# *"Open by Default"*

## Developing reproducible, computational research


Dani Arribas-Bel

@darribas

Martin Fleischmann

@martinfleis

UNIVERSITY OF LIVERPOOL

The Alan Turing Institute

Geographic Data Science Lab

# *Open research*

- Open! Open? "Open"; *Open*; Open
    1. Transparent
    2. Accessible
    3. Transferrable
- Open data is *not* necessary, *nor* sufficient

# *"Open by default"*

- Like happiness, it's about the journey not (only) the destination
- Treat research artifacts as OSS from Day 1
- Use (and nurture) the *existing* OSS infrastructure

# For example

# Urban Grammar

## Learning an urban grammar from satellite data through AI

Learn more ↓

Project status
**Ongoing**

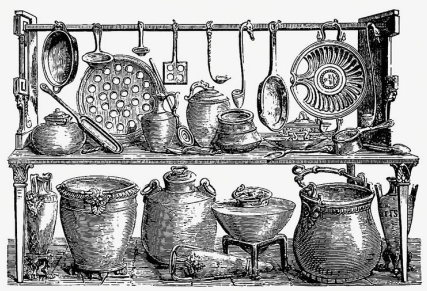
Related programmes
**Urban analytics**

# Urban Grammar

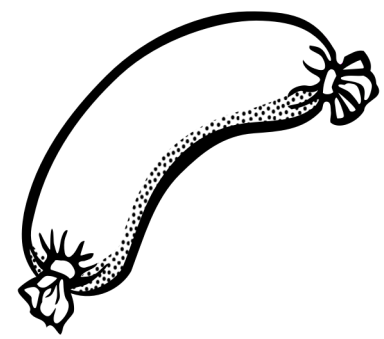
*A characterisation of space based on form and function designed to understand urban environments*

