











MapReader: v1.1.0

Katherine McDonough ^{1,2¶}, Daniel C. S. Wilson ¹, Andrew Smith ¹,
Kaspar Beelen ³, Kalle Westerling ¹, Rosie Wood ¹, and Kasra
Hosseini ^{1,4}

¹ The Alan Turing Institute, UK ² Lancaster University, UK ³ School of Advanced Study, University of London, UK ⁴ Zalando SE, Germany ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: 

Submitted: 15 December 2023

Published: unpublished

License

Authors of papers retain copyright
and release the work under a
Creative Commons Attribution 4.0
International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

MapReader is an interdisciplinary software library for processing digitized maps [[@Hosseini_mapreader](#)], but also other types of images, by ‘patching’ them into small, custom-sized cells, which are then classified according to the user’s needs. MapReader thus offers a flexible pipeline which can be used both for manual annotation of small datasets as well as for Computer Vision-based inference of large collections. As an example, in [[@Hosseini_mapreader](#)], we utilized MapReader’s interface to manually annotate 62,020 patches, used its functionalities to train a suite of computer vision models and performed model inference on approximately 30.5 million patches.

MapReader’s approach was inspired by methods in biomedical imaging, which were adapted for use by historians, and it is suitable for a wide range of applications in image analysis: it has, for example, been applied to an image classification problem in plant phenotype research [[@Corcoran](#)]. This cross-pollination between the humanities and the natural sciences was made possible by the open and reproducible research methods at the heart of MapReader.

MapReader pioneers a methodological shift in how historians interact with maps as primary sources. Sustained engagement with big collections of maps rarely moves beyond analysis of cartographic history. To change this, MapReader encourages historians to reflect on the content of maps and is designed to facilitate linking datasets representing visual map content with other historical geospatial data.

In this paper, we present the MapReader release at the conclusion of the Living with Machines project, which supported the development of the software and associated historical research. This release represents the culmination of extensive work to improve MapReader’s usability, especially through clear documentation and tutorials.

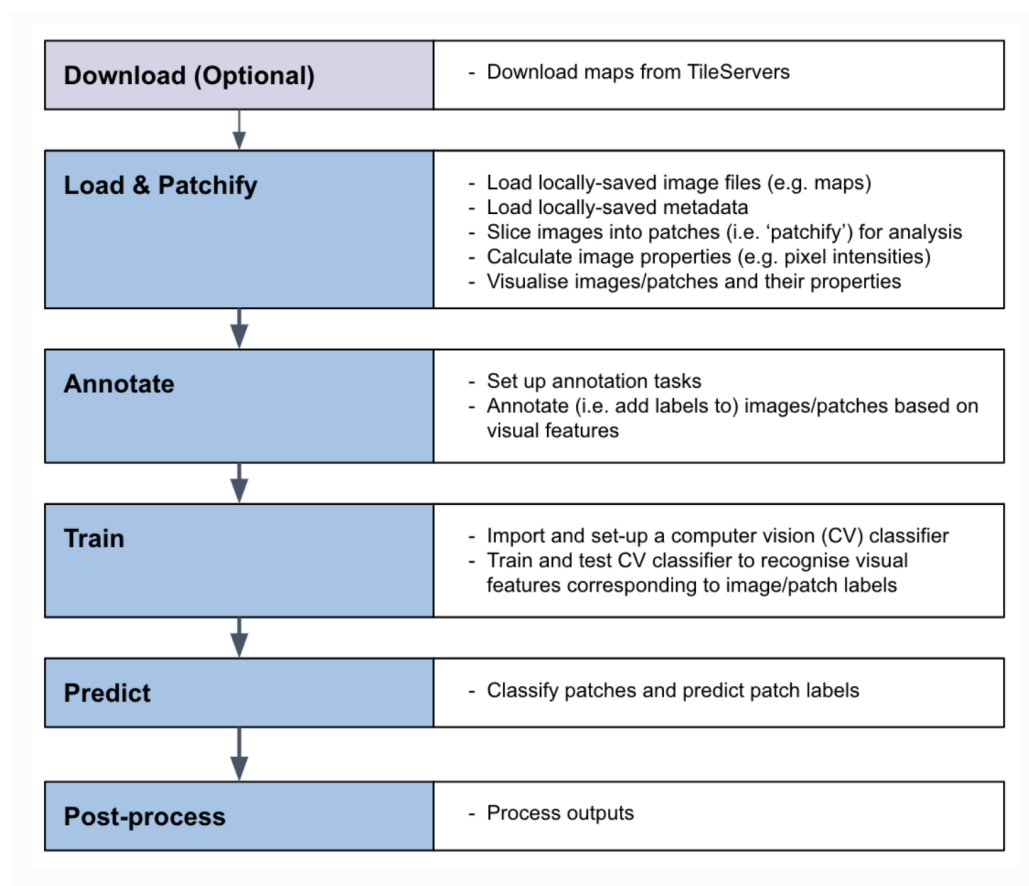


Figure 1: MapReader modules and input-outputs. Credit: Rosie Wood.

