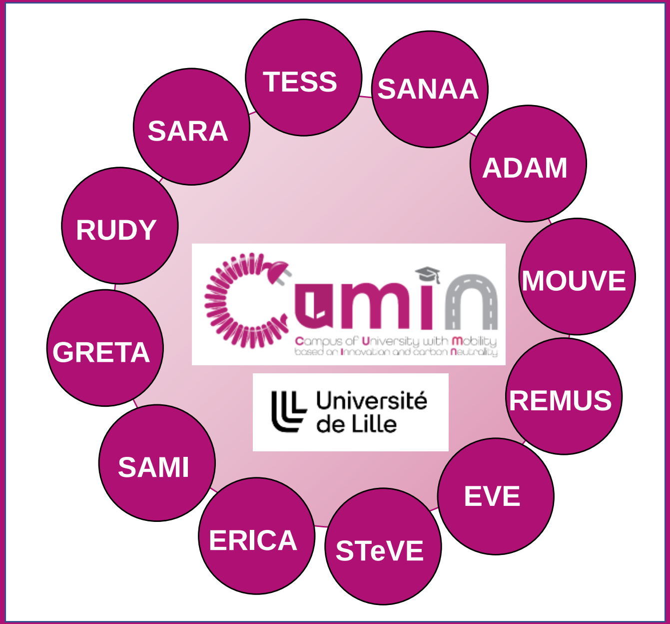
CUMIN - SARA

https://cumin.univ-lille.fr/



Open Data Plateform on e-Mobility

Q. Pochet, A. Fraisse and A. Bouscayrol

Gériico, L2EP, ULille

SARA Project : Social Acceptance of electric vehicles in Restricted Areas

- The goal of SARAH is to understand the opinions of users at the University of Lille (students and staff) regarding electric vehicles.

- The project aims to better identify potential challenges related to access to renewable transportation options while assessing participants' perceptions of electric cars.

Objective of the Internship : Vizualization of the trajectories of users on the Cité Scientifique University Campus Based on their Mode of Transport

- The analysis focuses on how users move around the scientific campus based on their mode of transport.

- This will provide insights into transportation habits and potential barriers to sustainable mobility.

Description of the Survey Dataset

- The survey consists of 3 phases:
  - **Phase 1:** Before driving, to collect participants' initial opinions on electric vehicles.
  - **Phase 2:** During the driving experience.
  - **Phase 3:** After driving, to gather final feedback.
- The questionnaire contains approximately **40 questions**.
- Two datasets are available:
  - **Phase 2 (May 15 - July 13, 2023):** 72 responses.
  - **Phase 3 (February 7 - June 25, 2024):** 44 responses.

Mapping Vizualization of Survey Results

- The objective of processing this questionnaire data is to create a **map-based visualization** of the daily movements of students and staff on the **Cité Scientifique campus**.

- This will help identify potential **patterns** and transform data into **actionable knowledge** about users' **transportation habits** when commuting to campus.

9 Modes of Transportation Used by Respondents

1. Walking
2. Metro / Tramway *(Grouped due to data limitations)*
3. Bus
4. Mechanical Bicycle
5. Electric Bicycle
6. Car
7. Electric / Hybrid Car
8. Carpooling
9. Hybrid Carpooling

Data Cleaning Process – Phase 1

Before visualizing the data, we first processed the original CSV file, which contained over **40 columns**. The dataset was **filtered** to retain only the most relevant values:

✅ **Mode of transportation**
✅ **Travel duration**
✅ **Distance traveled**
✅ **Respondent's place of residence**

However, the raw data was inconsistent, requiring standardization to ensure accurate analysis.

Data Cleaning Adjustments

Several modifications were applied to **harmonize** the dataset:

- Standardizing city names and correcting misspellings
- Converting inconsistent time values (e.g., “More than 1 hour” → 75 min)
- Rounding distance values (e.g., “between 2-5 km” → 3.5 km)
- Replacing place names with distances (e.g., metro stop names → estimated distance)

These transformations required interpretation, meaning some assumptions were made to ensure data consistency.