\+ Satellite imagery & AI

# Needs



- Data, data, data
- (New) methods, (new) code
- Infrastructure



- (Academic) dissemination
- Open Data Products
- Journaling/reporting

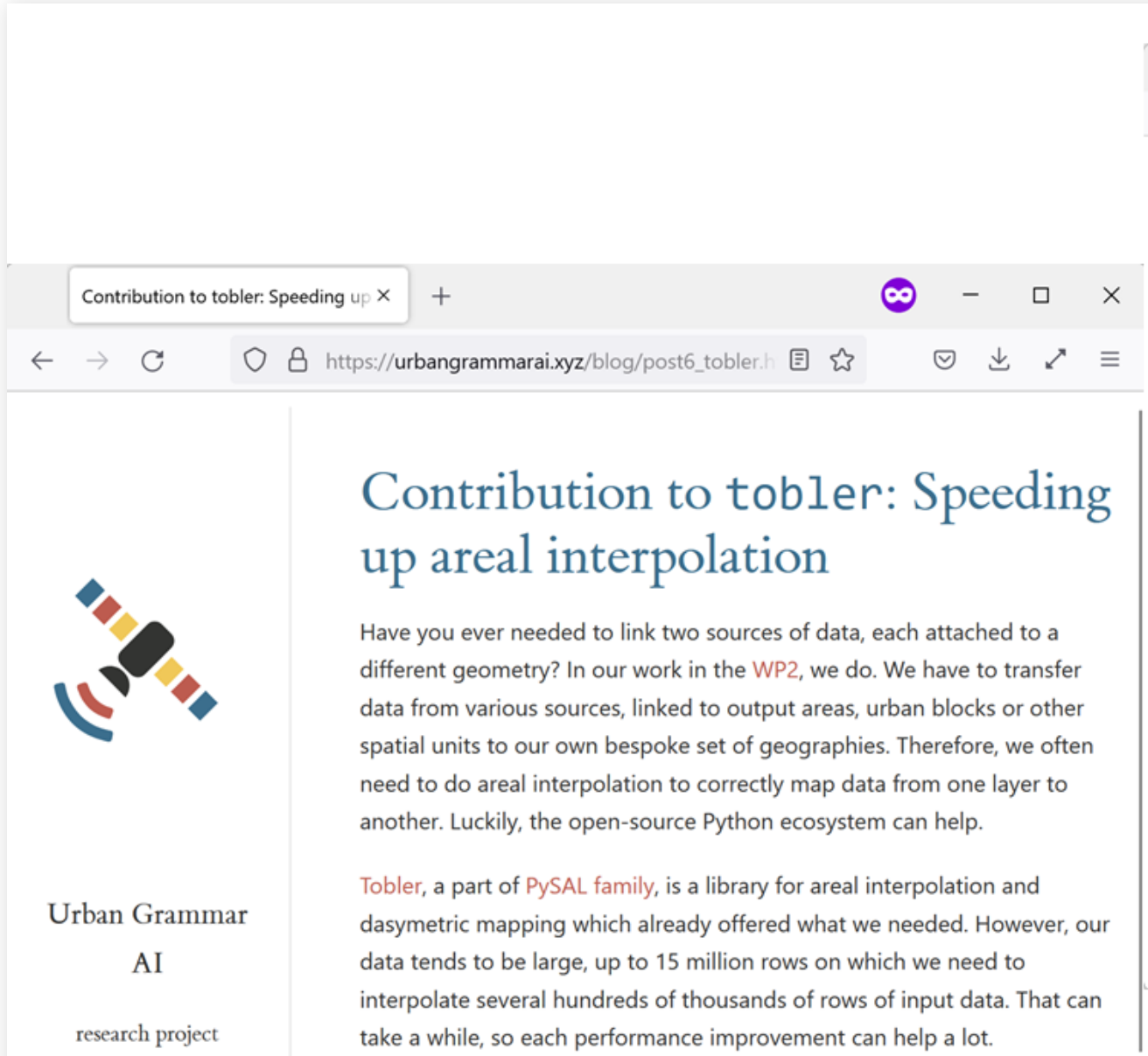
# The kitchen

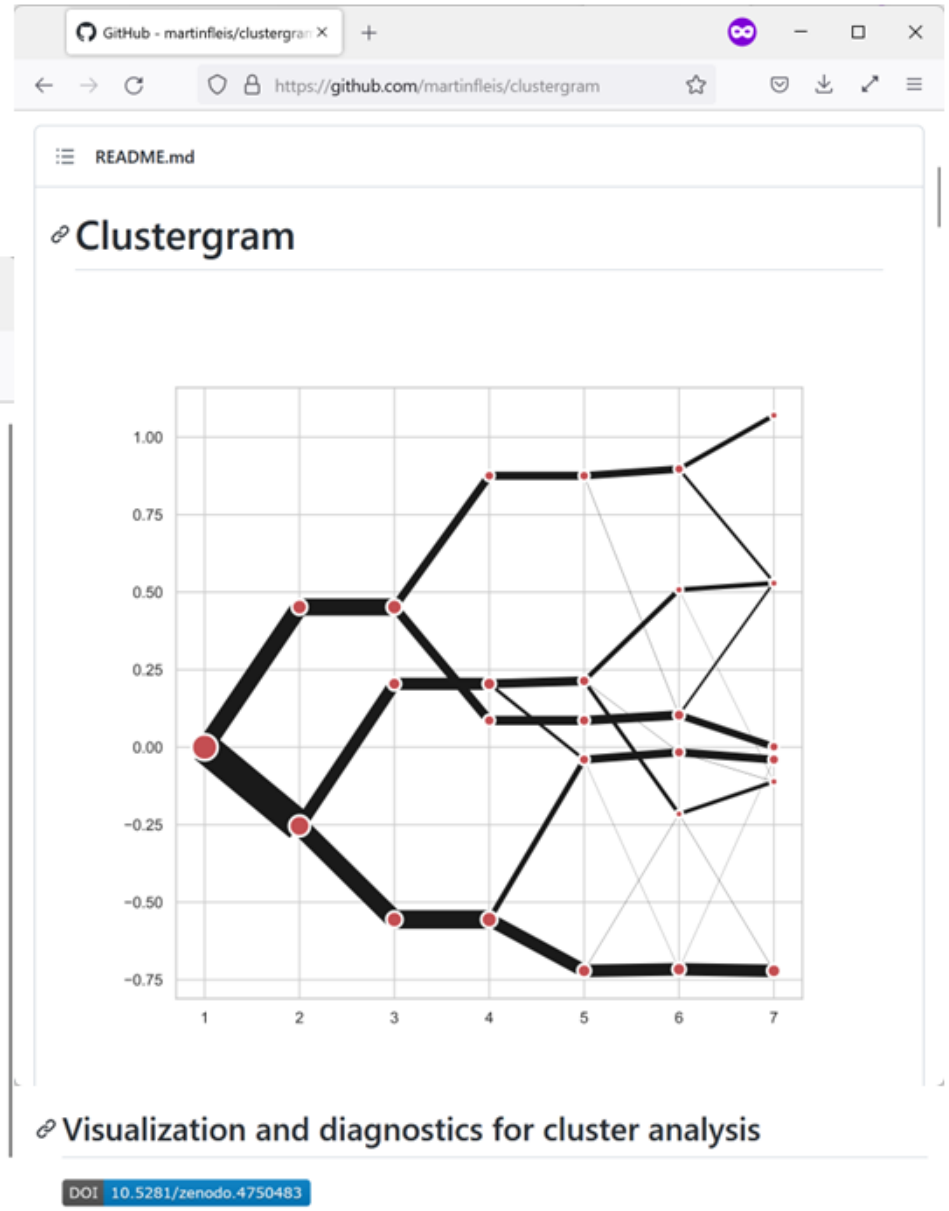
# Data

- Standard (formats) better than niche
- Open better than proprietary
- Static better than DB