Statement of need

Since the 1990s, map libraries have been scanning maps and creating digital collections of these images [Hosseini_maps]. In 2023, there are more than a million images of maps in digital libraries and archives around the world, and yet it is very difficult for anyone to do more than browse them in a web viewer.

MapReader makes it possible to ask questions of thousands of digitized maps at a time, a fundamentally different intellectual experience from both the traditional manner of viewing a few maps at a time on a reading room table and the act of visually scanning digital files sequentially. As an example, we used MapReader to process a collection of ~16K nineteenth-century Ordnance Survey map sheets (~30.5M patches) covering England, Wales and Scotland [Hosseini_mapreader]. Inspired by the possibility of seeing series maps stitched together in seamless layers, such as the National Library of Scotland Ordnance Survey map viewing interface, MapReader takes the next step by transforming the experience of working with maps from surface exploration to critical investigation [Hosseini_maps].

Related Work

MapReader is among the first end-to-end pipeline for processing historical maps and other images that was designed to lower barriers to experimenting with computer vision in answering research questions about large image datasets. Other projects are emerging which are performing similar research tasks with the visual content in historical map collections [Petitpierre;

49 @Combes], and of course other tools, like the Distant Viewing Toolkit [@Arnold], address
50 similar needs for other kinds of media. As part of a collaboration between Machines Reading
51 Maps and the David Rumsey Historical Map Collection, the Knowledge Computing Lab released
52 mapKurator [@mapkurator], which takes map image input and creates geojson files of all text
53 on each map.

54 Documentation

55 MapReader aims to build computational skills among historians. Our extensive work
56 on documentation and training, including substantial updates to MapReader since
57 @Hosseini_mapreader, reflect this commitment. As historians explore the possibilities of
58 computational methods for novel historical research, MapReader models how computational
59 tools can unlock difficult-to-use primary sources and how we can embrace open research
60 practices as a way to encourage learning. We welcome contributions and requests for new
61 documentation or tutorials.

62 Our documentation includes: - About MapReader: A basic introduction to the software and
63 its origins - Events: Activities where the community can engage with MapReader - Project
64 Curriculum Vitae: Papers, talks, workshops, etc. delivered by the MapReader team - Coding
65 Basics: Skills and tools that are useful for getting started with MapReader - Installation
66 Instructions: How to install MapReader - Input Guidance: What kind of maps and which
67 formats work well in MapReader - User Guide: Walkthrough of how to run MapReader -
68 Worked Examples: Jupyter notebooks demonstrating uses of MapReader for specific cases
69 (with data, e.g. @Hosseini_mapreader_data) - API Reference: Auto-generated API reference
70 documentation - Code of Conduct and Inclusivity: Our approach to ethical and inclusive
71 conduct - Contribution Guide: How to engage with MapReader - Developer's Guide: How to
72 update the MapReader version number and upload to package managers

73 Conclusion

74 Through its conceptual approach, modular structure, documentation, and worked examples,
75 MapReader enables researchers to ask questions of large collections of maps. It represents a
76 novel approach to digitizing map content, one which intentionally prevents the collection of
77 overly precise data from cartographic documents. MapReader embraces a humanistic approach
78 to data creation and curation, offering an alternative or complement to pixel-level image
79 segmentation.

80 Acknowledgements

81 This work was supported by Data/Culture (AHRC grant AH/XXXXX), Living with Machines
82 (AHRC grant AH/S01179X/1), and The Alan Turing Institute (EPSRC grant EP/N510129/1).
83 Living with Machines, funded by the UK Research and Innovation (UKRI) Strategic Priority
84 Fund, was a multidisciplinary collaboration delivered by the Arts and Humanities Research
85 Council (AHRC), with The Alan Turing Institute, the British Library and the Universities
86 of Cambridge, East Anglia, Exeter, and Queen Mary University of London. Maps and their
87 metadata in MapReader are reproduced with the permission of the National Library of Scotland
88 (<https://maps.nls.uk/index.html>). We also wish to thank participants in events in 2023 who
89 provided feedback on using MapReader.

90 Contribution Statement

91 Katherine McDonough wrote and revised this article, with substantial contributions from
92 Daniel C.S. Wilson and Rosie Wood. Andy Smith, Kalle Westerling, Kaspar Beelen and Kasra

⁹³ Hosseini reviewed the final manuscript. Please see contributions to the MapReader software
⁹⁴ library at <https://github.com/Living-with-machines/MapReader#contributors>, including work
⁹⁵ from all named authors.

⁹⁶ **References**

DRAFT