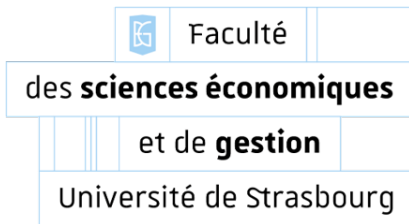


# Novelty Components of Scientific Productions

-

## Data Visualization



**Supervised by:** Diletta Abbonato

**Authors:** Valentin Barthel, Adrien Busché, Diego Fracassi

[valentin.barthel@etu.unistra.fr](mailto:valentin.barthel@etu.unistra.fr)

[adrien.busche@etu.unistra.fr](mailto:adrien.busche@etu.unistra.fr)

[diego.fracassi@etu.unistra.fr](mailto:diego.fracassi@etu.unistra.fr)

# Table of Contents

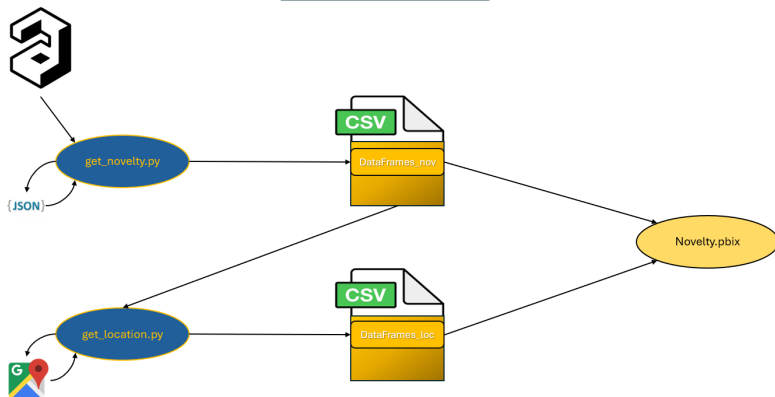
- 1 Introduction
- 2 Novelty Indicator
- 3 Geocoding
- 4 Our Dashboard Structure
- 5 Limitations

# Structure of the Project

## Arborescence of the Project :

```
C:.\n  get_location.py\n  get_novelty.py\n  main.py\n  Novelty.pbix\n  README.md\n  requirements.txt\nDataFrames_loc\n...\nDataFrames_nov\n...\nPresentation\nNoveltySchema.png\nNovelty_Support.pdf
```

### Our Program Diagram



⇒ Github

# Co-occurrence Matrix

- The co-occurrence matrix represents how frequently two references (journals) appear together in a scientific article.
- Each cell  $(i, j)$  contains the number of times journal  $i$  and journal  $j$  are cited together.

Reference	Journal A	Journal B	Journal C	Journal D
Journal A	10	3	2	0
Journal B	3	15	1	1
Journal C	2	1	12	0
Journal D	0	1	0	8

**Table:** Example of a co-occurrence matrix

$\Rightarrow$  *get\_novelty.py*

# Geocoding

The aim is to identify the geographical location of institutions linked to authors of scientific publications.

Functions:

- **get\_city\_state(*place\_name*)**: Fetches geographic details (city, region, state, latitude, longitude) for a given institution name (Google maps API).
- **get\_city\_from\_coordinates(*lat*, *lng*)**: Finds city, region, and state from latitude and longitude (Google maps API).
- **get\_continent\_from\_country(*country\_name*)**: Determines the continent for a given country (pycountry\_convert).

⇒ *get\_location.py*

# Our Dashboard Structure

This dashboard allows you to explore the novelty indicators for scientific productions on the 17 SDGs since 2016. It consists of 5 pages providing a complete analysis of the current and past situation.

- **‘Summary’**: This page provides information about a number of SDGs. We can see the number of citations, publications, and novelty indicator since 2016 for each author, area of expertise, type of production, and the evolution by year.
- **‘Year View’**: A year-by-year analysis of each SDG. We can see changes in key indicators compared with the previous year, as well as the most active SDGs by type and institution.
- **‘Mapping’**: This gives you a geographical view of the scientific output situation.
- **‘SDG Comparison’**: Allows you to compare the key indicators for different SDGs. We offer a view by year, by region, and analyze the differences in the link between the number of contributors and the novelty indicator.
- **‘Database’**: A page that allows the user to check the data by being able to sort it.

# Limitations of the Novelty Indicator

- **Dependence on citation patterns:** The indicator relies heavily on the availability and accuracy of citation data.
- **Field-specific dynamics:** It may not account for differences in citation practices across disciplines.
- **Computational limitations:** In very large datasets, constructing co-occurrence matrices and calculating frequencies can be resource-intensive.
- **Simplistic measure of novelty:** The indicator may not capture the full spectrum of creativity or innovation in the research being analyzed.



# Our Limitations

- **Keyword Queries:**

- The use of multiple queries per SDG is limited due to error 403 restrictions.
- The selected queries may not necessarily capture the most cited scientific production for the thematic area.

- **Computational and Sample Size Limitations:**

- Each query is constrained by the frequency of the selected keyword's occurrence in scientific papers.
- The available RAM (16 GB) is insufficient for the demands of a comprehensive scientometric approach.