E.g., parquet, COGs

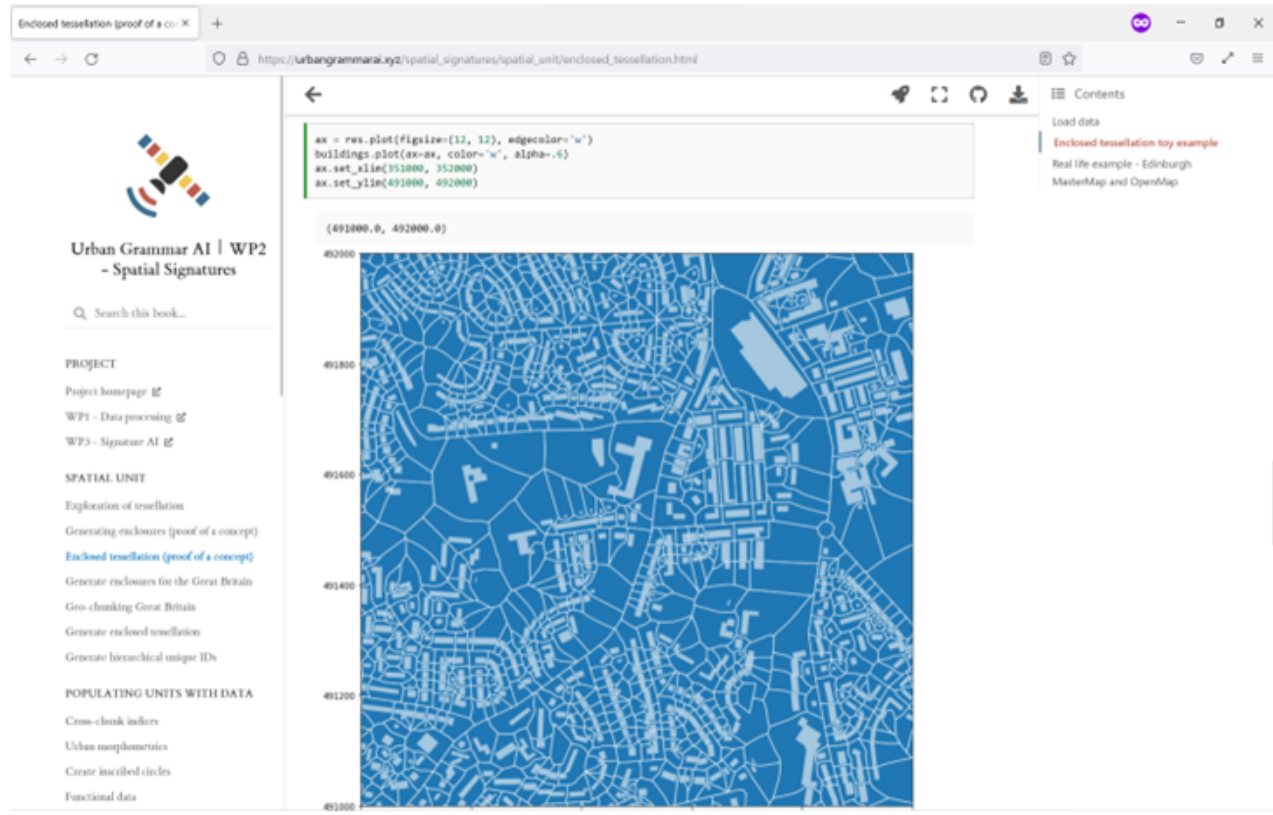
# New methods


Contribution to tobler: Speeding up
https://urbangrammarai.xyz/blog/post6_tobler.h
Urban Grammar AI
research project
Contribution to tobler: Speeding up areal interpolation
Have you ever needed to link two sources of data, each attached to a different geometry? In our work in the WP2, we do. We have to transfer data from various sources, linked to output areas, urban blocks or other spatial units to our own bespoke set of geographies. Therefore, we often need to do areal interpolation to correctly map data from one layer to another. Luckily, the open-source Python ecosystem can help.
Tobler, a part of PySAL family, is a library for areal interpolation and dasymetric mapping which already offered what we needed. However, our data tends to be large, up to 15 million rows on which we need to interpolate several hundreds of thousands of rows of input data. That can take a while, so each performance improvement can help a lot.

