

# MapReader

Katie McDonough, Daniel Wilson,  
Rosie Wood

**Environmental Digital Humanities  
Workshop**

**29 January 2025  
University of Manchester**

#### Our Partners



#### Our Funders



# **Workshop material and research applications co-developed with:**

Rosie Wood, Turing

Kalle Westerling, Turing

Daniel Wilson, Turing

Kaspar Beelen, SAS, University of London

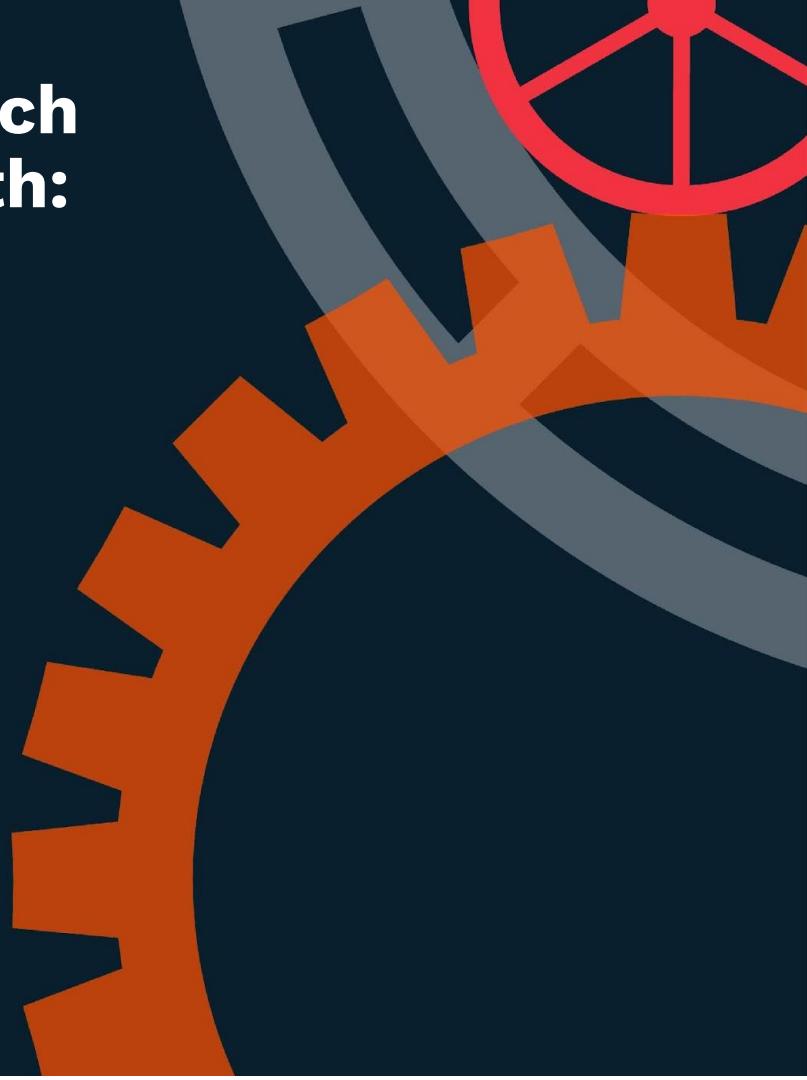
Tim Hobson, Turing

Kasra Hosseini, Zalando (formerly Turing)

Josh Rhodes, Durham

Jon Lawrence, Exeter

Andy Smith, Turing



Home + Research + Research projects

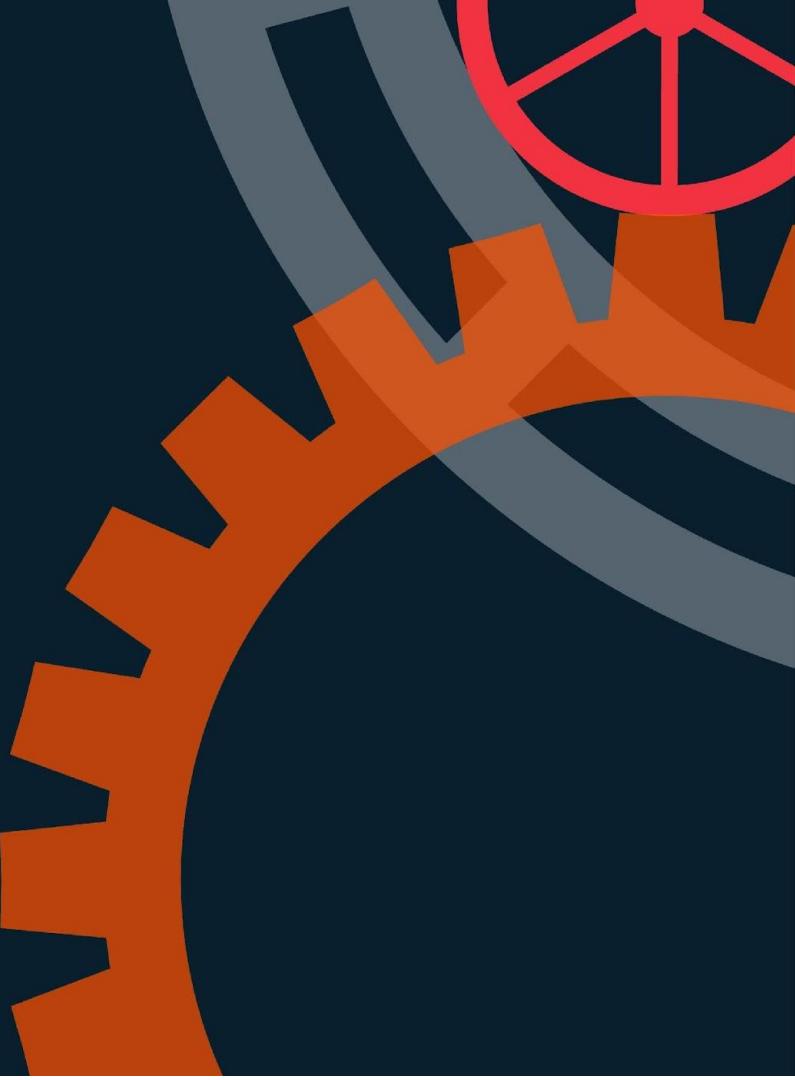
# Data/Culture: Building sustainable communities around Arts and Humanities datasets and tools

Learn more ↓

Project status  
Ongoing

**[https://www.turing.ac.uk/research/research-projects/  
dataculture-building-sustainable-communities-around-arts-and-humanities](https://www.turing.ac.uk/research/research-projects/dataculture-building-sustainable-communities-around-arts-and-humanities)**

# Why MapReader?



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)

