

MapReader

FOSS4G Lancaster

Living with Machines
7 Oct 2023

Katie McDonough
Lancaster University

Our Partners



Our Funders



UK Research
and Innovation



Living with Machines

- Digitised collections from the long 19th century
- Computational methods
- Multidisciplinary approach
- Digitised Ordnance Survey map collection from the National Library of Scotland

Our Partners



Our Funders



UK Research
and Innovation

The Living with Machines Team

Principal and Co-Investigators



Ruth Ahnert
(QMUL)



David Beavan
(Turing)



Emma Griffin
(UEA)

Project team



Claire Austin
Rights Assurance



Kaspar Beelen
Digital Humanities Senior
Research Associate



Mariona Coll Ardanuy
Computational Linguistics
Senior Research Associate



André Piza
Research Project Manager



Griffith Rees
Research Data Scientist



Kalle Westerling
Research Software Engineer



Timothy Hobson
(Turing)



Jon Lawrence
(Exeter)



Maja Maricevic
(British Library)



Lélie Demertzí
Programme Coordinator



Luke Hare
Research Data Scientist



Sherman Lo
Research Data Scientist



Daniel Wilson
History Senior Research
Associate



Rosie Wood
Research Data Scientist



Barbara McGillivray
(Turing / King's College
London)



Mia Ridge
(British Library)



Alan Wilson
(Turing)



Katie McDonough
Senior Research Associate



Federico Nanni
Senior Research Data
Scientist



Nilo Pedrazzini
Corpus-Based Digital
Humanities Research
Assistant

Our Partners



The
Alan Turing
Institute

UNIVERSITY OF
CAMBRIDGE

UEA University of
East Anglia

UNIVERSITY OF
EXETER

Queen Mary
University of London

Our Funders

Arts & Humanities
Research Council

UK Research
and Innovation



1888

How do researchers work with maps?



Digitized maps can be more than sheets to browse in a virtual reading room. But how?

Ordnance Survey
maps of England,
Wales, and
Scotland

6 inches to 1 mile
1888-1913
(2nd edition)

~15K sheets



Scale: Making trustworthy historical claims based on *lots* of maps

case studies → ‘high resolution’ archival research
or anecdotes from printed materials

MAPS AS HUMANITIES DATA

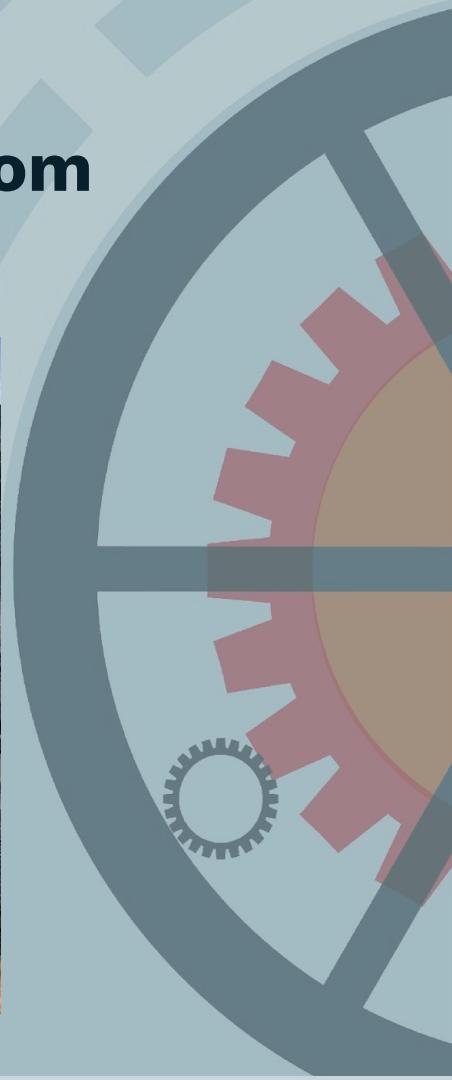
aggregated statistics → ‘low resolution’
regional/national

Opportunity: Computer Vision + Machine Learning

Iterative and reproducible: Historical OS maps tell us how Victorians represented the British landscape, and how that landscape was changing, but they are not a ‘ground truth’. We want to ***use CV to advance interpretation, not define truth.***