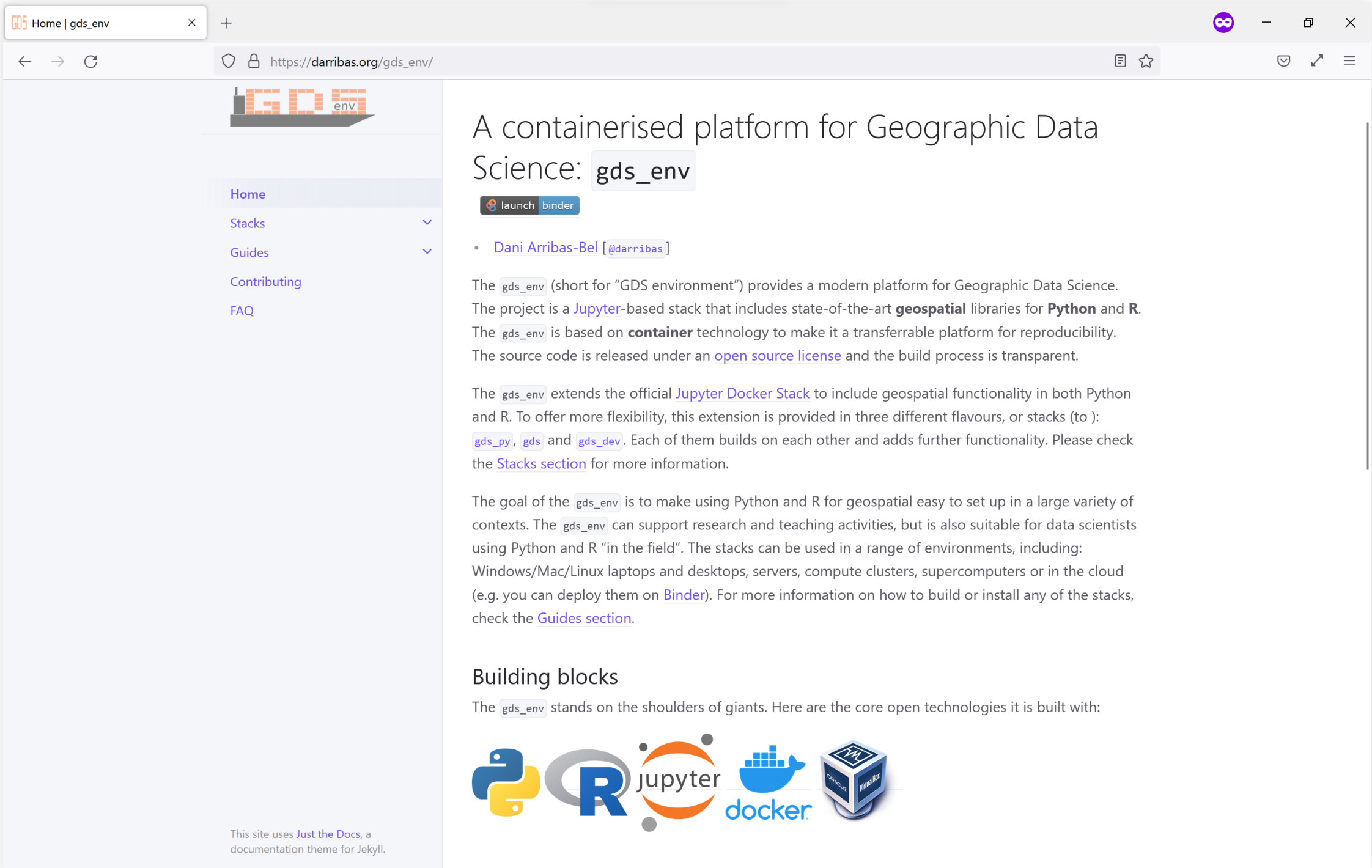

GitHub - martinfleis/clustergram
https://github.com/martinfleis/clustergram
README.md
Clustergram
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
1
2
3
4
5
6
7
Visualization and diagnostics for cluster analysis
DOI 10.5281/zenodo.4750483

# New code

python
jupyter
jupyter{book}

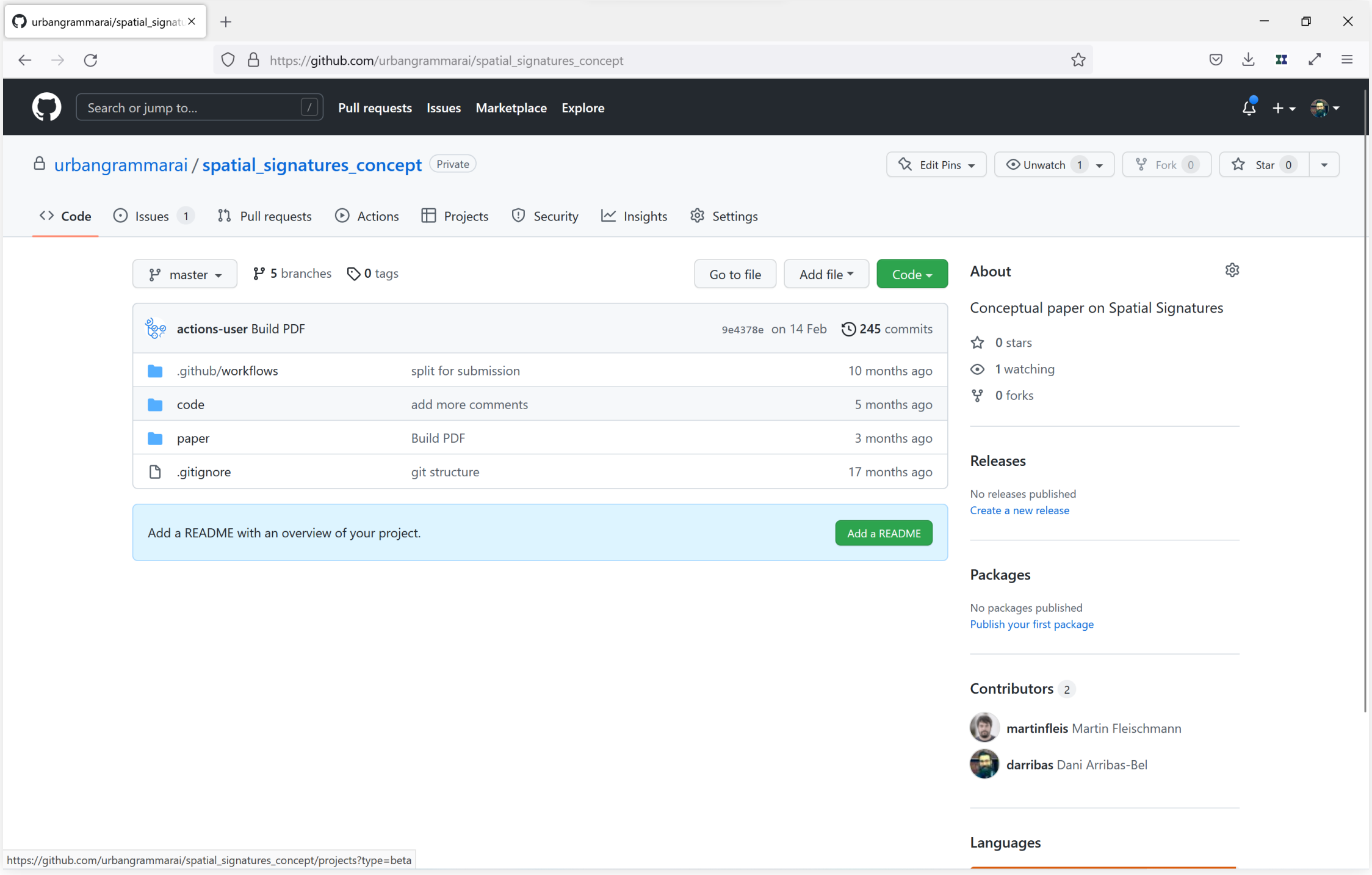

Urban Grammar AI | WP2 - Spatial Signatures
Search this book...
PROJECT
Project homepage
WP1 - Data processing
WP3 - Signature AI
SPATIAL UNIT
Exploration of tesellation
Generating enclosures (proof of a concept)
Enclosed tesellation (proof of a concept)
Generate enclosures for the Great Britain
Geo-chunking Great Britain
Generate enclosed tesellation
Generate hierarchical unique IDs
POPULATING UNITS WITH DATA
Cross-chunk indices
Urban morphometrics
Create inscribed circles
Functional data
Contents
Load data
Enclosed tesellation toy example
Real life example - Edinburgh
MasterMap and OpenMap
(491000.0, 492000.0)
492000
491800
491600
491400
491200
491000