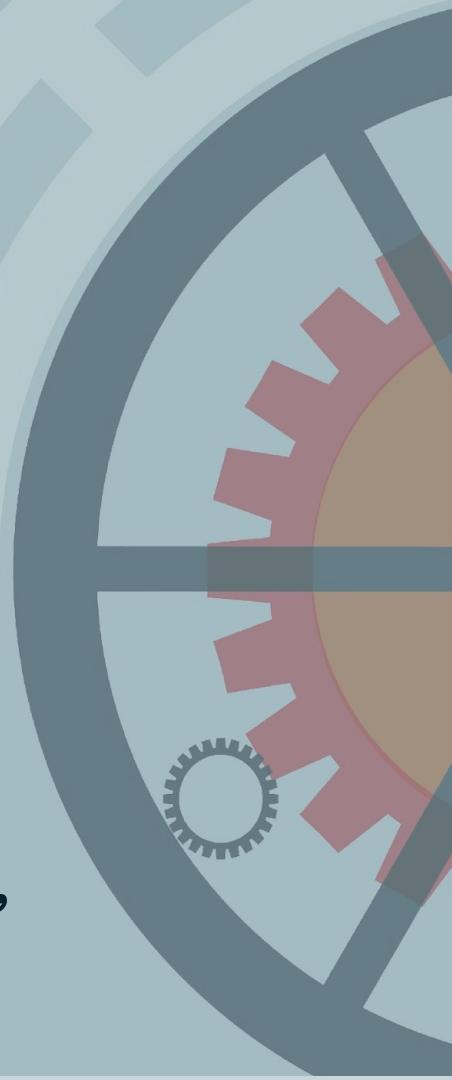
**~15K sheets**

# **Make trustworthy claims based on *thousands* of maps**

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

**aggregated statistics** → ‘low resolution’ regional/national

**Ordnance Survey (series) maps as non-aggregated, high-resolution, national-coverage sources**



# You Cannot Ground Truth the Past

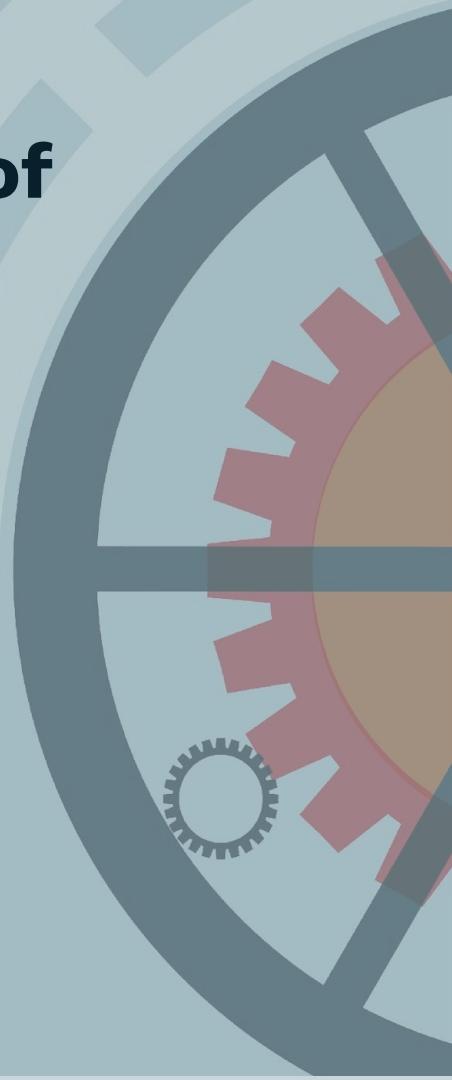
Historical OS maps tell us how Victorians represented the British landscape, and how that landscape was changing, but they are not a ‘ground truth’.

We want to ***use CV to advance interpretation, not define truth.***

# **Step away from the GIS paradigm of vectorizing everything on a map**

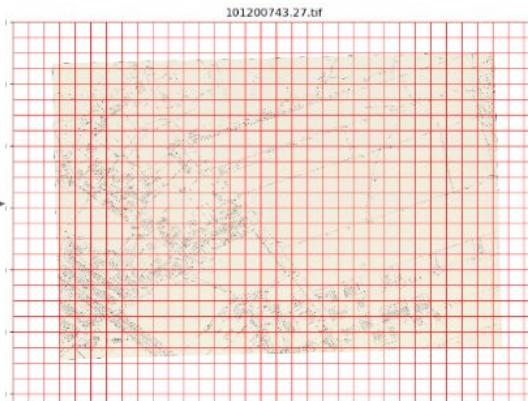
**Stop** historical **data “mining”/“extraction”**  
and insisting on overly-precise data

**Start** developing methods that permit  
**critical distance between scholars and sources** & question-appropriate data collection and exploration

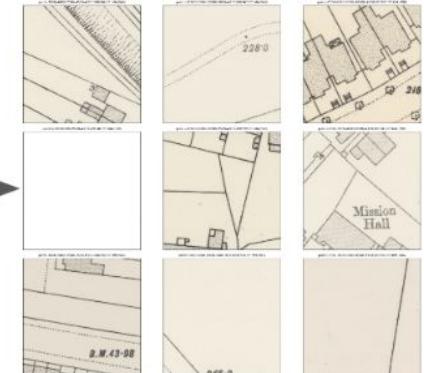


# Solution: ‘Patches’ as a new shape for historical research

Parent image



Patches



# Image Classification for Map Patches

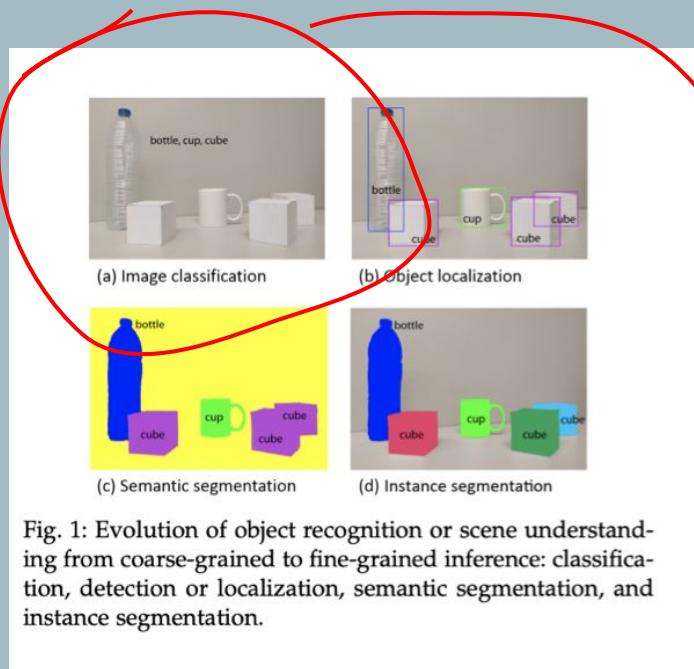
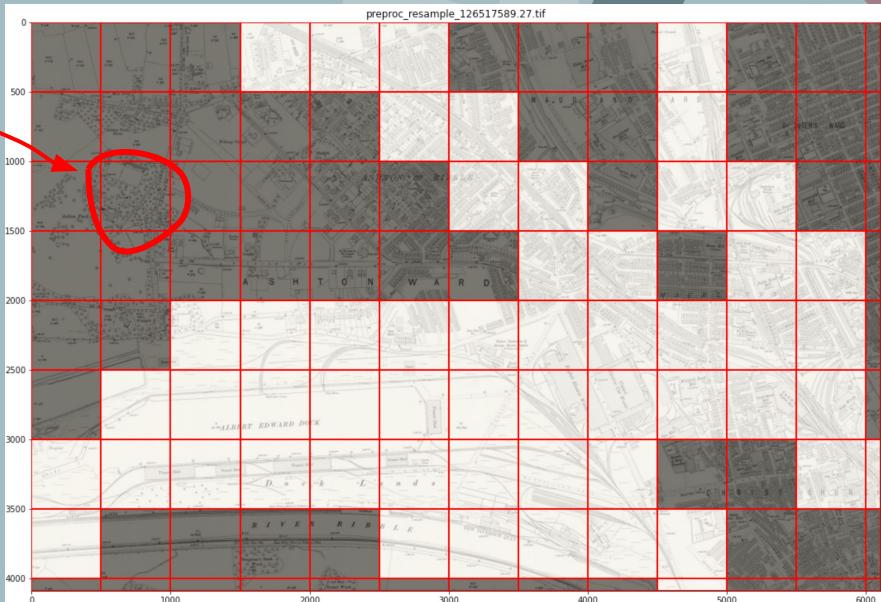
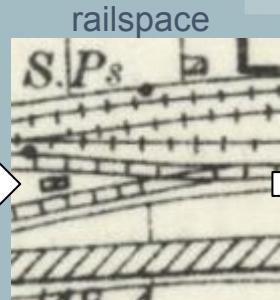
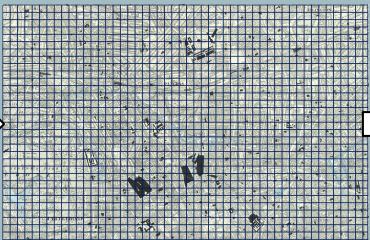
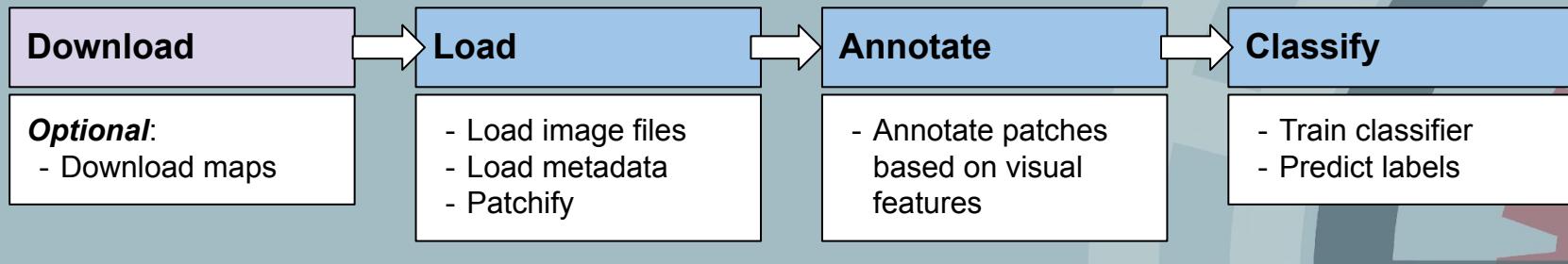


