

Alexandre Nobajas¹

Targeted crowdsourced vectorisation of historical cartography

Keywords: crowdsourcing, gamification, systematic vectorisation, training set creation

Summary: Many institutions have systematically digitised their cartographic documents, so there are now millions of maps which have been digitised and can be accessed on-line. The next step many map libraries have undertaken has been to georeference their digitised maps, so they can now be used within modern digital cartographic dataset. However, many of these applications still use the raster version of the maps, so they hamper the full potential working with historical cartography has, as it limits the analytical capacity and impedes performing spatial analysis. Even though there have been timid steps to overcome this limitation, the discipline seems to be struggling to advance into the next step, the systematic vectorisation of historical maps. Vectorisation is the process of converting pixel-based images –raster- into node or point based images –vector, which can then be queried or analysed by individual components. Vector based maps have a series of advantages when compared to raster maps, as they allow scale changes without a loss in detail, classifying map features by type or performing spatial queries just to mention a few. Therefore, by vectorising a historical map we are providing an unprecedented level of usability to it and allowing detailed inquiries to the information contained into it. Although there have already been some successful experiences in vectorising historical maps, they have been either expensive or not systematic, so there is scope to build upon these successful experiences and achieve a widespread method of vectorisation which can be applied to map collections across the world.

There are different approaches to overcome this technological and methodological hurdle, most of them linked to the use of automation techniques such as using remote sensing land classification methods or, more recently, the use of machine learning and artificial intelligence techniques. However, a large number of training sets are necessary in order to train the automation systems and to benchmark the outputs, so large datasets are still needed to be generated. The use of crowdsourcing methods has been proven as a successful means of achieving the same goal, as experiences such as the ones developed by the New York Public Library demonstrate. If given the possibility and the right tools, people will help to vectorise historic documents, although Nielsen's 90-9-1 rule will apply, meaning that other ways of increasing participation are necessary in order to speed the vectorisation up.

This paper proposes a framework in which by using a variety of gamification methods and by targeting two different groups of people, retirees and school children, historic maps can be massively vectorised. If fully deployed, this system would not only produce a wealth of vectorised maps, but it would also be useful in teaching school children important disciplines such as cartography, geography and history while at the same helping older generations keeping them mentally active. Finally, an intergenerational dialogue that would enrich both participant groups would be possible thanks to their common participation in the vectorisation of historical documents.

¹ School of Geography, Geology and the Environment, Keele University, United Kingdom
[a.nobajas@keele.ac.uk]