Data Visualization – Anonymization Process

To **preserve anonymity**, we apply a small random offset to all location points on the map.

✅ A function called **"add_noise"** is used on **Shapely's** "point" object.

✅ The noise level is set to **~1 km**, ensuring minimal distortion while protecting identities.

✅ The algorithm introduces **random shifts** in longitude (**x**) and latitude (**y**) to simulate anonymization.

This allows us to maintain **privacy** while enabling insightful **spatial analysis**.

Mapping Mobility Flows & Transportation Modes

To visualize movement patterns, we create **two dictionaries**:

- Transport Modes → Separate dataframes for each travel mode.
- Color Assignments → Each mode is color-coded for clarity.

These visual elements help identify mobility trends within the dataset.

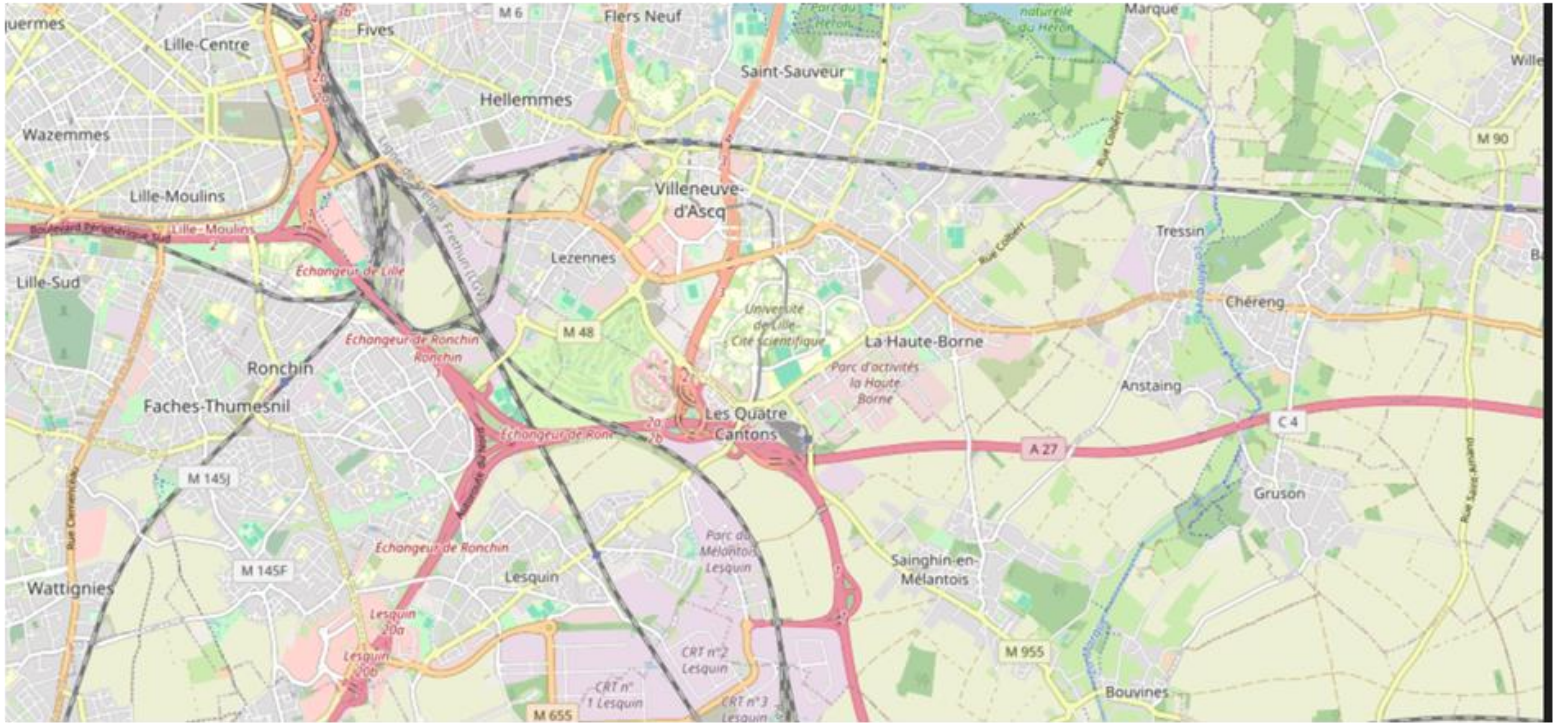
Transport Mode	Color
Walking	Green
Metro/Tram	Purple
Bus	Pink
Mechanical Bike	Cyan
Electric Bike	Blue
Car	Red
Electric/Hybrid Car	Orange
Carpooling	Brown
Electric Carpooling	Yellow