# Infrastructure

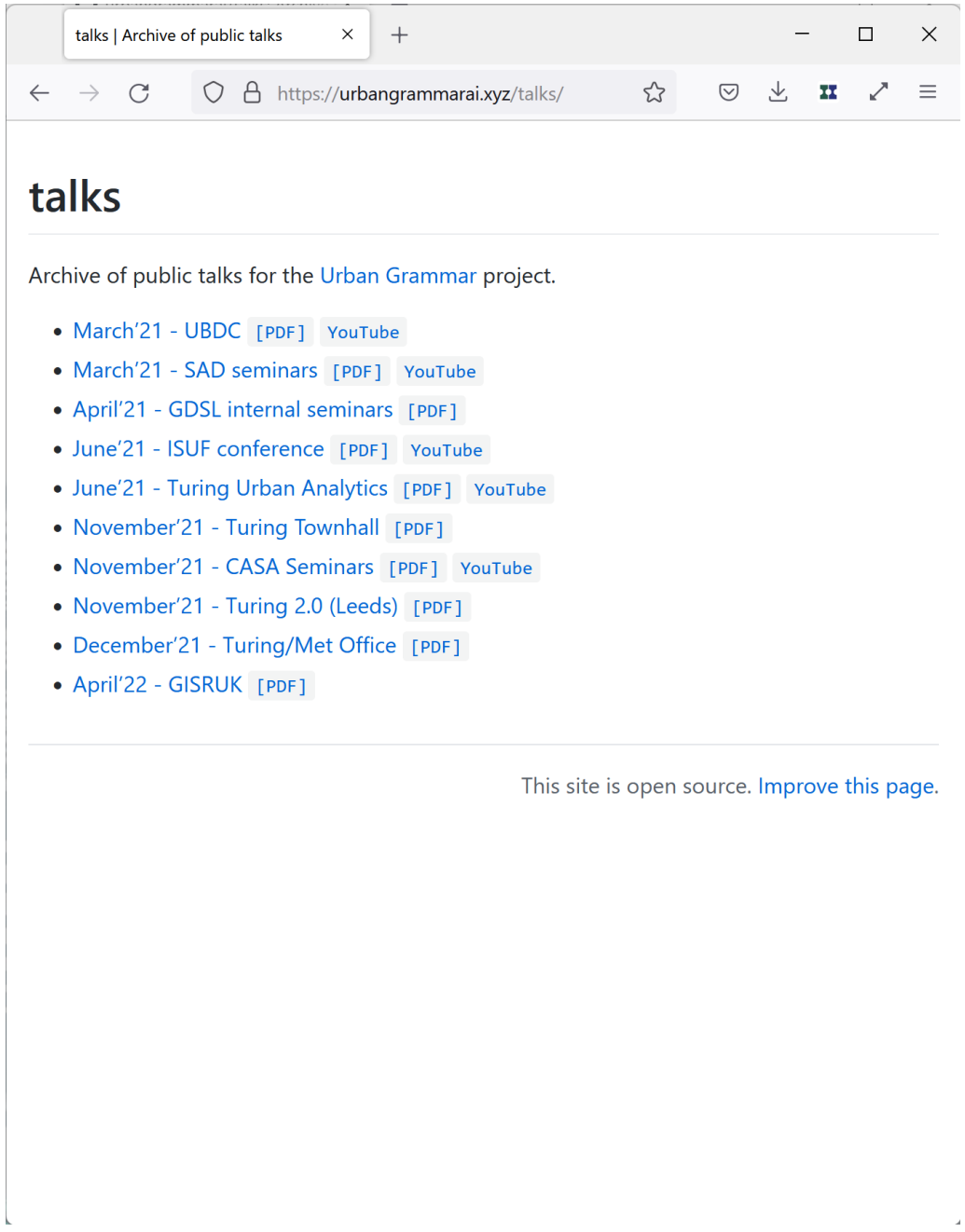


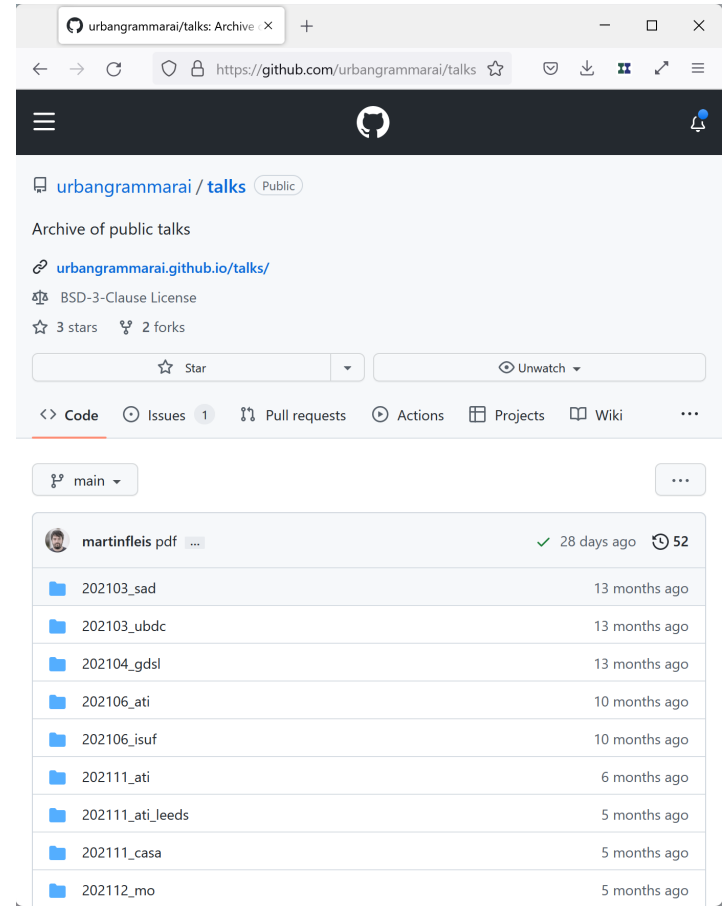
# … and the “sausage”