Fig. 1: Evolution of object recognition or scene understanding from coarse-grained to fine-grained inference: classification, detection or localization, semantic segmentation, and instance segmentation.

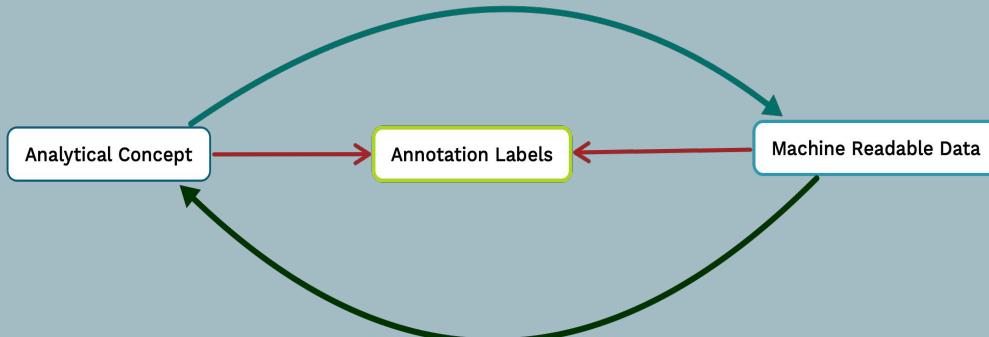


Map patches annotated as training data

# Classification pipeline



# Annotating patches: What is a good label?



Rail Space <sup>1</sup> No Rail Space <sup>2</sup> ← back <sup>j</sup> → next <sup>k</sup>

<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 buttons for 'Rail Space <sup>1</sup>' (green), 'No Rail Space <sup>2</sup>' (blue), '← back <sup>j</sup>', and '→ next <sup>k</sup>'. Below this is a caption: '<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 a detailed map of a railway station area with tracks, platforms, and surrounding buildings, labeled with 'B.M. 981', 'PRIORITY', 'ABBREV.', and 'INGOT ST.'. The bottom version is a coarser, lower-resolution version of the same area. A red rectangular box highlights a specific section of the map, likely indicating a region of interest or a difference between the two versions.

# What was it like to live *with* railways?

## Railspace as the total footprint of rail infrastructure

- Railspace label inferred on ~15K late nineteenth-century 6" OS maps (NLS)
- ~30.5 million 100x100m patches
- ~62k expert-annotated patch dataset now available on Zenodo/[Hugging Face](#)

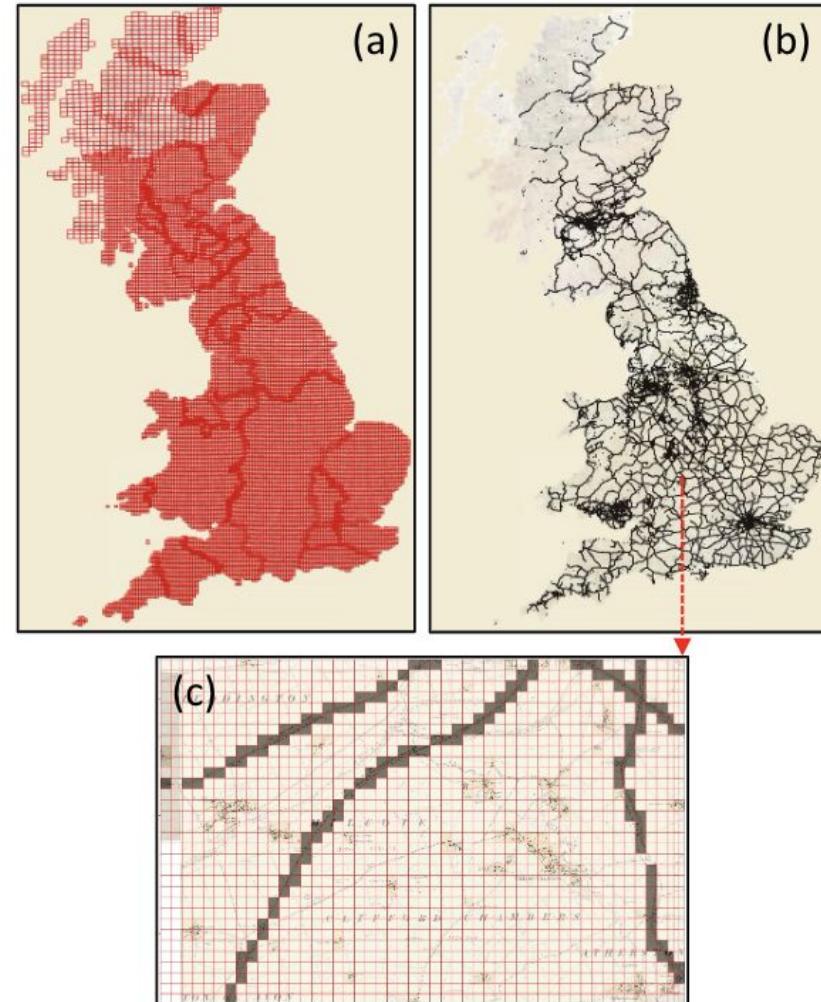
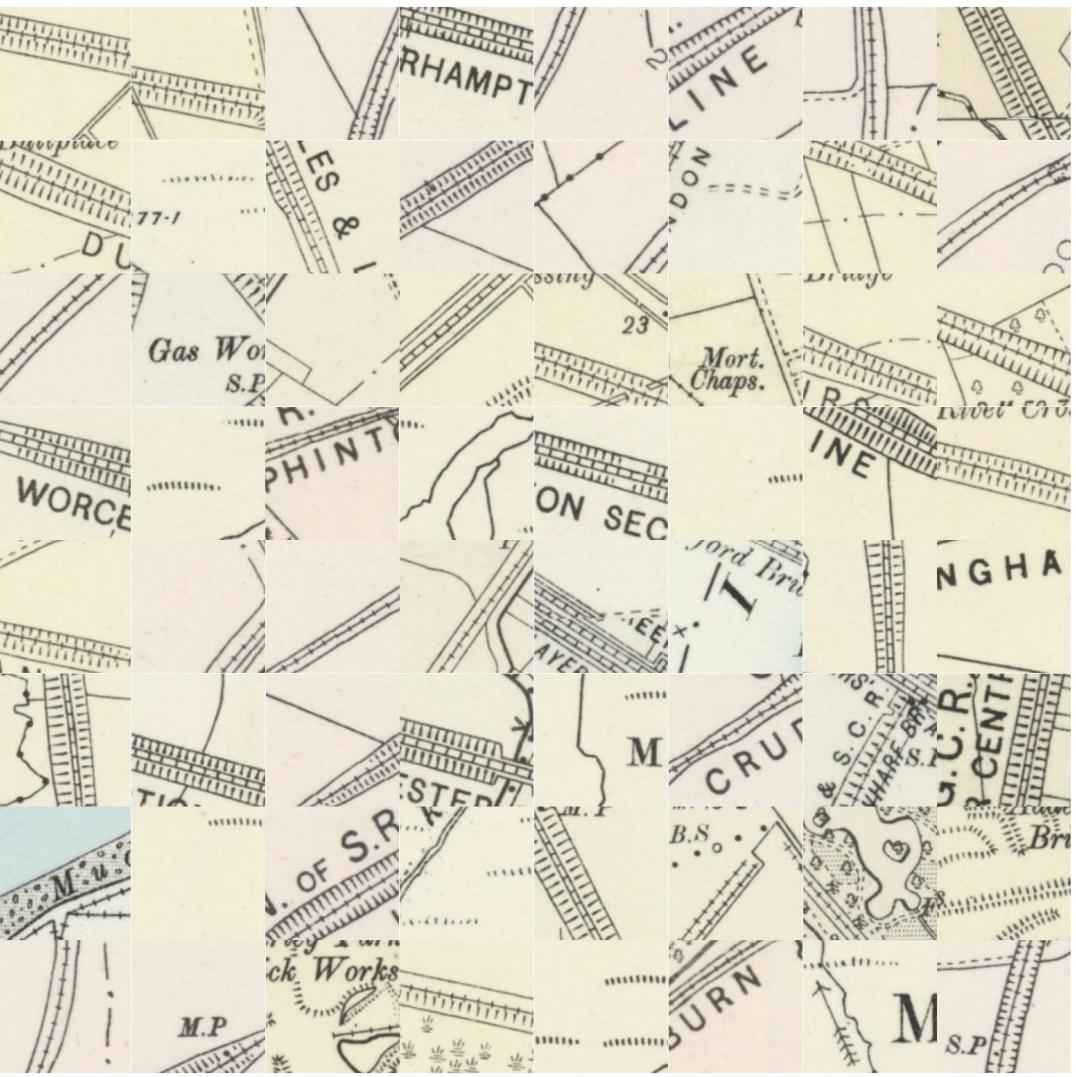