Required Libraries for Data Vizualization

- **Pandas**: dataframe manipulation (data tables).
- **NumPy**: numerical calculations.
- **Shapely**: shapes on maps and other geographic visualizations
  Folium: interactive maps.
- **Scikit-learn**: machine learning algorithms.
    - For our work, we only need one specific functionality:
        - **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)**: clustering with a small amount of noise added. This helps automatically create clusters based on specified parameters while preventing the exact geographic position of users from being revealed and ensures compliance with **GDPR** (General Data Protection Regulation).

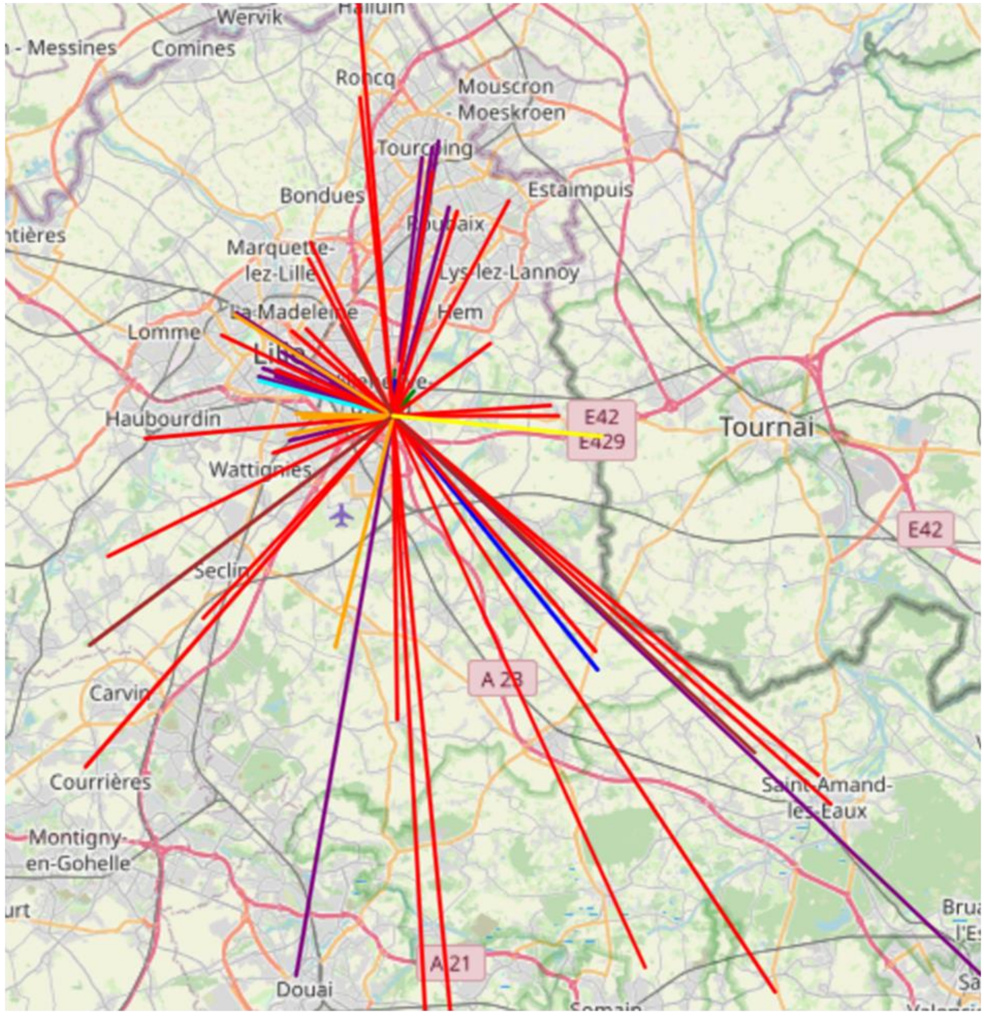
Intreractive Map of Cité Scientifque

```
#Création de la carte vierge (sans les flows)
m = folium.Map(location=campus_coords, zoom_start=13)
✓ 0.0s
```