# Dissemmination (I): Papers

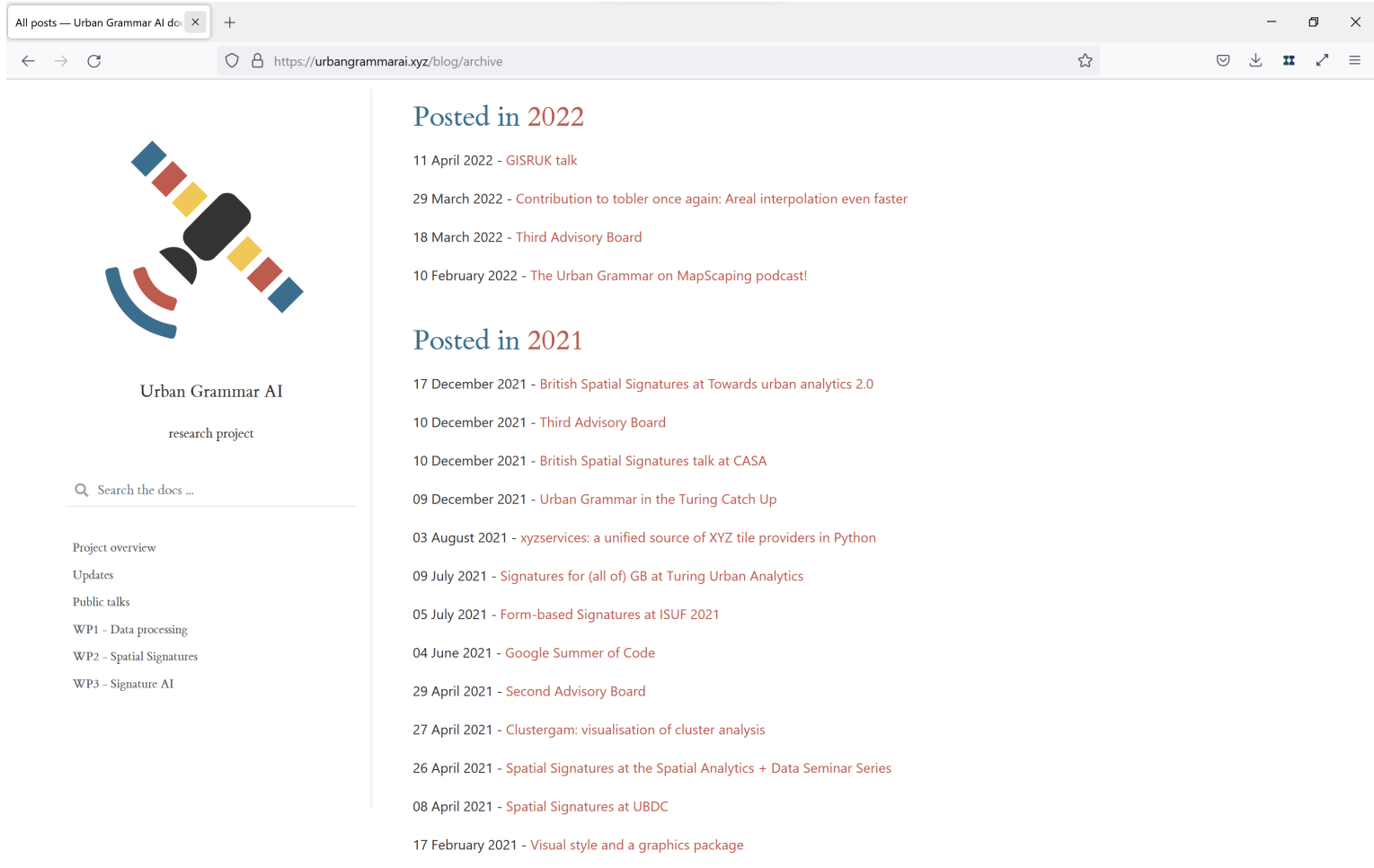


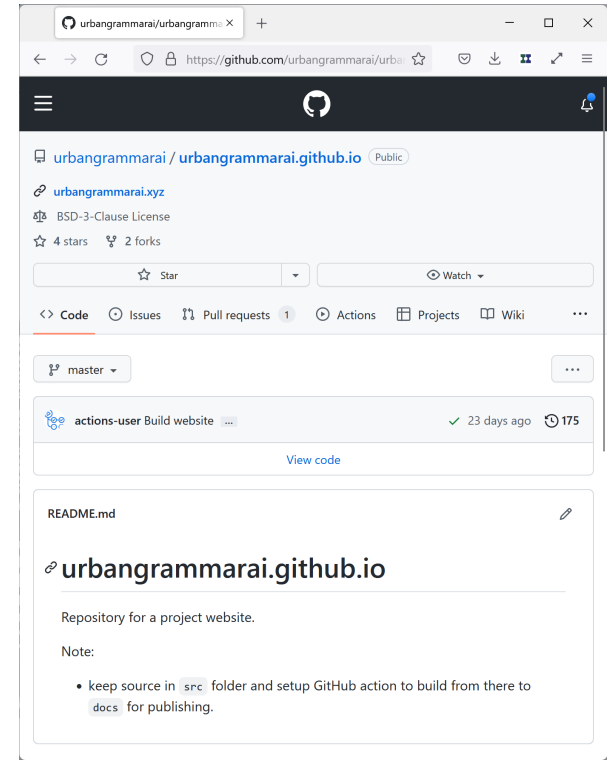
# Dissemmination (II): Talks



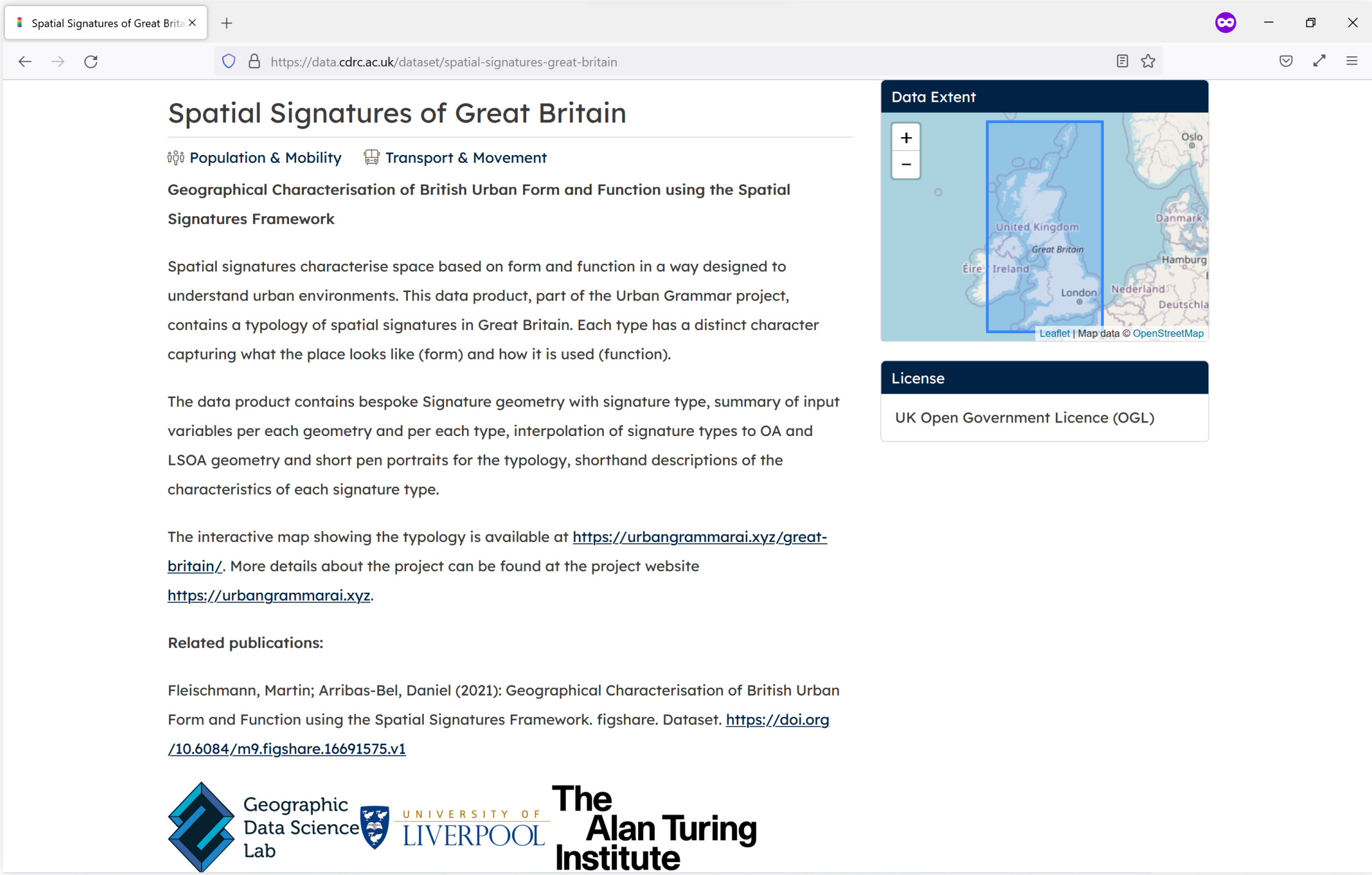



# Dissemmination (III): Other

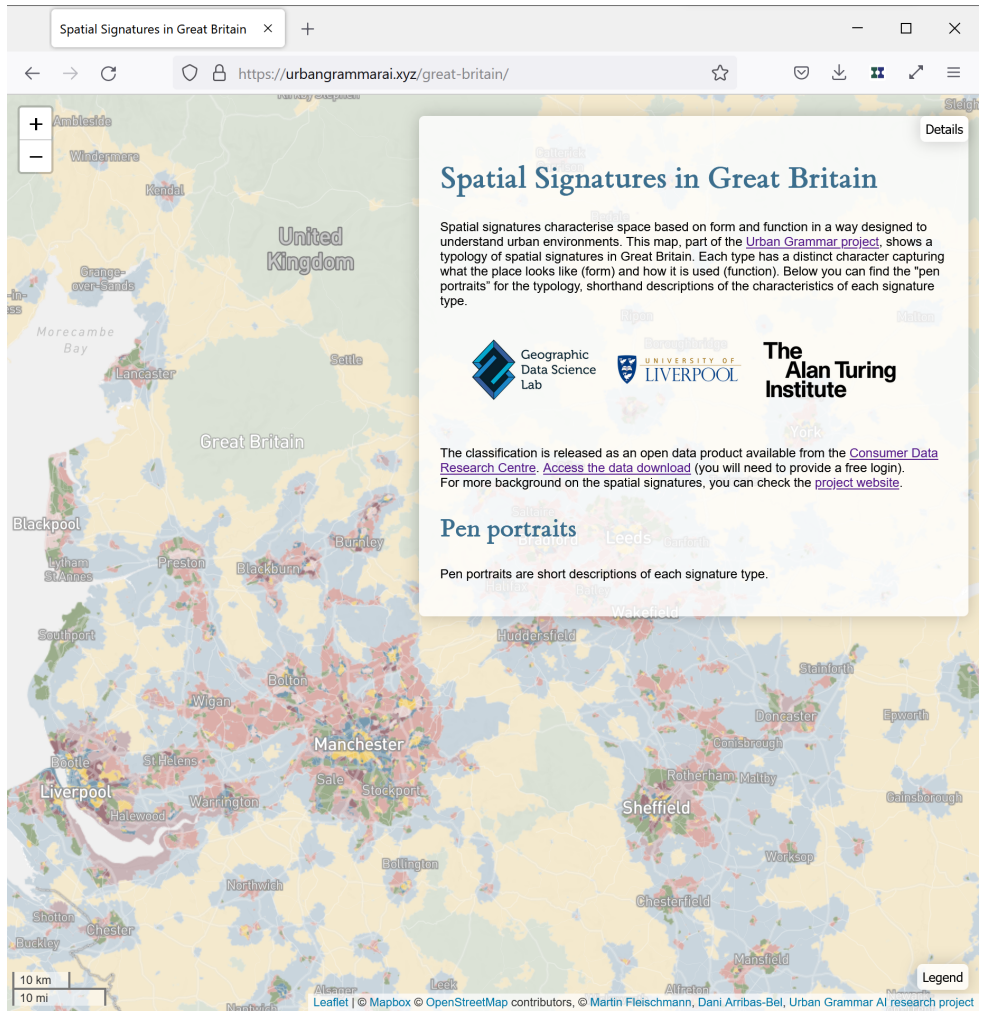


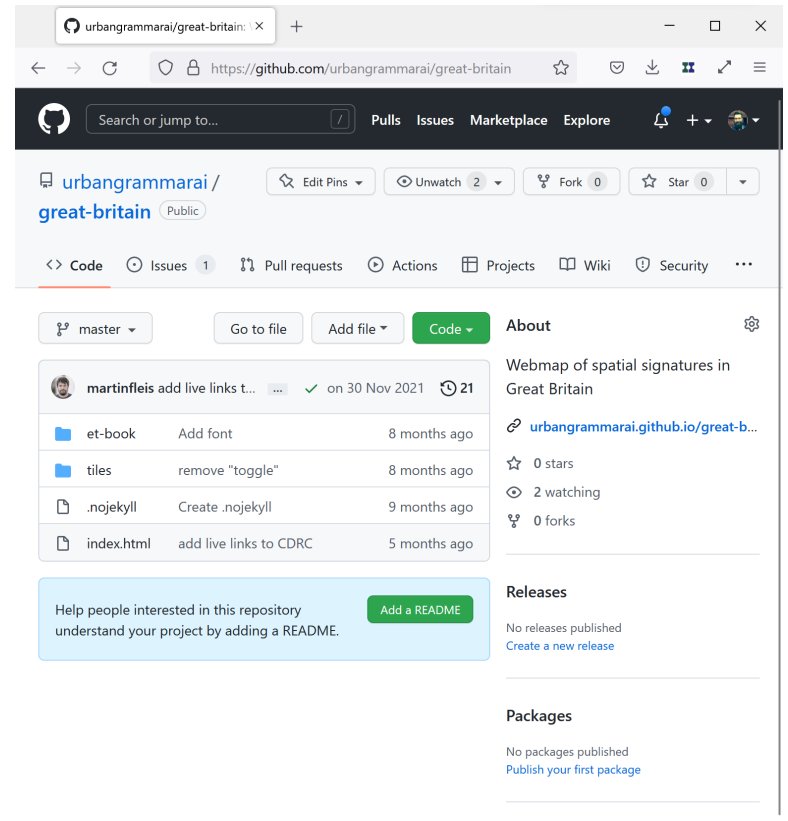



# Open Data Products (I): Data



# Open Data Products (II): Exploratory tools



# The takeaway

- “Release early, release often”
- Make open a feature, not a bug
- *Share your knowledge, it is a way to achieve immortality* (Dalai Lama)

# A bit more to read

Rey, S., Arribas-Bel, D., & Wolf, L. (*in press*). Computational Tools for Geographic Data Science. Chapter 2 in *Geographic Data Science with Python*. Available at: https://geographicdata.science/book

Boeing, G., & Arribas-Bel, D. (2021). GIS and Computational Notebooks. *Geographic Information Science & Technology Body of Knowledge*, 2021(Q1).

Arribas-Bel, D., Green, M., Rowe, F., & Singleton, A. (2021). Open data products-A framework for creating valuable analysis ready data. *Journal of Geographical Systems*, 23(4), 497-514.

https://urbangrammarai.xyz

# *"Open by Default"*

## Developing reproducible, computational research


Dani Arribas-Bel
@darribas

Martin Fleischmann
@martinfleis

The Alan Turing Institute

Geographic Data Science Lab