Image credit: Kasra Hosseini

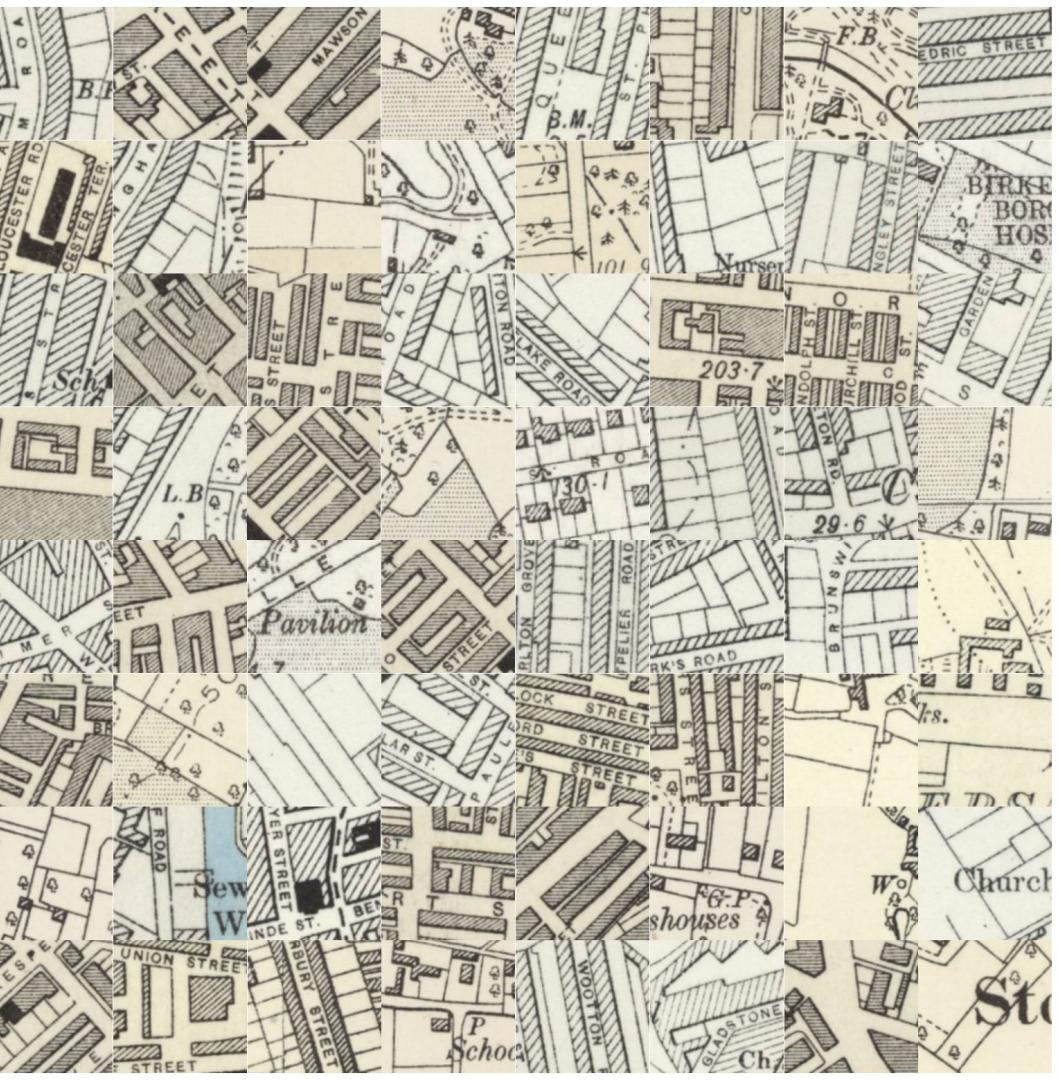
# Railspace (rural)



# Railspace (urban)



# Buildings





## Living with Machines

*Computational Histories of the Age of Industry*

by Ruth Ahnert, Emma Griffin, Jon Lawrence,  
The Living with Machines Team

*Living With Machines* is a data-driven history of the coming of the machine age in Britain in the long nineteenth century. Featuring an innovative open access edition enhanced with interactive maps, datasets and visualisations, digital notebooks, video, audio and images, this book harnesses the combined power of massive digitised historical collections and computational analytical tools to examine the ways in which technology altered the very fabric of human existence on a hitherto unprecedented scale.



ABOUT DIGITAL EDITIONS

**Beyond the Tracks: re-connecting people, places and stations in the history of late-Victorian railways**

Joshua Rhodes, Jon Lawrence, Kaspar Beelen, Katherine McDonough, Daniel C.S. Wilson

# MapReader in 2025

MapReader performs 2 tasks:

1. Classify map regions (patches) with concepts that have a **visual** signal
2. Find and transcribe **text**

[github.com/maps-as-data/  
MapReader](https://github.com/maps-as-data/MapReader)

The screenshot shows the GitHub repository page for 'maps-as-data / MapReader'. The main area displays a list of commits, showing frequent updates to files like '.github', 'conda', 'docs', 'mapreader', 'paper', 'tests', and 'worked\_examples'. The sidebar on the right provides details about the repository, including its purpose ('A computer vision pipeline for exploring and analyzing images at scale'), tags ('machine-learning', 'computer-vision', 'deep-learning', 'article', 'maps', 'pytorch', 'digital-humanities', 'spatial-data', 'hut23', 'hut23-96'), and metrics like 97 stars, 7 watching, and 13 forks.