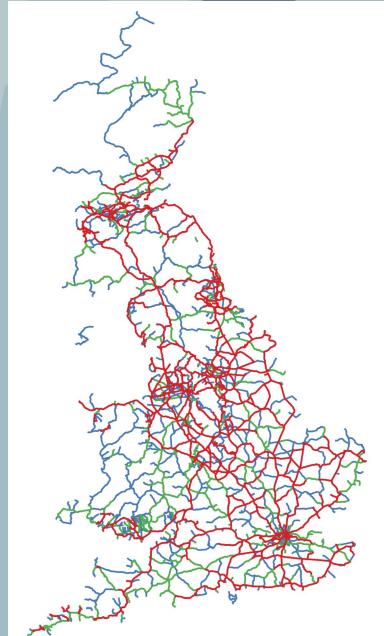
Fast: There are too many maps to examine with close reading. ML makes it possible to ***work with large collections quickly.***

AI for the humanities: Let's move on from data 'mining'



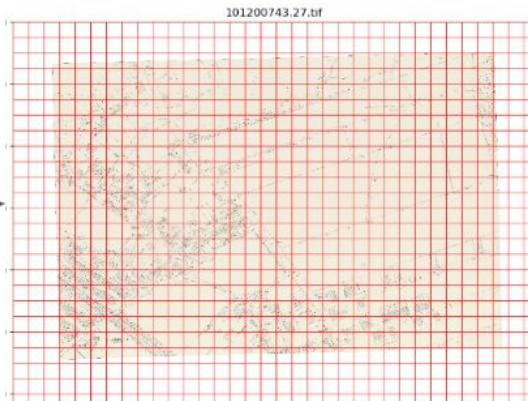
Cultural Spatial Data Creation

Can we identify and search for new shapes as intentionally imprecise **visual signs** forming new **patterns** rather than reproducing manual vector data collection?



Solution: ‘Patches’ as a new shape for historical research

Parent image



Patches



Image Classification with Maps: Raster Patches

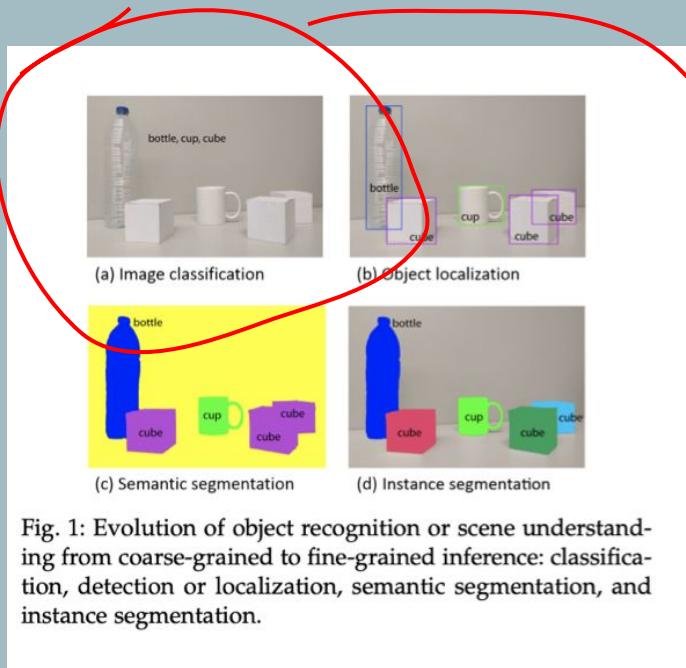
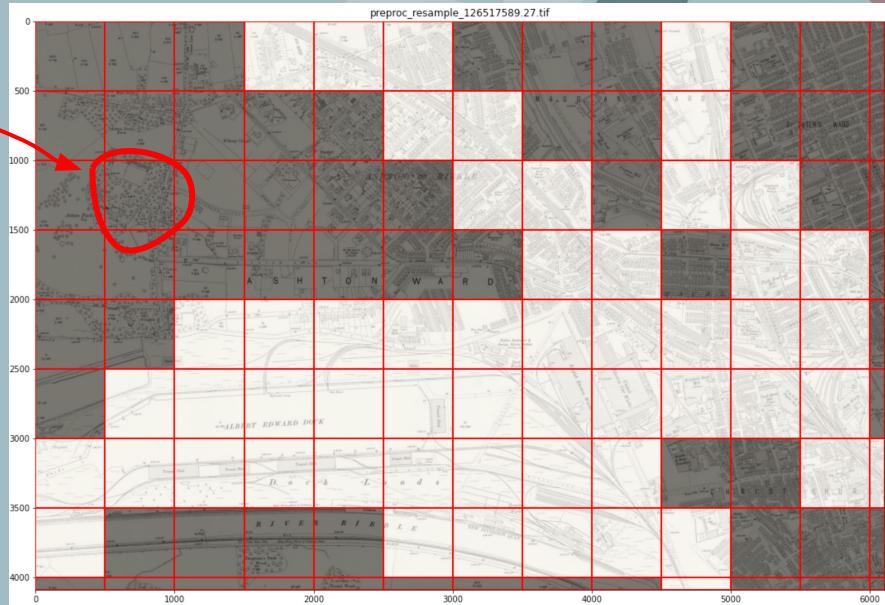