Issue with the Initial Visualization

- *Before applying clustering*
- Representation of unique individual movement flows
- Chaotic and unreadable result
- Risk of revealing personal information despite adding an error margin
- Individual trajectory visualization is not usable



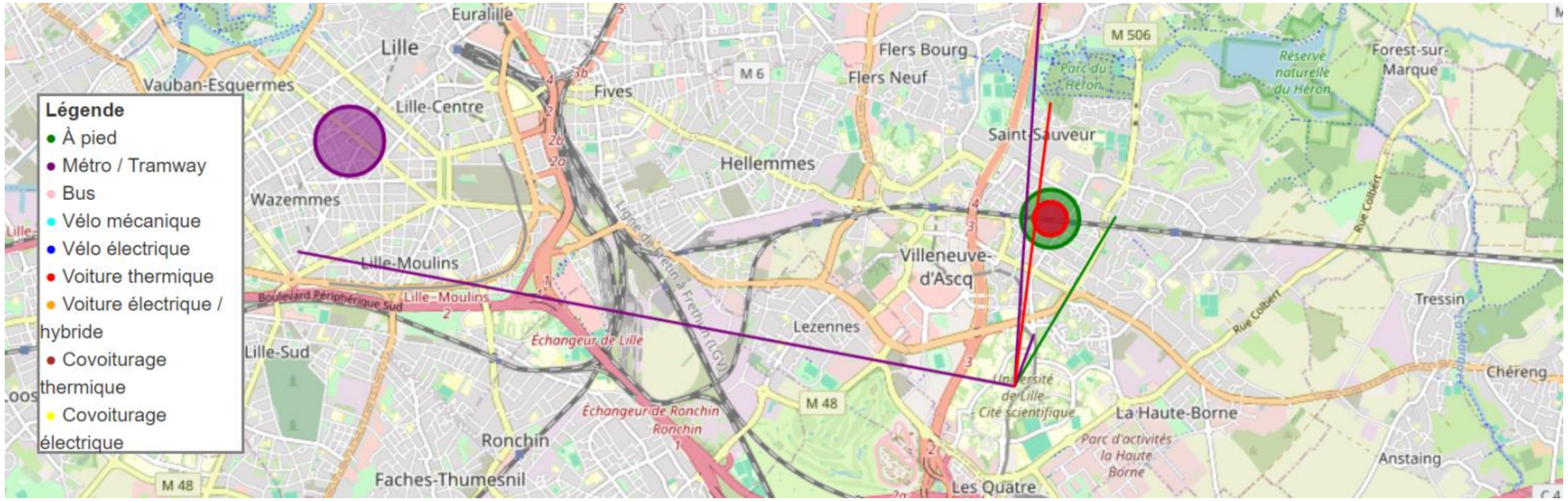
Clustering Approach With DBSCAN

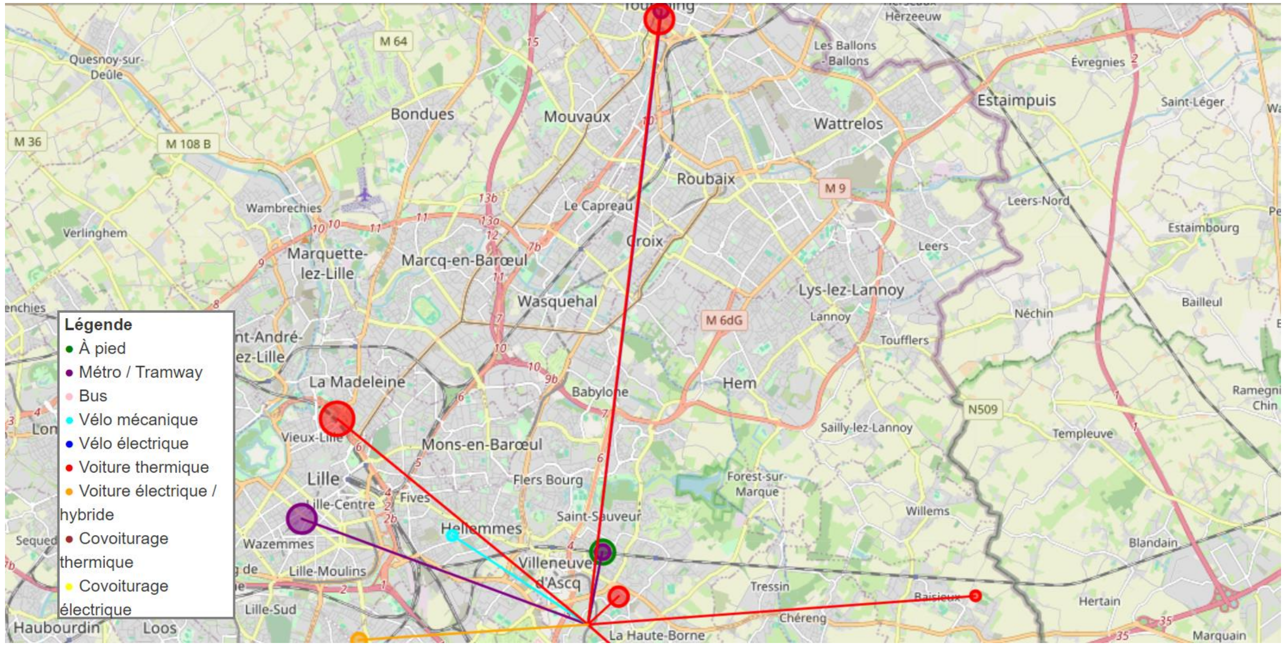
Grouping users into clusters (groups of similar flows)
Algorithm used: **DBSCAN** (from scikit-learn)
Parameters:

- **eps**: Maximum distance between users to consider them in the same group
- **min_samples**: Minimum number of points required to form a cluster (here, 3)

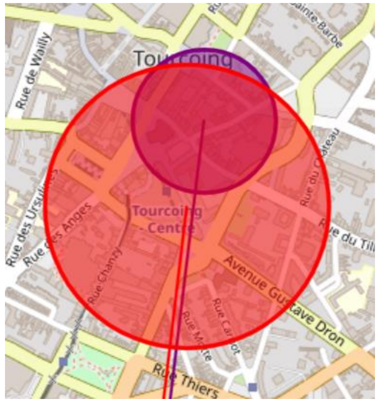
```python
#Application du clustering DBSCAN
clustering = DBSCAN(eps=eps, min_samples=min_samples).fit(points)

#Récupération des clusters
clusters = {}
for i, label in enumerate(clustering.labels_):
    if label == -1:  # Label -1 signifie bruit (point isolé)
        continue
    if label not in clusters:
        clusters[label] = []
    clusters[label].append(points[i])
```