**About**  
A computer vision pipeline for exploring and analyzing images at scale  
mapreader.readthedocs.io/en/latest/  
machine-learning computer-vision  
deep-learning article maps  
pytorch digital-humanities  
spatial-data hut23 hut23-96

**Code**  
main 15 Branches 37 Tags Go to file Add file Code

**Commits**

- rwood-97 update changelog 16 hours ago 1,872 Commits
- .github Update publish-to-conda-forge.yml 3 weeks ago
- conda v. minor tidy-up of conda specification 2 weeks ago
- docs Merge branch 'main' into text\_dev 2 weeks ago
- mapreader Merge branch 'main' into text\_dev 2 weeks ago
- paper Update paper.md 4 months ago
- tests Merge branch 'main' into text\_dev 2 weeks ago
- worked\_examples update worked examples and two minor fixes 3 months ago
- .all-contributorsrc docs: update .all-contributorsrc [skip ci] 2 weeks ago
- .gitattributes Add versioneer 2 years ago
- .gitignore add worked example outputs to gitignore 3 months ago
- .pre-commit-config.yaml update python version in files 10 months ago
- .readthedocs.yaml pre-commit run all 2 years ago
- .ruff.toml update python version in files 10 months ago
- CHANGELOG.md update changelog 16 hours ago
- CITATION.cff Update CITATION.cff 2 years ago
- CODE\_OF\_CONDUCT.md Create CODE\_OF\_CONDUCT.md 3 years ago
- LICENSE americanize docs and readme 2 years ago
- README.md docs: update README.md [skip ci] 2 weeks ago
- contributors.md Rename Contributors.md to contributors.md 2 years ago
- setup.cfg changes versioneer style 2 years ago
- setup.py Merge branch 'main' into improve\_show 3 months ago
- text-requirements.txt reorder text-requirements.txt 3 months ago
- versioneer.py run all 2 years ago

**Releases**  
v1.3.10 - maps-as-data/MapR... [Latest] on Sep 3, 2024 + 8 releases

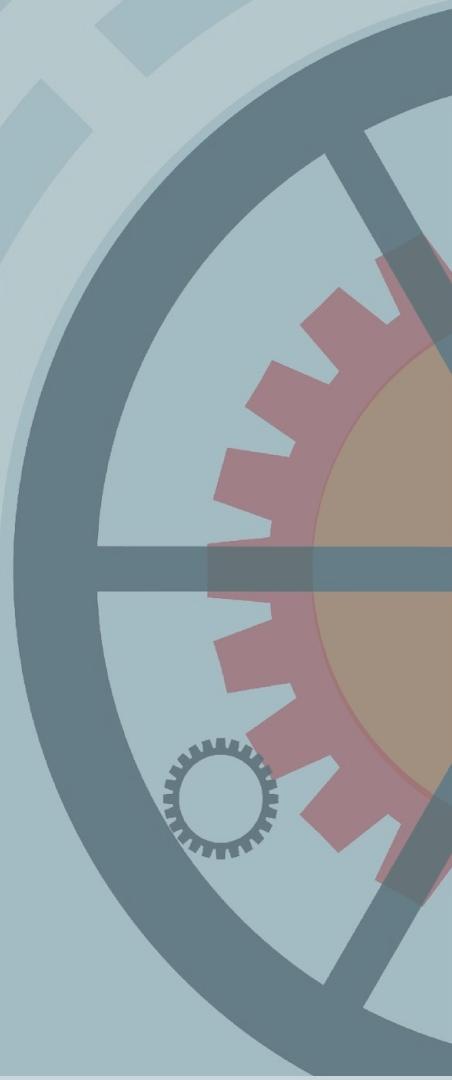
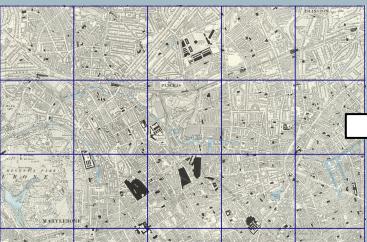
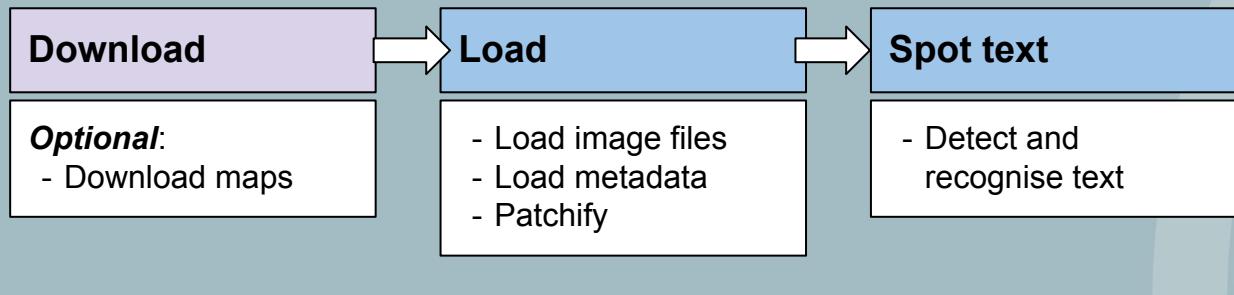
**Packages**  
No packages published Publish your first package

**Contributors** 15

**Deployments** 328  
github-pages 16 hours ago + 324 deployments

**Languages**  
Jupyter Notebook 99.3% Other 0.7%

# Text spotting pipeline



# **MapReader & AI**

**MapReader is a sandpit for thinking with map content and metadata to critically create structured data for onward linking and analysis.**

**We have privileged software design and methods that:**

- Illuminate understudied cartographic elements, like text**
- Encourage interpretation over extraction, with patches**
- Foster stronger connections with collections & curators**

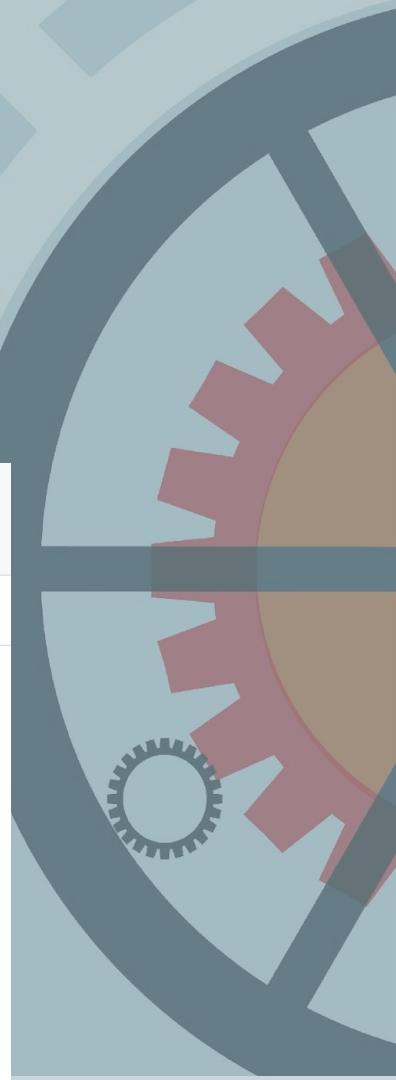
**<https://github.com/maps-as-data/MapReader>**

# Using MapReader



# MapReader Github Repository

<https://github.com/maps-as-data/MapReader>



Screenshot of the MapReader GitHub repository page.

