMBOLS: USING TEXT ON MAPS TO INVESTIGATE THE ANTIQUITIES ON THE ORDNANCE SURVEY MAPS OF GREAT BRITAIN

1. Introduction

This paper will use the representation of antiquities on the Ordnance Survey (OS) Maps of Great Britain as a case study to discuss some of the opportunities and limitations of applying Machine Learning (ML) to digitise the text on historical maps, and how the resulting dataset can be used to address archaeological research questions. While maps are increasingly recognised as rich historical sources (Hosseini *et al.* 2021) and digitised map collections become more available, it is still difficult to harness the potential of analysing the text on maps as a digital *corpus*. Historical maps are mostly accessed through close reading of single sheets, making systematic study of an entire series or collection time- and resource-consuming. Even though Optical Character Recognition (OCR) returns increasingly good results for newspapers and manuscripts (Moudgil *et al.* 2022), maps present distinct challenges, namely rich and complex backgrounds, and highly variable text featuring different fonts, spacing, and orientation.

Exploring maps through their textual components has been partly addressed by pioneer projects such as 'GB1900' (Aucott, Southall 2019; https://maps.nls.uk/projects/gb1900) and 'The Map of Early Modern London' (Jenstad 2011), which relied on volunteers to transcribe the text on selected maps. The need to scale up this approach and make it more sustainable drove the international project 'Machines Reading Maps' (MRM; https://machines-reading-maps.github.io/). Initial map *corpora* were provided by the National Library of Scotland, the Library of Congress, and the British Library. Thanks to the ML pipeline mapKurator (Kim *et al.* 2023; https://github.com/knowledge-computing/mapkurator-system), MRM harvested the text on the maps in the test collections and output it as GeoJSON. While this result appealed to researchers with computational literacy, the format was not user-friendly for most researchers in the humanities. In the second phase of MRM, collaboration with the David Rumsey Map Collection (DRMC, https://www.davidrumsey.com/) and its in-house data visualisation team (Luna Imaging) made a 'search by word' function available in the DRMC in a more accessible and impactful way (Vitale 2023). Both MRM's outputs, the GeoJSON datasets and the visual interface, will be proposed as research tools.

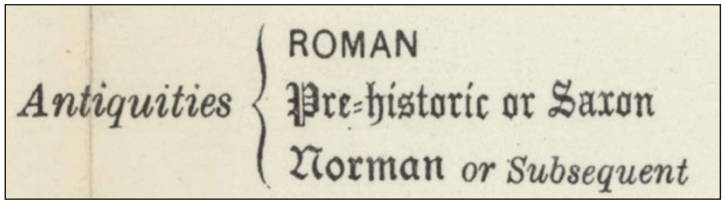

Fig. 1 – Characters of the writing on the engraved Six Inch Ordnance Maps of Great Britain, 1897, National Library of Scotland.

2. The Os Map series and its antiquities

The OS Maps were a national mapping endeavour that started in the 18th century for explicit military purpose, but then gradually became part of the life of the nation (Hewitt 2011), and almost a symbol of cultural identity (as well as of colonial power and acculturation, as Doherty 2004 remarks). The project involved, through the years, a number of map series and subsequent revisions. Among them, those most relevant to our case study were the County Series (25 inches to the mile). One of the peculiarities of the OS Maps was their detailed record of sites of historical relevance or, as they were referred to, antiquities (Close 1931). While the over-representation of antiquities on OS Maps can be partly explained with the popularity of antiquarianism in Victorian England (Phillips 1959; Goldhill 2011), their relevance in a non-specialist publication remains remarkable. This emphasis was reinforced by the appointment, in 1920, of O.J.S. Crawford as OS archeology officer. The reason why the antiquities on OS Maps are interesting both from a cultural heritage and a ML perspective is that they stand out very clearly as a distinct class of features.

Semiotically, this result is achieved through the combined use of three elements (not always present at the same time): dedicated fonts, dedicated keywords, a dedicated symbol. A specific symbol to mark historical sites (in our case a slightly ornate cross) is not an usual choice; examples can be easily found in both historical and contemporary maps. More interesting are the dedicated words (for example 'ruins' or 'site of') that may accompany the symbol (or act as one, as we will discuss), and, even more striking, the choice of three dedicated fonts (Fig. 1), each associated to a specific historical macro-period: Prehistoric or Saxon, Roman, and Norman (Ordnance Survey 1920). The use of bespoke fonts specifically created for the OS antiquities highlights the function that these words perform in the map: not simply labels that accompany and disambiguate symbols and icons (as they are defined by Schlichtmann 2018) but, to an extent, as symbols themselves or, better, as text performing additional symbolic function.

