PyData Rome 4th Meeting, May 8th 2024

Immobiliare Lab, Rome - Italy

Geospatial Data Integration with Jupyter Notebook and Open Source Software

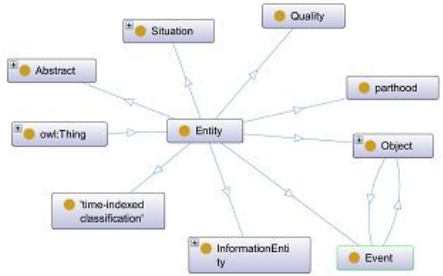
Luigi Selmi
lselmi@datiaperti.it

Who I am: working experience and background

Working Experience: software engineer in enterprises and research institutions in Italy, Switzerland, Germany

Academic Background: Atmospheric Physics (Sapienza University of Rome)

Interests: satellite imagery, algorithms and machine learning applied to geospatial and climate data



datiaperti

PyData

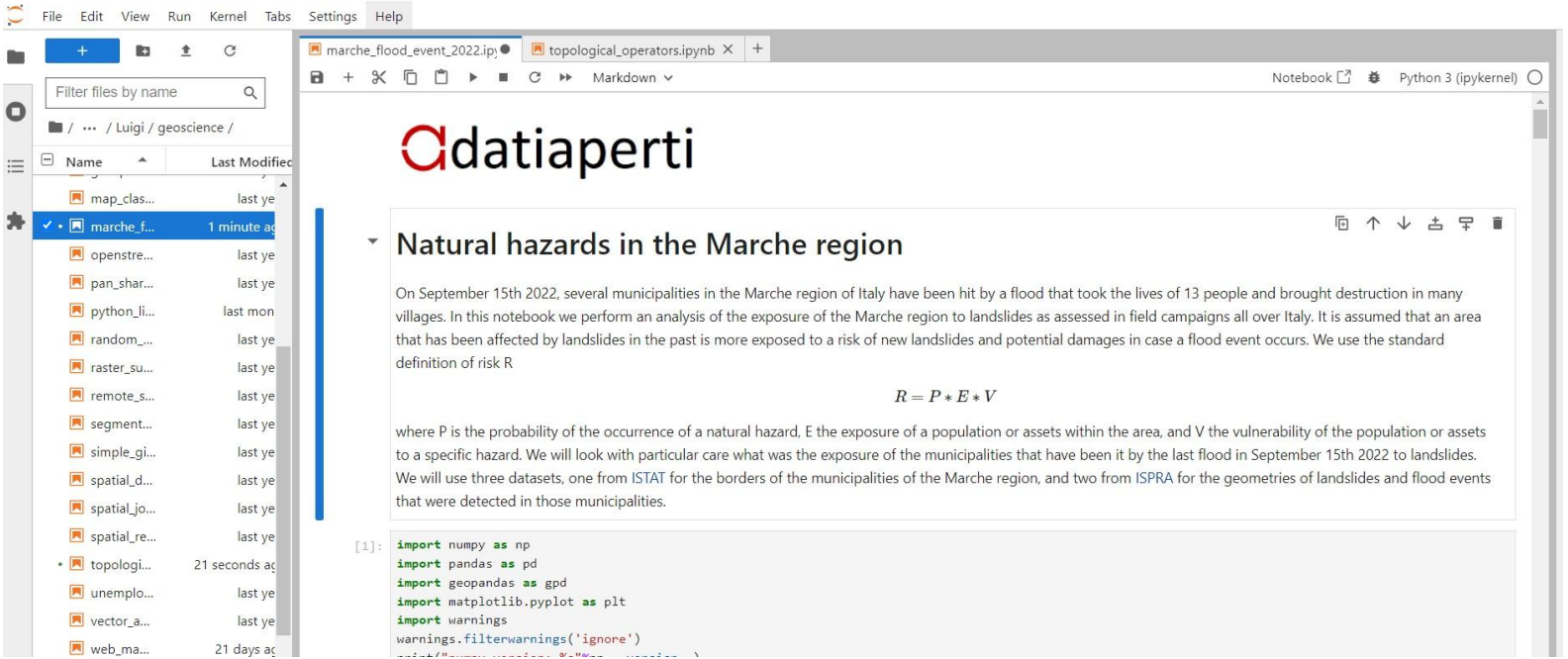
- Communities of open source software users and developers: 230 groups in 80 countries, 6 in Italy
- Supported by NumFOCUS (Jupyter, NumPy, Pandas, Matplotlib,..)
- PyData Rome started in Nov. 2022, 102 members, 4 meetups

Outline

- Jupyter Notebook as an IDE for prototyping
- Creating new data products from open datasets: raster, vector, statistics
- Topological operators for spatial reasoning
- Risk assessment: landslides and floods in the Marche region
- Conclusion

Jupyter Notebook

Without further ado..