**Code** | Issues 85 | Pull requests 5 | Discussions | Actions | Projects 3 | Wiki | Security | Insights | Settings

**MapReader** Public

main ▾ 15 Branches 37 Tags

Go to file

kmcdono2 add modelcard template · 4 hours ago · 1,876 Commits

.github Update publish-to-conda-forge.yml · last month

conda v. minor tidy-up of conda specification · 2 years ago

docs Merge branch 'main' into text\_dev · last month

mapreader fix explore methods · last month

paper Update paper.md · 4 months ago

tests Merge branch 'main' into text\_dev · last month

About

A computer vision pipeline for exploring and analyzing images at scale

mapreader.readthedocs.io/en/latest/

machine-learning computer-vision  
deep-learning article maps  
pytorch digital-humanities  
spatial-data hut23 hut23-96

Readme  
View license  
Code of conduct

# MapReader Documentation

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

The screenshot shows the MapReader documentation website. At the top left is the MapReader logo and a search bar. To the right is the page title "MapReader" and a "Edit on GitHub" link. The main content area features a large heading "MapReader" and a sub-section "What is MapReader?". Below this is a paragraph describing MapReader as an end-to-end computer vision pipeline for exploring and analyzing images at scale. To the right of the text is a map of London with red and purple overlays. At the bottom of the main content area is a section titled "Overview". A sidebar on the left contains a navigation menu with links to "Introduction to MapReader", "Getting Started", "Using MapReader", "In-Depth Resources", and "Community and contributions". Below the menu is an "On-demand webinar" advertisement for "TIDELIFT". The advertisement includes a play button icon, the text "How to reduce your organization's reliance on 'bad' open source packages.", a "WATCH NOW" button, and the "TIDELIFT" logo. At the very bottom of the sidebar is the text "Ad by EthicalAds".

MapReader

latest

Search docs

Introduction to MapReader

Getting Started

Using MapReader

In-Depth Resources

Community and contributions

ON-DEMAND WEBINAR

How to reduce your organization's reliance on "bad" open source packages.

WATCH NOW

TIDELIFT

On-demand webinar: Reduce your organization's reliance on "bad" open source packages. WATCH NOW!

Ad by EthicalAds

## MapReader

### What is MapReader?

MapReader is an end-to-end computer vision (CV) pipeline for exploring and analyzing images at scale.

## Overview

MapReader was developed in the [Living with Machines](#) project to analyze large collections of historical maps but is a **generalizable** computer vision pipeline which can be applied to *any images* in a wide variety of domains.

MapReader is a groundbreaking interdisciplinary tool that emerged from a specific set of geospatial

<https://github.com/maps-as-data/edhs-jan-2025>

maps-as-data / edhs-jan-2025

Type  to search

Code Pull requests Actions Projects Security Insights Settings

**edhs-jan-2025** Public  
forked from [maps-as-data/scai-mapreader-jan-2025](#)

Edit Pins Watch 0 Fork 1 Star 0

main 1 Branch 0 Tags Go to file Add file Code

This branch is 3 commits ahead of maps-as-data/scai-mapreader-jan-2025:main.

Contribute Sync fork

**rwood-97** update screenshots 514ddcf · 3 hours ago 44 Commits

images	update screenshots	3 hours ago
maps	add maps and metadata	5 months ago
.gitignore	update gitignore and saving geojson	4 months ago
Dockerfile	add dockerfile and docker instructions	4 months ago
README.md	update for edhs workshop	20 hours ago
mapreader_edhs_jan_2025.ipynb	fix plot_metrics	3 hours ago
requirements.txt	test just mapreader in requirements.txt	5 months ago