3. Labels and symbols: OS visual and textual code

Thanks to the combined use of these elements (the symbol, the fonts, the keywords) to mark the presence of antiquities, the OS Maps expand the usual scope of representing the contemporary landscape by including some unusual content:

– The invisible: contrary to common practice, the OS records the position of sites and monuments that have disappeared. Their status as nonextant is identified by the use of parentheses and dedicated words (for example '(site of)').
– The hypothetical: even more peculiarly, the OS includes sites and monuments of uncertain and/or unconfirmed location (for example 'supposed site of' or 'reputed site of') (Fig. 2).
– Archaeological events: dedicated vocabulary is used on OS maps to mark the position of, and/or add details about, archaeological excavations and the related findings (Fig. 3). This choice makes the series especially stand up as, rather than simply recording information about places, the OS surveyors added abundant data about discrete events (the act of finding Roman arrows or tripods, for example). The focus on the archaeological event is further emphasised by the mention of the year in which the finding took place.
– A chronological timeline: the use of three different, dedicated fonts not only makes the information about the antiquities easily findable on the map sheets, but also enables the reader to perceive, at a glance, the different chronological

Fig. 2 – Examples of the representation of non extant heritage on early OS maps, 1:6000 (composite screenshots from GB1900 portal).

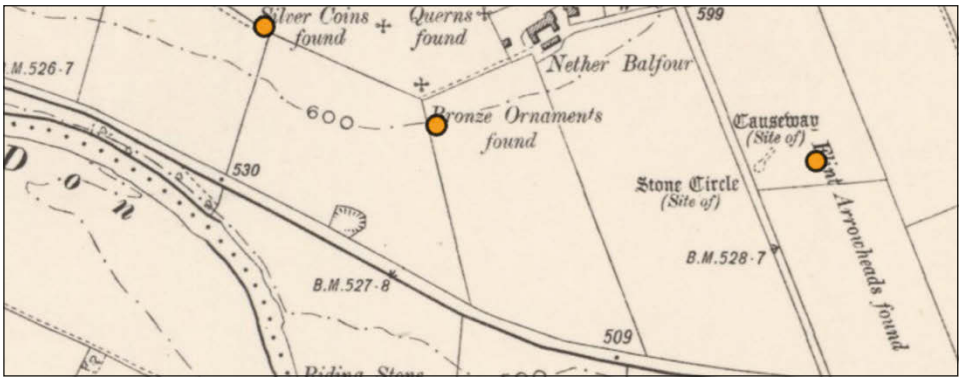

Fig. 3 – Example of the representation of archaeological events on early OS maps, 1:6000 (from GB1900 portal).

layers that coexist in the landscape, rural and urban, generating an effective visual timeline. A further chronological dimension is added by the references to the findings, expanding the chronology to include the recent past of the archaeological excavations (most commonly through the keyword 'found').

The level of archaeological expertise required to spot antiquities (including earth mounds and other less obvious features), label them, assign them to the correct macro-category (i.e. either roman, mediaeval or prehistoric), and mark the excavation's site is hardly to be expected in surveyors of the OS who were mostly trained in engineering and geography (Davidson 1987). What the official guidelines, compiled by Crawford (1922), show is that the surveyors were invited to supplement their knowledge making use of two types of external sources: published materials (including archaeological maps) and local knowledge (i.e. from the owners of the lands where the excavations took place and local heritage experts). The articulated representation of antiquities that can be observed on the early OS maps becomes, then, a valuable insight on the hidden components of the cartographic process, hinting at the many levels of authorships involved in the creation of this complex cultural object. But it may also guide us towards a critical re-examination of the several layers of biases and inaccuracies that are likely to have accumulated as a result.

4. Analysing text on maps

Such re-examination is truly possible only when two conditions are met: 1) the textual information on the maps has been digitised and made ready for computational analysis; 2) the analysis can be undertaken at scale, rather than one map at a time. Processing digitised maps with mapKurator

(or similar ML pipelines) produces a new kind of *corpus* that is, at the same time, textual and spatial, and, as such, can be usefully investigated with tools coming from the spatial humanities as well as computational linguistics. For example, using a combination of keywords especially employed to identify extant or disappeared sites ('ruins of', 'site of', 'remains of' among others) and keywords relating to archaeological discoveries ('found', but also words that are likely associated with findings such as 'urn', 'burial', 'tripod', 'arrow'), it is possible to create a sub-*corpus* of the textual component of the OS Maps that is specific to the representation of the antiquities. Once such a sub-*corpus* has been created, it enables informative comparisons with other datasets.