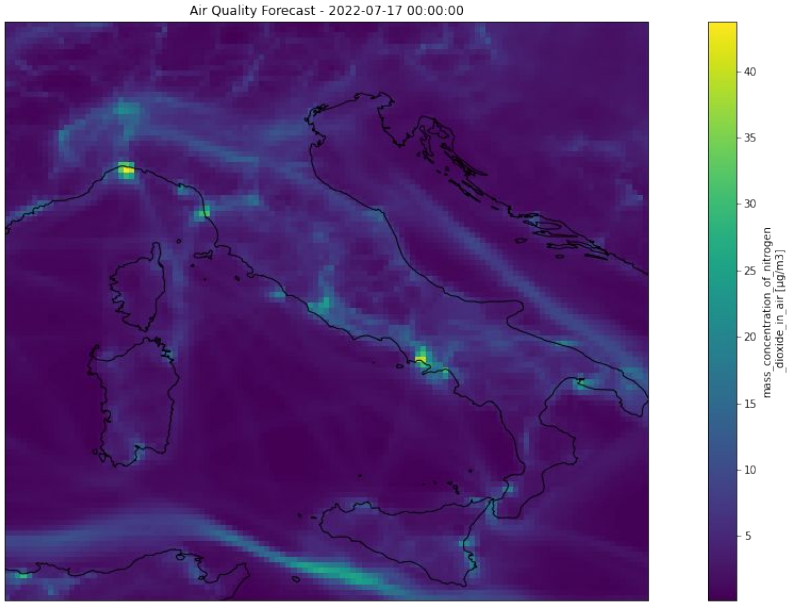


Open Data

The EU Directive 2019/1024 on open data and the re-use of public sector information established the following as high-value datasets

- Geospatial
- Earth observation and environment
- Meteorological
- Statistics
- companies and company ownership
- mobility