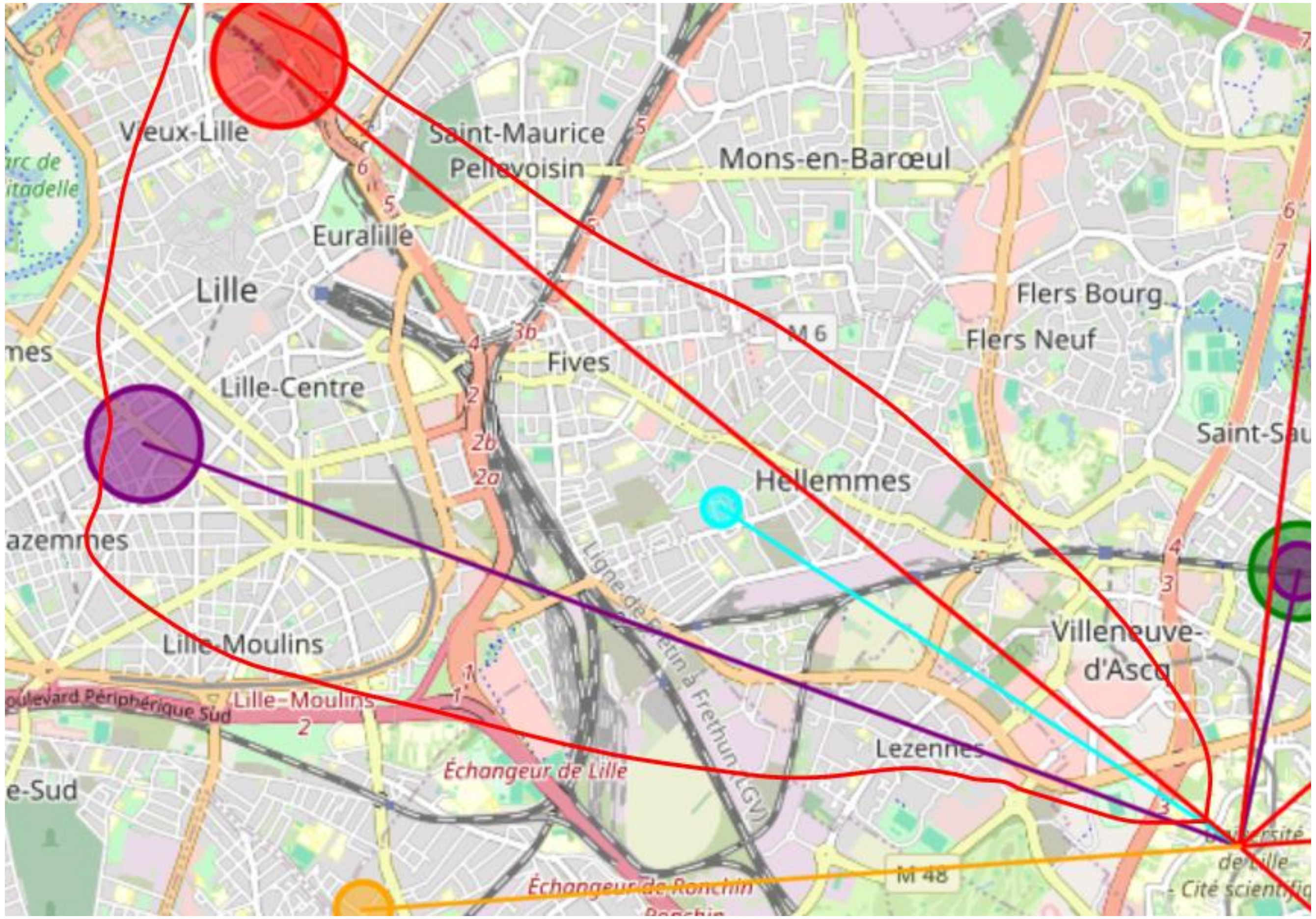
User Cluster's Flow Map





Distance Similarity Between 2 Different Travel Mode





Visualization representing a similarity in distance between metro/tramway and car trips.

Result Analyis and Usage Overview

After data classification, we identified the most common modes of transportation among the respondents:

- Thermal car: 32 respondents (44%), the most used mode.
- Metro / Tramway: 19 respondents (26%), in second place.
- Walking & Electric/Hybrid Cars: 5 respondents each (7% each), tied for third place.
- Carpooling (thermal): 3 respondents (4%).
- Bicycle (electric & mechanical): 2 respondents each (3% each).
- Bus & Electric carpooling: 1 respondent each (1% each).

Comparison With University’s Mobility Survey

University of Lille 2023 Mobility Survey:

- Metro/tramway most used for daily commuting, followed by thermal car.
- **Walking**: Third most popular mode of transport.
- **Bicycles** and **buses** were more popular in their survey than in our dataset.
- **Train**: Not represented in our data, but used by around 10% of Cité Scientifique campus respondents.