About

A repo for the EDHS MapReader session

Readme Activity Custom properties 0 stars 0 watching 1 fork Report repository

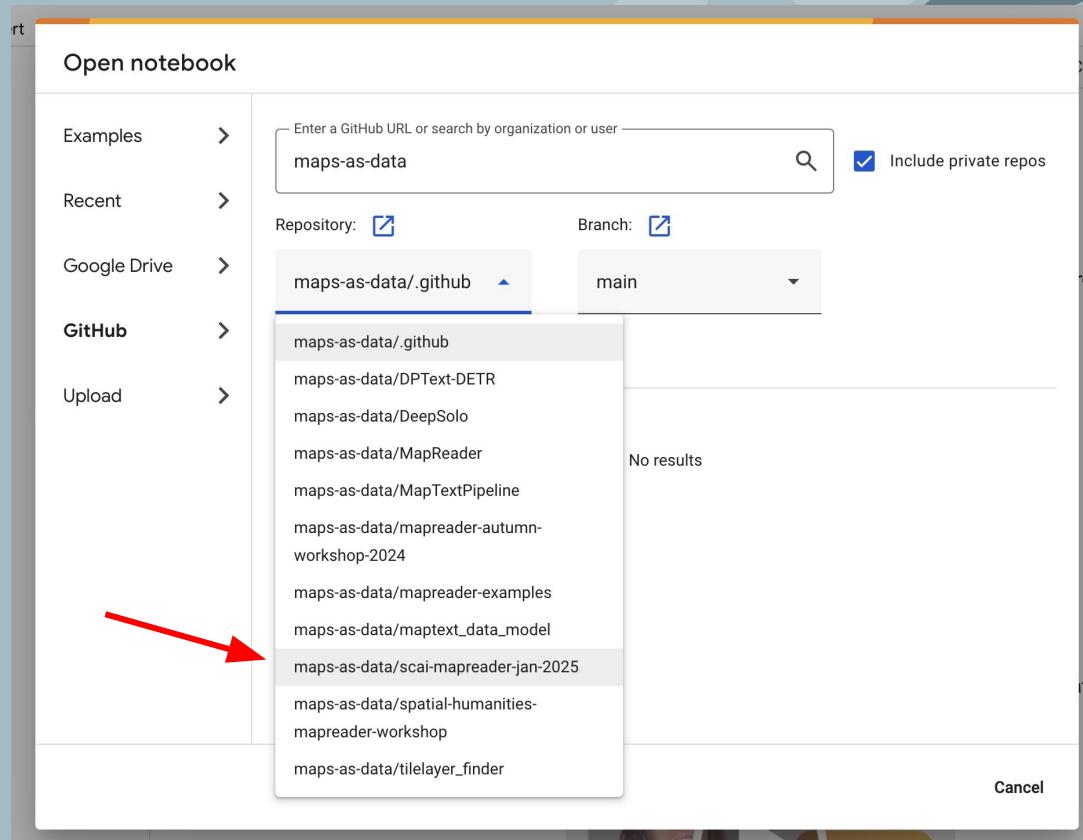
Releases

No releases published Create a new release

Packages

# Setting Up

Follow the  
instructions for  
Google Colab in the  
README



# Jupyter Notebook Basics

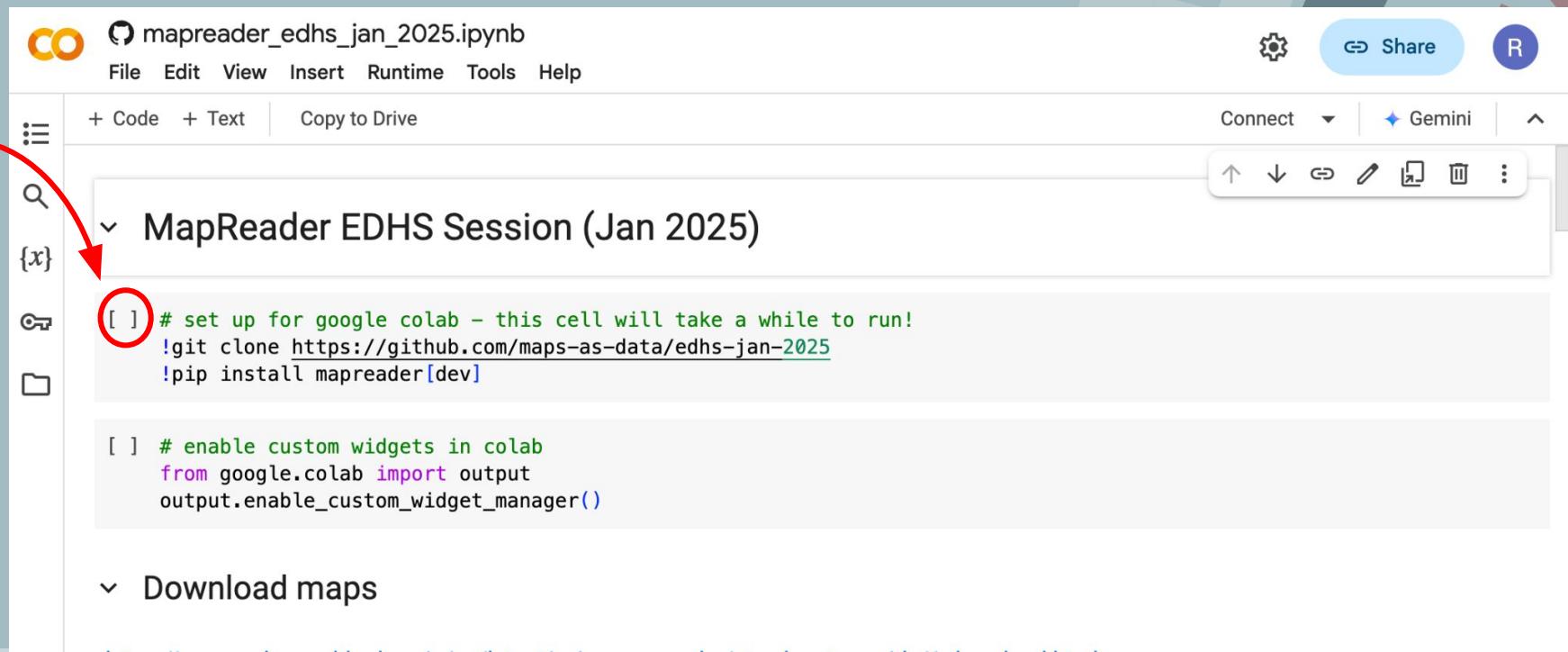
- Two different cell types - **Code or Text**
- Running cells - Press “▶” or  
“⌘/Ctrl/Shift+Enter”

**WARNING: Google Colab will time out after a while  
so make sure you save your work!**

# Notebook Sections

- **(No need to download maps, we have provided them!)**
- **Install dependencies**
- **Load & Patchify maps**
- **Annotate patches with your label of choice**
- **Fine-tune a model**
- **Review model performance**
- **Infer labels on remaining patches**
- **Export labeled patch data**

# Open notebook & run first code cell



The screenshot shows a Google Colab notebook titled "mapreader\_edhs\_jan\_2025.ipynb". The interface includes a toolbar with icons for file operations, a search bar, and a runtime selector. The notebook content is organized into sections: "MapReader EDHS Session (Jan 2025)" and "Download maps". The first code cell, which sets up the environment for the session, is highlighted with a red circle and a red arrow pointing to it from the left sidebar.

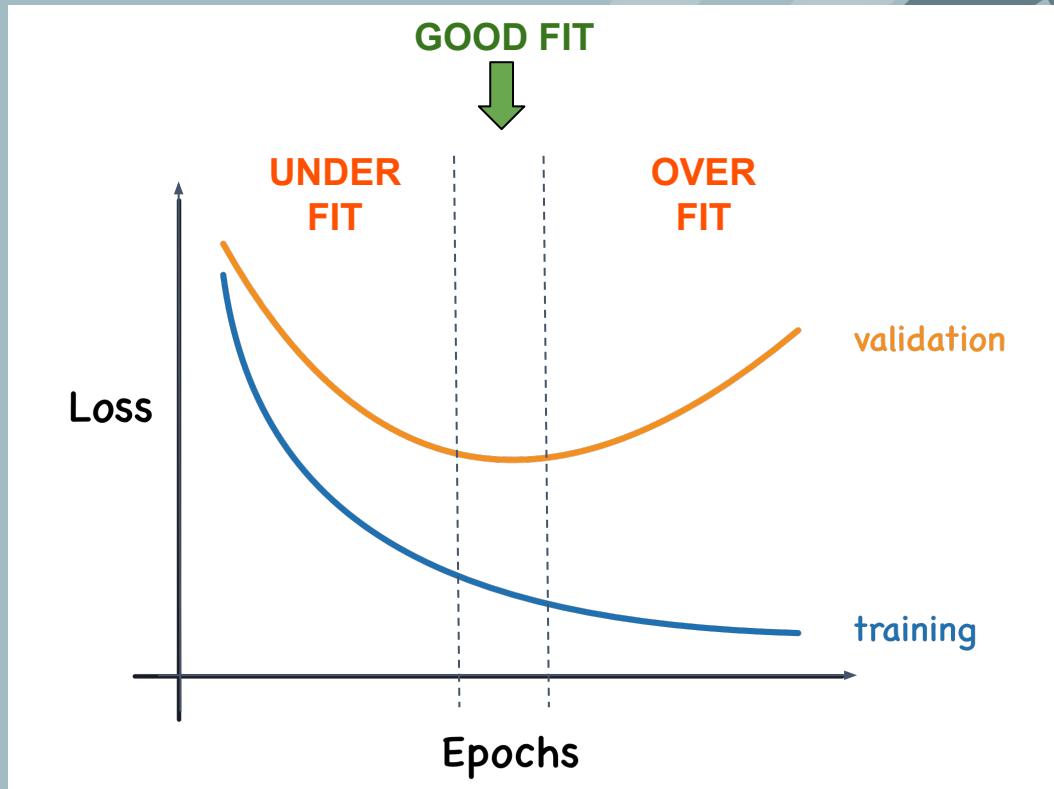
```
[ ] # set up for google colab - this cell will take a while to run!
!git clone https://github.com/maps-as-data/edhs-jan-2025
!pip install mapreader[dev]

[ ] # enable custom widgets in colab
from google.colab import output
output.enable_custom_widget_manager()
```

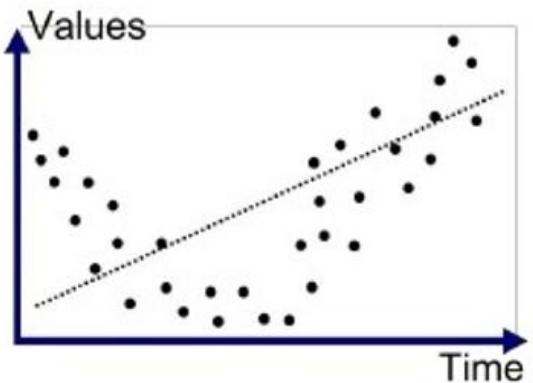
# Training

Annotations split into train, validation and test datasets

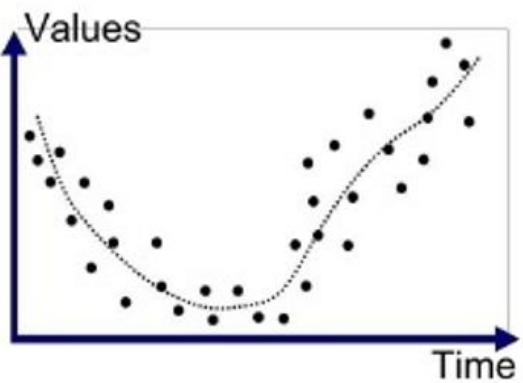
Goal is to minimise both training and validation losses