Open Data - Examples



Istat | Istituto Nazionale di Statistica

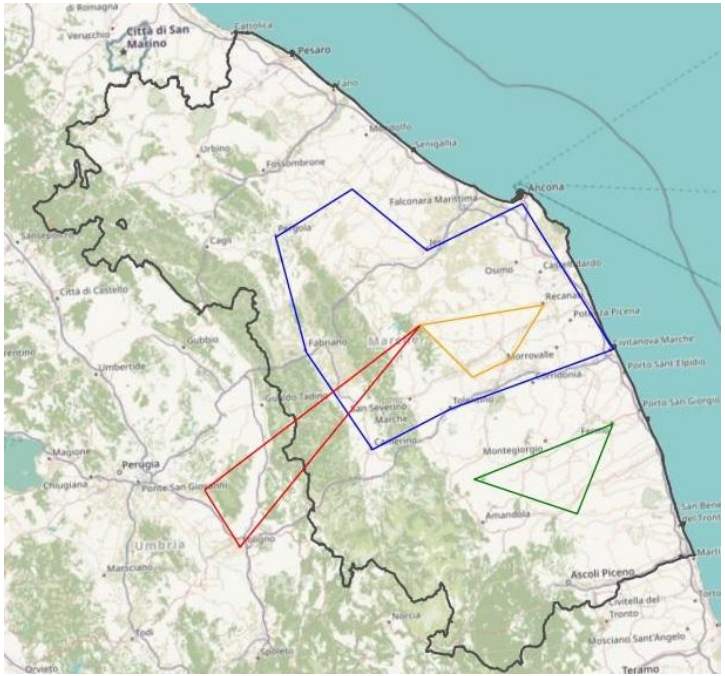
Copernicus Europe's eyes on Earth

ISPRA Istituto Superiore per la Protezione e la Ricerca Ambientale

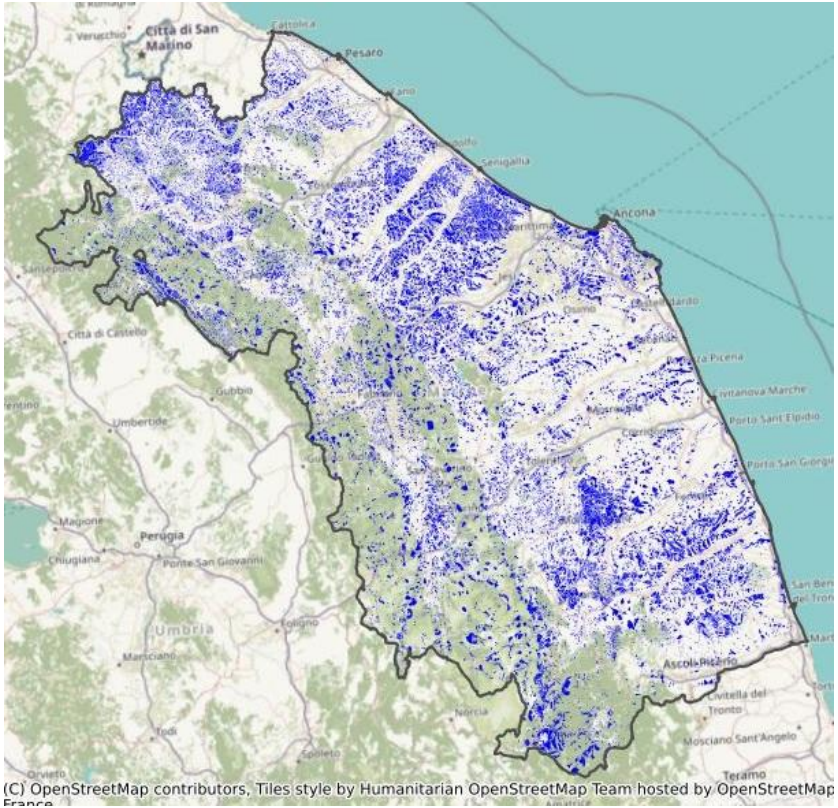
SPQR Open Data

Spatial Reasoning: Topological Operators

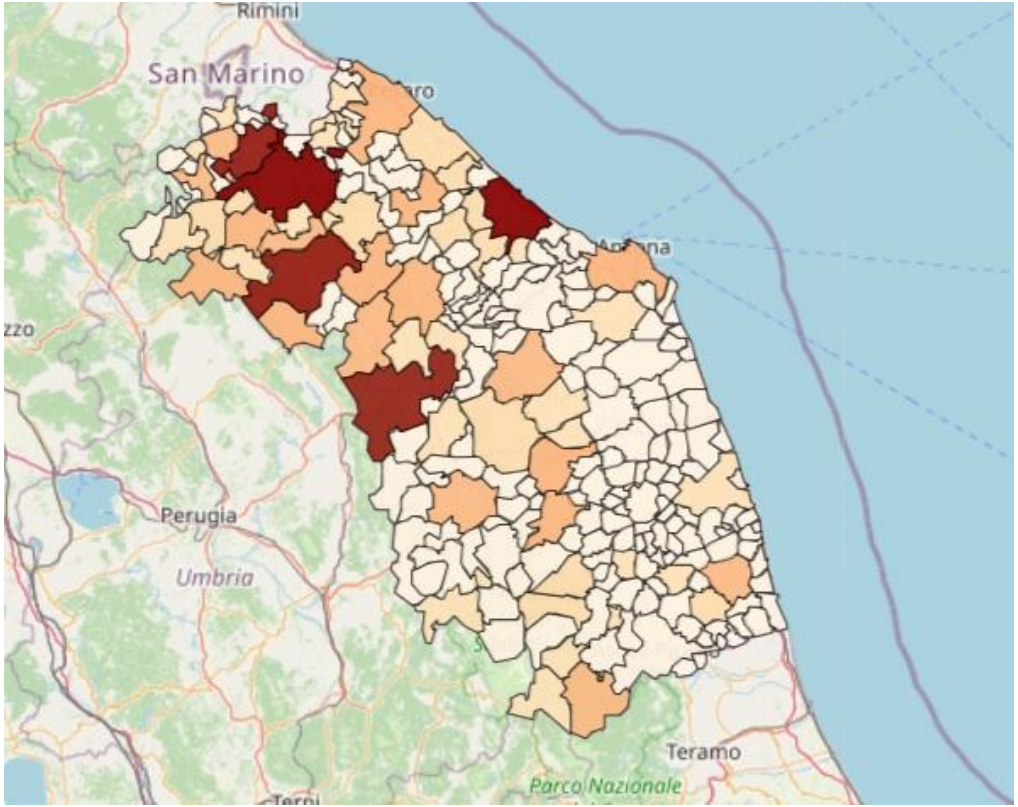
within, intersects, touches, covers, overlaps