Differences: Our sample size (72 responses) is much smaller than the university's (9,584 responses), so the results aren’t fully representative.

Limitations and Hypotheses

- **Sample size**: The small number of responses (72 vs. 9,584) makes it difficult to generalize to the entire campus population (78,000 students and 7,200 staff).

- **Context of the Survey**: The dataset was influenced by the target audience for the electric vehicle trial:
  - Mainly **thermal car users** curious about electric vehicles.
  - **Electric car users** might not have felt the need to participate, leading to skewed data.

- These factors help explain why thermal car users dominate our sample.

Insights from the Flowmap Visualization

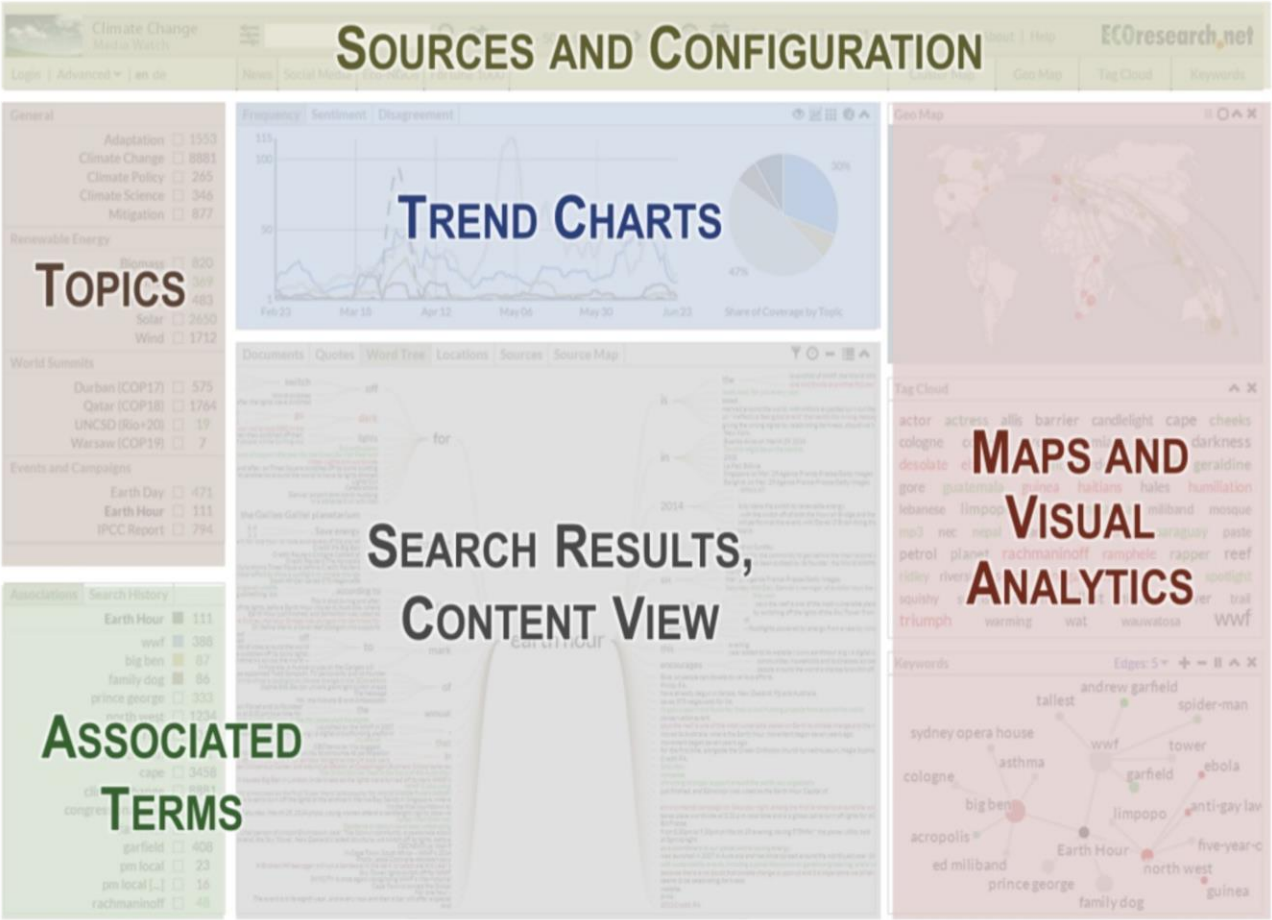
- **Thermal car travel**: Most diversified, with 5 clusters present, reflecting the variety of residential areas (27 different municipalities).