Fig. 1: Evolution of object recognition or scene understanding from coarse-grained to fine-grained inference: classification, detection or localization, semantic segmentation, and instance segmentation.



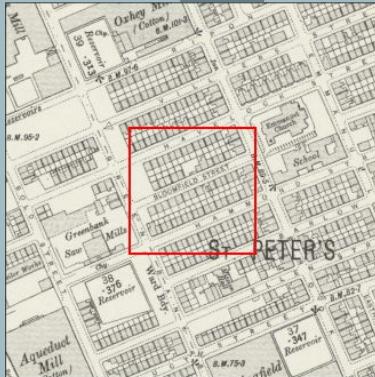
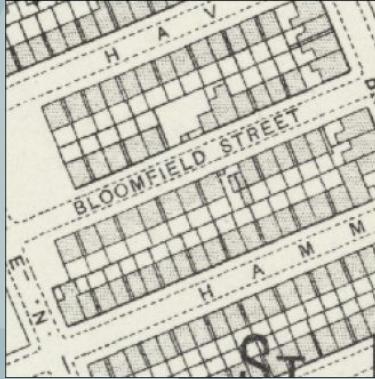
Map regions (patches) annotated as training data

Patches as humanistic data

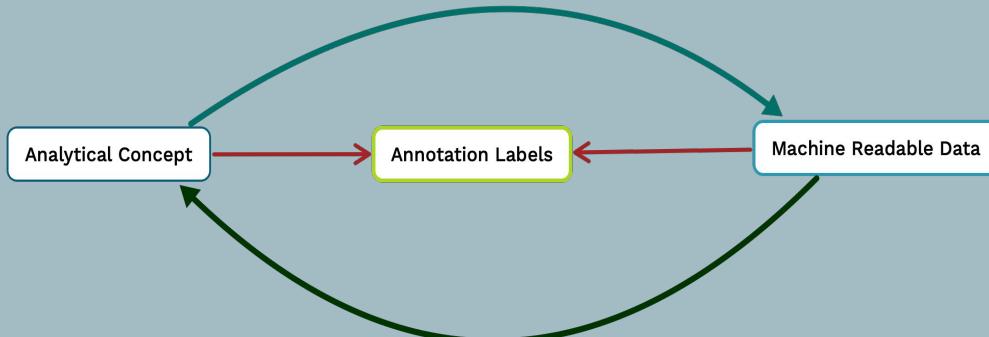
Image classification tasks on
'patches' are common in
medical imaging research.

Can we adapt this approach for
humanistic inquiry and search?

tile-5250-500-5500-750-#preproc_resample_126517589.27.tif#.PNG



Annotating patches: What is a good label?



Rail Space ¹ No Rail Space ² ← back ^j → next ^k

<Figure size 432x288 with 0 Axes>

tile-4500-2000-5000-2500-#preproc_resample_126517589.27.tif#.PNG

The screenshot shows a digital map interface. At the top, there are two buttons: 'Rail Space ¹' (green) and 'No Rail Space ²' (blue). Below the buttons are navigation links: '← back ^j' and '→ next ^k'. A status bar at the bottom indicates '<Figure size 432x288 with 0 Axes>' and the file name 'tile-4500-2000-5000-2500-#preproc_resample_126517589.27.tif#.PNG'. The main area displays two versions of a map patch. The top version is labeled 'Rail Space' and shows a map with several tracks and labels like 'B.M. 881', 'PRIORITY', 'ABBREV.', and 'INGOT ST.'. The bottom version is labeled 'No Rail Space' and shows the same map without the tracks. A red box highlights a specific area in both versions. Below the maps, the text 'Additional info:' and a URL 'URL to the NLS map: <https://maps.nls.uk/view/126517589>' are visible.