Risk Assessment: landslides in the Marche region



Number of landslides



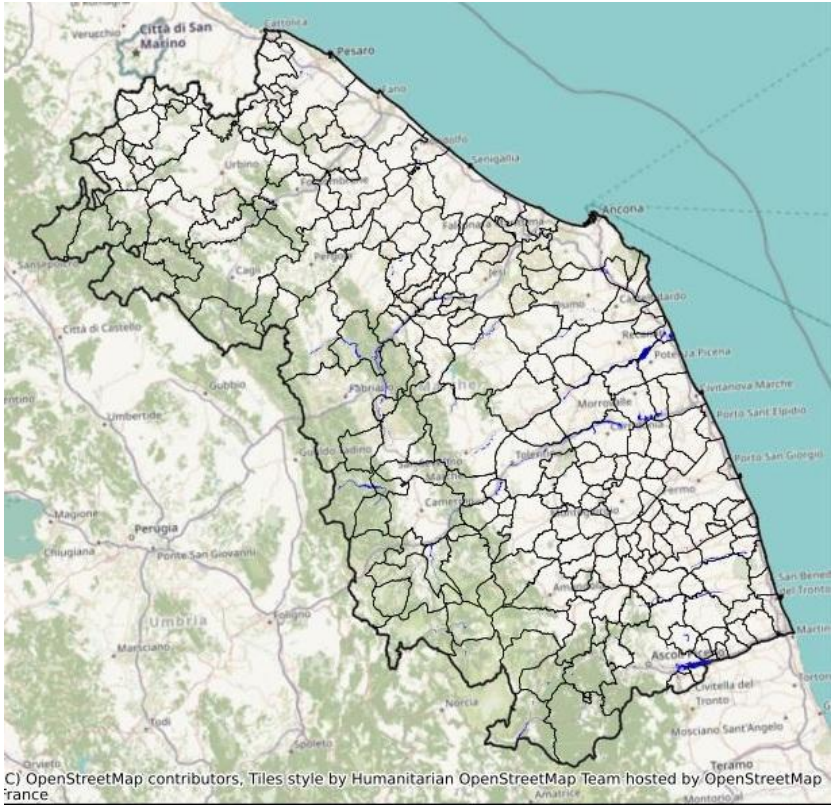
Flood risk in the Marche region

There are 158 areas at risk of flood but the dataset does not provide the toponyms so we need to add them to it using the topological operators and the dataset of the region’s municipalities