- **Cluster observations**:
  - **Tourcoing clusters**: Overlapping, suggesting potential areas of transportation overlap.
  - **Closest cluster to Cité Scientifique**: The smallest distance, indicating possible alternatives to thermal cars, such as public transport (metro, bus, tram), biking for shorter distances, carpooling, or electric vehicle adoption.

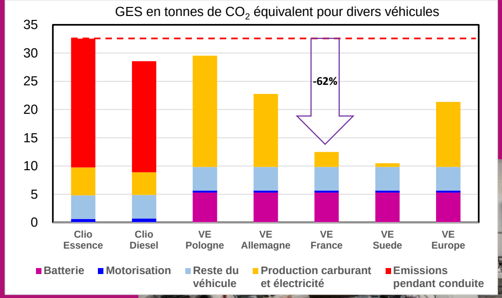
- **Villeneuve-d’Ascq**:
  - A low-carbon footprint with walking and metro/tramway as main modes of transport.
  - The campus is <10 km for most residents, making metro and walking logical choices.

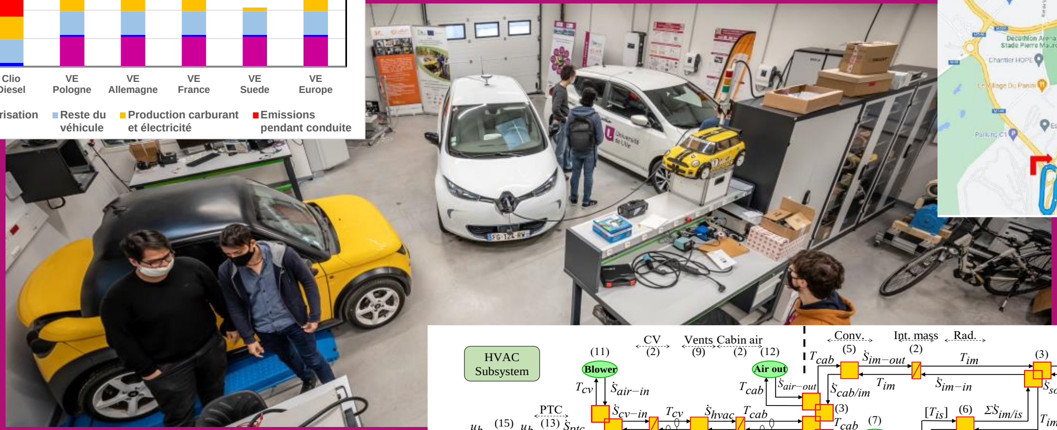
Limitations and Future Improvements

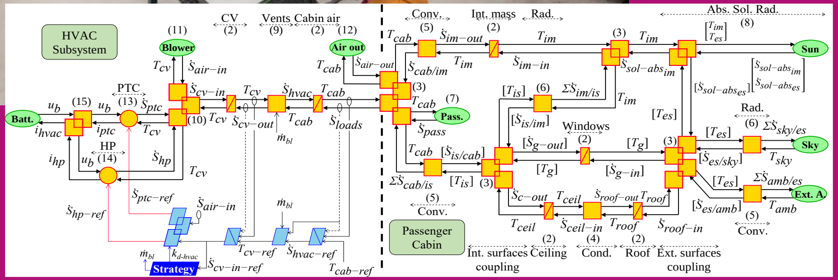
- **Data limitations**:
  - Small sample size (72 responses), not enough to fully represent campus mobility flows.
  - Lack of **Big Data** (GPS data, connected vehicles, sensors), limiting the depth of analysis.



https://cumin.univ-lille.fr/







Our university as an exciting living lab towards eco-cities through an innovative transdisciplinary framework !