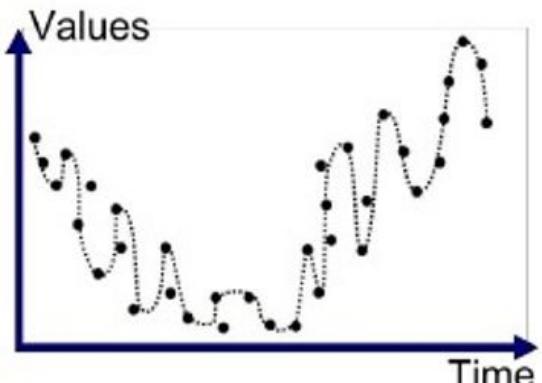
# Training



Underfitted



Good Fit/Robust



Overfitted

# Metrics

- Loss - defined by loss function
- Precision - **proportion of positive predictions that are correct**

$$\text{Precision} = \frac{TP}{TP + FP}$$



		PREDICTED VALUES	
		POSITIVE	NEGATIVE
ACTUAL VALUES	POSITIVE	TRUE POSITIVE	FALSE NEGATIVE
	NEGATIVE	FALSE POSITIVE	TRUE NEGATIVE

# Metrics

- Loss - defined by loss function
- Precision - proportion of positive predictions that are correct
- Recall - **proportion of actual positives correctly identified**

$$\text{Recall} = \frac{TP}{TP + FN}$$

		PREDICTED VALUES	
		POSITIVE	NEGATIVE
ACTUAL VALUES	POSITIVE	TRUE POSITIVE	FALSE NEGATIVE
	NEGATIVE	FALSE POSITIVE	TRUE NEGATIVE

# Metrics

- Loss - defined by loss function
- Precision - proportion of positive predictions that are true
- Recall - proportion of actual positives correctly identified
- F-scores - **combination of precision and recall**

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

		PREDICTED VALUES	
		POSITIVE	NEGATIVE
ACTUAL VALUES	POSITIVE	TRUE POSITIVE	FALSE NEGATIVE
	NEGATIVE	FALSE POSITIVE	TRUE NEGATIVE

# More on training CV models

Try these *Programming Historian* Lessons:

[https://programminghistorian.org/en/lessons  
/computer-vision-deep-learning-pt1](https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt1)

[https://programminghistorian.org/en/lessons  
/computer-vision-deep-learning-pt2](https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt2)

# More MapReader example notebooks!

<https://github.com/maps-as-data/MapReader-examples>

The screenshot shows a GitHub repository page for 'mapreader-examples'. The repository is public and has 1 branch and 0 tags. The commit history is listed below, showing updates from 'rwood-97' over the last month. The README file is also visible at the bottom.

Author	Commit Message	Date
rwood-97	Update check_notebooks.yml	025f4f2 · last month
.github/workflows	Update check_notebooks.yml	last month
notebooks	update deprecated method	last month
.dockerignore	rename worked_examples to notebooks	2 months ago
.gitignore	rename worked_examples to notebooks	2 months ago
Dockerfile	install libgl1	2 months ago
README.md	rename worked_examples to notebooks	2 months ago
requirements.txt	Update requirements.txt	2 months ago

**MapReader Examples**

This repo contains worked examples of using the MapReader library for various geospatial and non-geospatial tasks.

# Stay in touch: Community Calls

<https://mapreader.readthedocs.io/en/latest/community-and-contributions/index.html>

- **Connect with the Team:** Ask questions and get real-time updates from MapReader developers.
- **Learn Best Practices:** Pick up tips and workflows from experts and users alike.
- **Hear Success Stories:** Discover how others are using MapReader in their research.
- **Influence Development:** Share your ideas to help shape future features.
- **Build Community:** Network with fellow researchers and developers.

The screenshot shows the MapReader documentation website. At the top, there's a navigation bar with a home icon, 'MapReader', and 'latest'. Below it is a search bar labeled 'Search docs'. The main content area has a dark sidebar on the left with links to 'Introduction to MapReader', 'Getting Started', 'Using MapReader', and 'In-Depth Resources'. The main content area is titled 'Community and contributions' and contains links to 'Joining the community', 'Events', 'Share Your MapReader Story', 'Contribution Guide', and 'Code of Conduct and Inclusivity'. A large black rectangular area is positioned below the contribution guide and code of conduct sections.

[Home](#) / Community and contributions

## Community and contributions

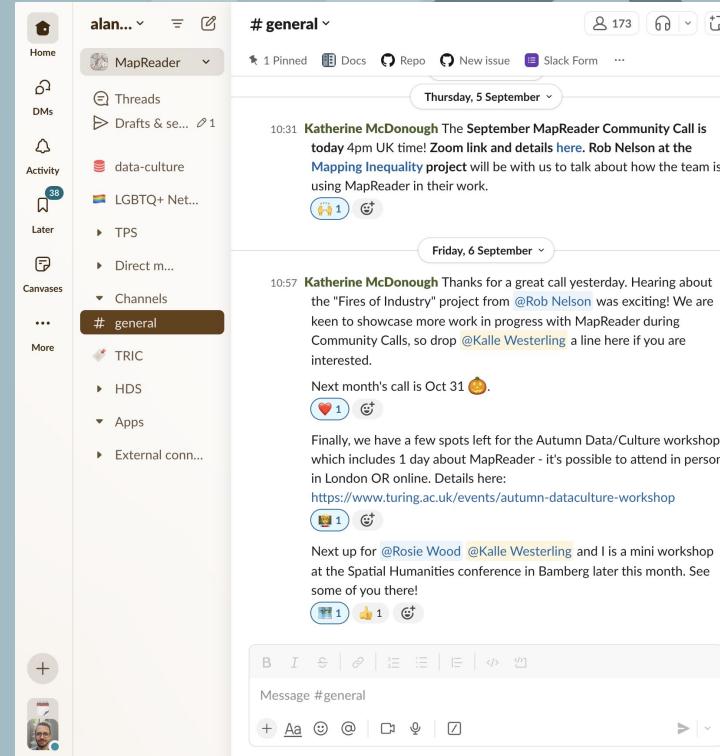
- [Joining the community](#)
- [Events](#)
  - [Community Calls](#)
  - [Workshops](#)
- [Share Your MapReader Story](#)
  - [Why Share Your Story?](#)
  - [How to Share Your Story](#)
  - [Examples of User Stories](#)
- [Contribution Guide](#)
  - [Pre-requisites](#)
  - [Ways to contribute](#)
- [Code of Conduct and Inclusivity](#)

[Previous](#)

# Stay in touch: Slack

## MapReader slack sign up form

- **Get Help Fast:** Ask questions and get immediate support from the team and community.
- **Stay Informed:** Hear about new releases, features, and events first.
- **Share Ideas:** Discuss, learn, and exchange real-world use cases with others.
- **Shape the Future:** Contribute feedback, suggestions, or code to improve MapReader.
- **Connect with Experts:** Network with researchers and developers using MapReader.



# Stay in touch: Github

<https://github.com/maps-as-data/MapReader>

- **Share Feedback:** Report bugs, suggest features, or give general input.
- **Improve Docs:** Help make guides and tutorials clearer by raising issues.
- **Track Progress:** Follow development updates in GitHub discussions.
- **Contribute Code:** Submit pull requests to add features or fix bugs. (For those who are more code-inclined!)
- **Collaborate:** Join discussions to shape the future of MapReader.

The screenshot shows the GitHub repository page for 'maps-as-data / MapReader'. The main header shows 81 issues, 4 pull requests, and 1 discussion. Below the header, the 'Code' tab is selected, showing a list of branches and commits. The 'main' branch has 13 branches and 33 tags. A commit by 'rwood-97' titled 'update changelog' is highlighted, showing it merged 'main' into 'test\_text\_spotting' two weeks ago. Other commits include 'v. minor tidy-up of conda specification' (last year), 'add docs for searching text' (2 weeks ago), and 'use tqdm auto' (last week). The sidebar on the right provides links to 'mapreader.readme', 'machine-learning', 'art', 'pytorch', 'digital-hu', 'spatial-data', 'Readme', 'View license', 'Code of conduct', 'Cite this repository', 'Activity', 'Custom properties', '80 stars', '8 watching', '11 forks', 'Report repository', 'Releases', 'v1.3.0 - maps-a...', '+ 8 releases', 'Packages', 'No packages published', 'Publish your first package', and 'Contributors'.

# Thank you!