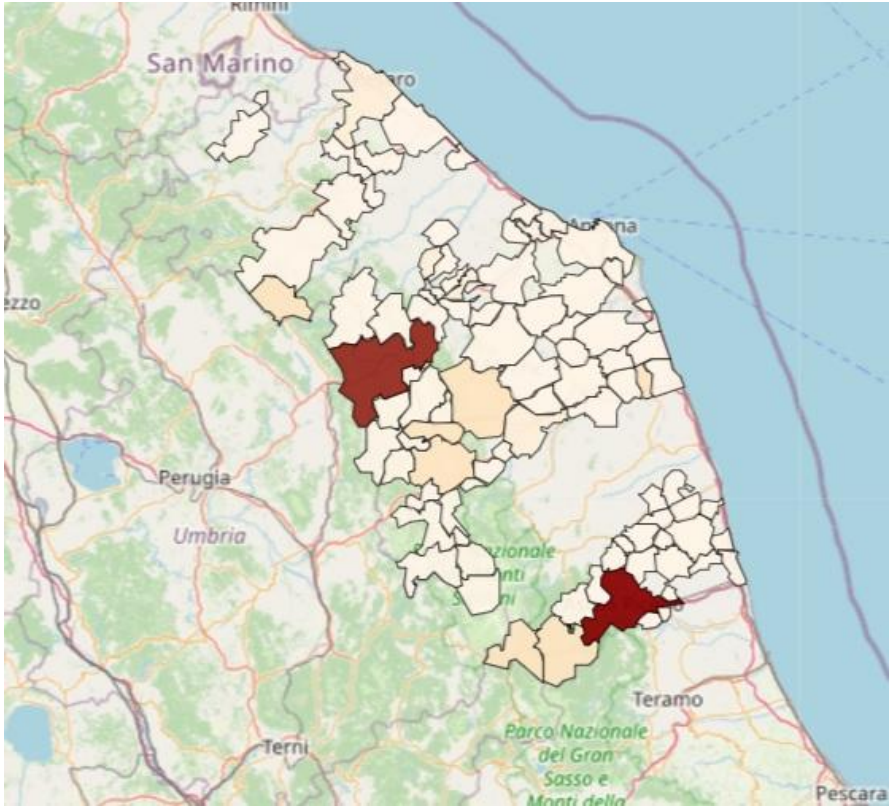
	scenario	geometry
0	Pericolosita' idraulica bassa - LowProbability...	POLYGON ((816476.192 4173280.111, 816449.779 4...
1	Pericolosita' idraulica bassa - LowProbability...	POLYGON ((807674.627 4200250.348, 807691.313 4...
2	Pericolosita' idraulica bassa - LowProbability...	POLYGON ((810203.964 4208515.556, 810266.895 4...
3	Pericolosita' idraulica bassa - LowProbability...	POLYGON ((812030.591 4214361.886, 812019.495 4...
4	Pericolosita' idraulica bassa - LowProbability...	POLYGON ((811810.404 4215280.789, 811849.523 4...

Risk Assessment: floods in the Marche region

Using the ***within*** and the ***intersects*** operators between the areas at risk and the polygons of the Marche region and of its municipalities we can map them and have some useful statistics, e.g.: 90 municipalities out of 225 are at risk of flood

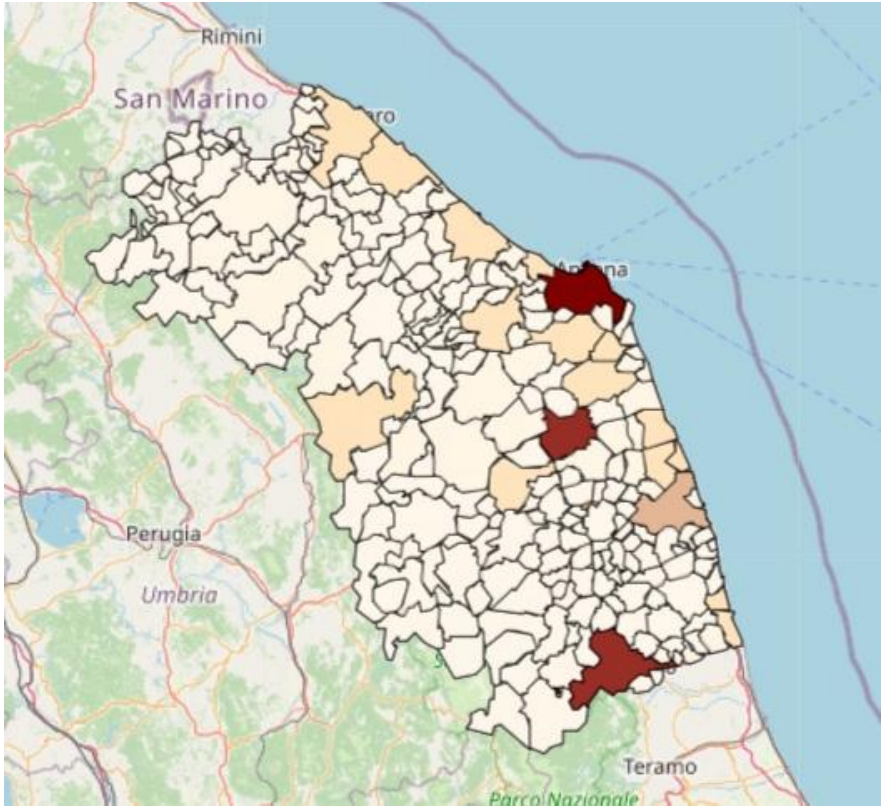


Risk Assessment: floods in the Marche region



Population exposed to natural hazards

Risk is associated to events that have a negative impact on the population or on their assets that are exposed to them, depending on their vulnerability, i.e. their capacity to avoid or reduce those impacts.