The most immediate option is, probably, that of comparing the earliest editions of the OS Maps with subsequent ones, looking for possible discrepancies. Highlighting, on a national scale, the antiquities that seem to have disappeared in later maps of the same series may be a powerful tool to spot lost and destroyed heritage, turning the earlier maps into visual (and surprisingly granular) archives of ancient sites and monuments (Close 1931, 149) that often predate the creation of formal national institutions for heritage preservation. Crawford himself (1922, 245) credits his predecessors for preserving the knowledge of the location of several long barrows that had already become hardly visible at the beginning of the 20th century.

At the same time, the disappearances of antiquities in later OS editions may be seen as an improvement due to a more rigorous approach to the recording of ancient heritage. Not only Crawford «should prefer to say nothing about the archaeological mistakes in the Ordnance Maps, because it is agreed by all that such exist» (1922, 247), but, more specifically, he points out the presence of Roman names that are «pure inventions» (1922, 248) and, as such, should be soon removed from subsequent editions.

A linguistic analysis of the text on maps across different OS editions would not have to be limited to place names. The close-reading of OS Maps has already made apparent, for example, how the archaeological language itself has evolved through time, deprecating terms such as 'antediluvian' or 'druidical'. An analysis at scale could offer even deeper insights on the history of archaeology and classical reception.

In its current state, the dataset generated by processing the OS Maps with mapKurator only takes into account the literal value of the text collected. The further layer of meaning added by the use of different fonts (or, in other words, the symbolic functions performed by the text of map) in invisible in the *corpus*. This case study on the OS antiquities magnified the extent of this limitation and has brought researchers in the MRM team to pursue more complex multimodal approaches that could look simultaneously at the words as textual labels and as symbols. Recognising automatically the antiquities' different fonts would add, to a dataset that already features names

and locations of ancient sites and monuments, also information about their interpretation at the time of the map creation. This further dimension would enable more nuanced comparisons with external datasets (such as English Heritage, Historic Environment Scotland, or Archaeology Data Service) that would move beyond the presence or absence of a site but could also evaluate changes in its archaeological and historical interpretation.

Beresford (1992) for example, evokes the case of the earthworks from mediaeval villages in Yorkshire that, recorded in the first edition of the OS, became candidates for deletion in a subsequent revision. Even Crawford (1922), in his honest evaluation of the archaeological value of the OS mapping endeavour, recognises that the interpretation of antiquities (via the choice of one of the three available fonts) may reveal more than a few mistakes in subsequent revisions. In his view, though, recording the location of a piece of ancient heritage was more important than getting its interpretation right, as interpretations could be corrected in the future «with the stroke of a pen» (1922, 251) while failing to record the presence of a site may contribute to its permanent loss. Beside rectifying incorrect interpretations, as Crawford would have hoped, an analysis at scale of these mistakes would support the discovery of potential systemic bias, for example highlighting the possible over popularity of one particular class of antiquities.

Linguistic analysis of the text on OS Maps could be extended to digital archives of archaeological scholarship (such as JStore) to facilitate the identification of the published sources that the OS surveyors were invited to rely upon in their work. Comparisons between the text on OS Maps and these publications may be used as a starting point to investigate the direction and intensity of knowledge exchange between the OS and academia, analysing to what extent the OS were mere duplications of reputable archaeological publications (including their inherent biases) and to what extent they generated new knowledge through integration of multiple sources (Davidson 1987).

Finally, the integration of a visual interface, like the one available on DRMC's website, enables us to reintroduce, partially, the symbolic component of the text on maps by showing the original map context. While lacking the computational sophistication of other approaches, a 'search by word' across maps can yield useful results. For example, a search for key words commonly associated with antiquities (such as 'ruins') may help us frame, culturally and cartographically, the practice of representing historical sites on maps. Bearing in mind that the DRMC, while sizable, is also heterogeneous and idiosyncratic like all private collections, it could still support the exploratory search for other examples of dedicated fonts used to mark antiquities on non-specialist maps. So far, the search through relevant keywords suggests that the approach is peculiar to the OS Maps, including their coverage of British colonies and protectorates.