Patch sizes and label choices

E.g.



1000 pixel
Too big



200 pixel
Woods



100 pixel
Trees

Railspace patches

- British **railspace** as predicted by MapReader from ~15K late nineteenth-century 6" OS maps (NLS)
- ~30.5 million patches total
- Railspace = the totality of rail infrastructure in the landscape
- Captured on 50m x 50m patches
- ~62k expert-annotated patch dataset now available on [Hugging Face](#)

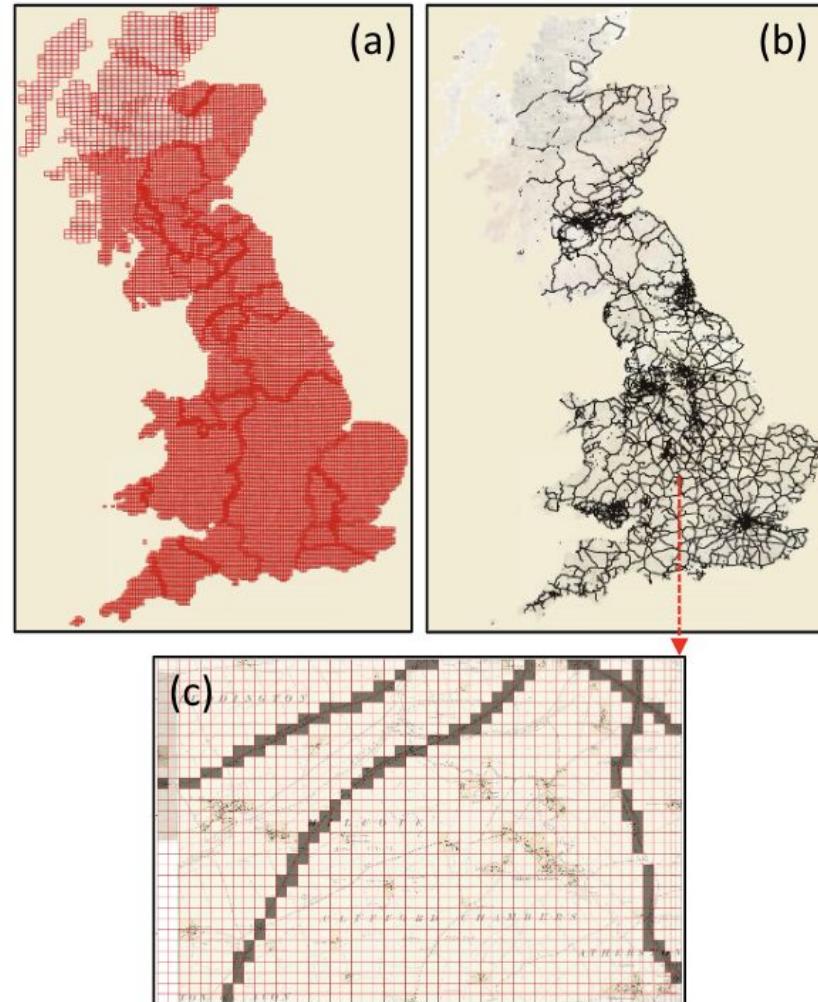


Image credit: Kasra Hosseini



Living with Machines: railspace and building datasets

Download and explore data derived from NLS 2nd-edition six inch to one mile Ordnance Survey maps of England, Wales, and Scotland.

This dataset was created by members of the **Living with Machines** digital history project (The Alan Turing Institute/British Library) while developing the **MapReader** computer vision pipeline. Living with Machines is a research project that rethinks the impact of technology on the lives of ordinary people during the Industrial Revolution. MapReader creates datasets for humanities research using historical map scans and their metadata as input. Here we share gold standard annotations and outputs from the earliest MapReader experiments.

Thank you

Find out more:

<https://livingwithmachines.ac.uk/latest/>
@LivingwMachines

<https://mapreader.readthedocs.io/>

Our Partners



Our Funders



UK Research
and Innovation