Risk: a definition

$$R = \sum_i P_i * E_i * V_i$$

Probability = [0, 1], probability of floods or landslides (aka extreme events)

Exposure = [0, 1], population or assets

Vulnerability = [0, 1], infrastructures or safety systems (or lack of them)

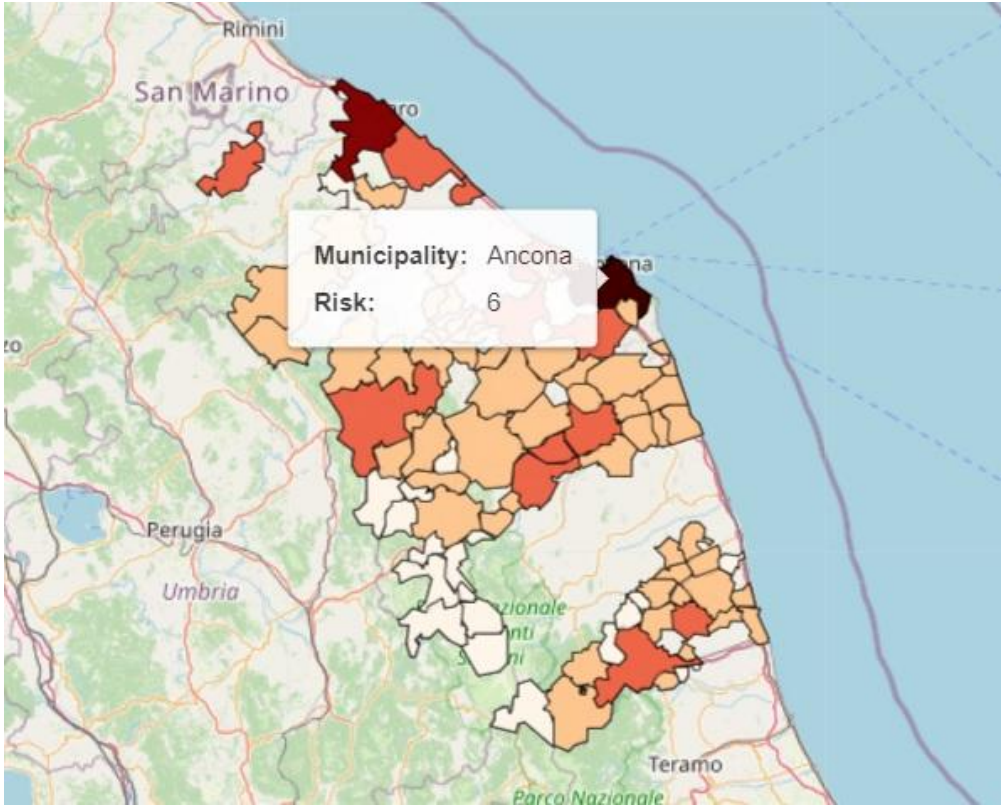
Ranking the municipalities by risk

We can order the municipalities by their exposure to floods and landslides using a formula such as

$$Risk = log(\frac{Floods + Landslides}{Area} * Population)$$

This is of course a very rough estimation, its only purpose is to show how to compute a ranking.

Risk map


Rimini
San Marino
Municipality: Ancona
Risk: 6
Perugia
Umbria
Teramo

Conclusion

We have shown how to make up a very rough idea of risks using a Jupyter Notebook, some open datasets, and some topological operators to merge those datasets. More precise and useful estimations of risks can be computed with more detailed datasets and procedures.