5. Conclusions

This case study highlights how digitised map collections may be used as tools to reinvestigate the history of archaeological practice and cultural heritage reception. *Corpora* created by collecting text on maps through ML, open up the opportunity for the application of linguistic tools and systematic comparisons with variant editions, other collections, as well as external datasets. However, the symbolic value carried by words on maps through their visual characteristics (such as fonts) is currently overlooked. While visual interfaces have shown the potential of a combined approach, a new ML workflow is needed to fully leverage the value of OS (and other) maps as complex textual objects linking together different sources of archaeological knowledge.

Valeria Vitale
University of Sheffield
v.vitale@sheffield.ac.uk

Katherine McDonough
University of Lancaster
k.mcdonough@lancaster.ac.uk

REFERENCES

Aucott P., Southall H. 2019, *Locating past places in Britain: Creating and evaluating the GB1900 Gazetteer*, «International Journal of Humanities and Arts Computing», 13, 1-2, 69-94.

Beresford M. 1992, *The spade might soon determine it': The representation of deserted medieval villages on ordnance survey plans, 1849-1910*, «The Agricultural History Review», 64-70.

Close C. 1931, *The Ordnance Survey*, «Antiquity», 5, 18, 149-160 (https://doi.org/10.1017/S0003598X00005913).

Crawford O.G.S. 1926, *Archæology and the Ordnance Survey*, «Geographical Journal», 66-68.

Davidson J.L. 1987, *The collection of antiquarian information for the early Ordnance Survey maps of Scotland*, in *Proceedings of the Society of Antiquaries of Scotland*, 116, 11-16 (https://doi.org/10.9750/PSAS.116.11.16).

Doherty G.M. 2004, *The Irish Ordnance Survey: History, Culture and Memory*, Dublin, Four Courts Press.

Goldhill S. 2011, *Victorian Culture and Classical Antiquity: Art, Opera, Fiction, and the Proclamation of Modernity*, Princeton, University Press.

Hewitt R. 2011, *Map of a Nation: A biography of the Ordnance Survey*, London, Granta Publications.

Hosseini K., McDonough K., van Strien D., Vane O., Wilson D.C. 2021, *Maps of a nation? The digitized ordnance survey for new historical research*, «Journal of Victorian Culture», 26, 2, 284-299.

Jenstad J. 2011, *Using early modern maps in literary studies: Views and caveats from London*, «GeoHumanities», 112-119.

Kim J., Li Z., Lin Y., Namgung M., Jang L., Chiang Y. 2023, *The mapKurator system: A complete pipeline for extracting and linking text from historical maps*, in *SIGSPATIAL '23: Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, art. 35, 1-4 (https://doi.org/10.1145/3589132.3625579).

Moudgil A., Singh S., Gautam V. 2022, *An overview of recent trends in OCR systems for manuscripts*, in *Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021*, 291, 525-533.

Ordnance Survey 1920 (and later editions), *A description of the large scale maps of Great Britain produced and published by the Ordnance Survey, with specimens, symbol sheet, and diagrams*, Southampton, Ordnance Survey.

Phillips C.W. 1959, *Archaeology and the Ordnance Survey*, «Antiquity», 33, 131, 195-204.

Schlichtmann H. 2018, *Background to the semiotic study of maps*, «Meta-carto-semiotics», 11, 1, 1-12.

Vitale V. 2023, *Searching maps by words: How Machine Learning changes the way we explore map collections*, «Journal of Cultural Analytics», 8, 1.

ABSTRACT

This paper discusses the application of a machine learning pipeline to automatically digitise text on historical maps and make it searchable, as explored by the 'Machines Reading Maps' project. Looking at the Ordnance Survey Maps of Great Britain as a case study, we will suggest ways in which this new kind of open datasets, of both a textual and spatial nature, offers the unprecedented opportunity to study maps at scale, analysing map collections as digital corpora. These new approaches facilitate the use of map as historical sources in humanities research, and their investigation as complex cultural objects that combine heterogeneous knowledge. In particular, we will focus on the uncommonly detailed representation of ancient sites on the Ordnance Survey maps, and how a further layer of information around them is delivered not by the words' literal meaning but by their appearance. We will propose ways in which this peculiarity could be digitally leveraged to retrace, investigate, and perhaps re-interpret the archaeological information on the Ordnance Survey maps. We will conclude by reflecting on the need for new, more sophisticated workflows that take into account the richness of information delivered by visual clues in words printed on maps.