Code and more details

Topological operators:

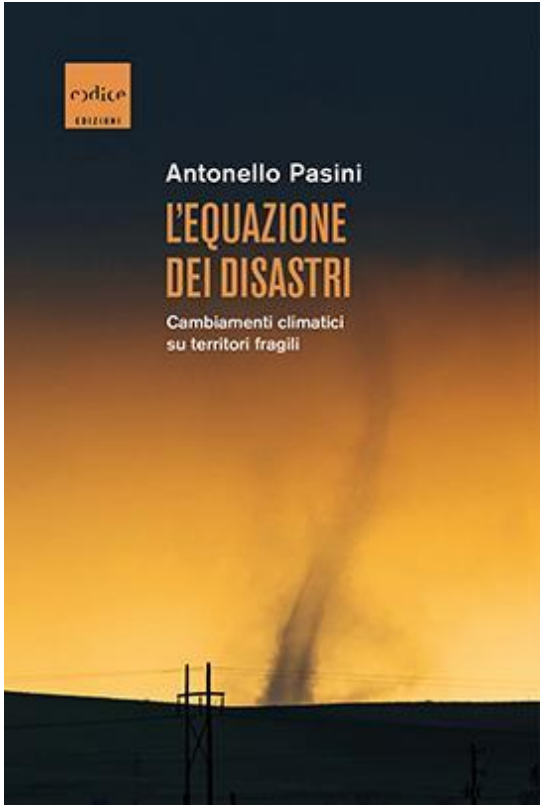
https://github.com/luigiselmi/geoscience/blob/main/topological_operators.ipynb

Landslides and floods in the Marche region

https://github.com/luigiselmi/geoscience/blob/main/marche_flood_event_2022.ipynb

Some references

- Pasini - L’equazione dei disastri
- FEMA - National Risk Index
- Bousquet - Extreme Value Theory with Applications to Natural Hazards



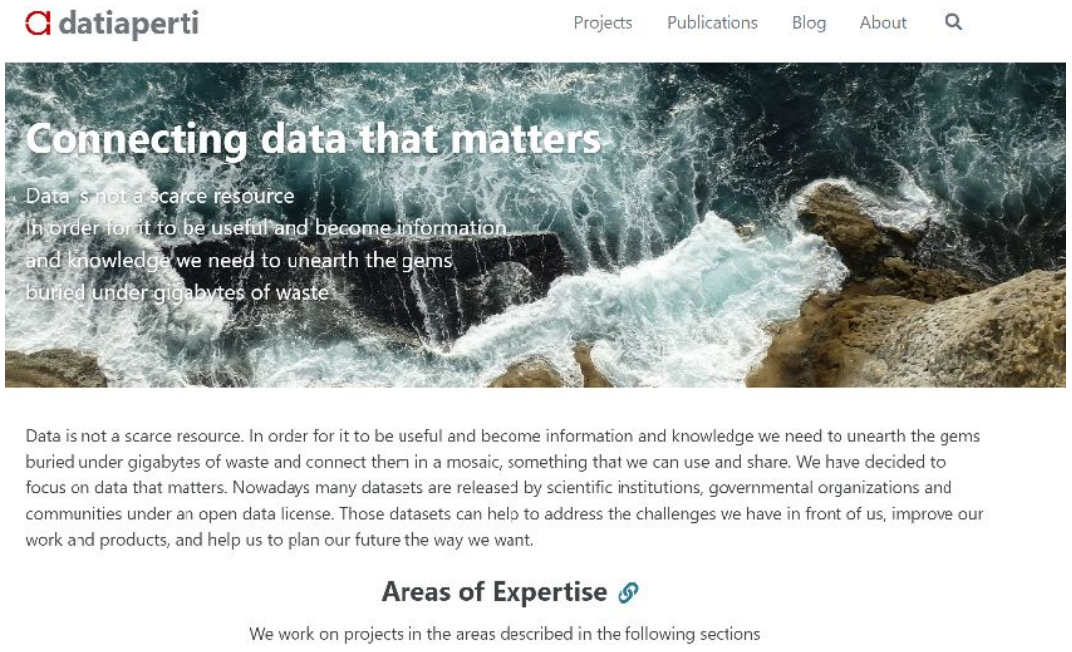
Thank you for your attention!

Email: lselmi@datiaperti.it

Personal website:
https://www.luigiselmi.eu/

Company website:

https://www.datiaperti.it/



datiaperti