**Internship:**

# Harnessing the Power of Open Data for Accessible Future Generation Transportation

## Context

Accessibility [1] measures how well a certain place is connected to the surrounding urban area. A place enjoys high accessibility when it can be easily and rapidly reached from any other place in the area. Accessibility is a key indicator for a town or a district, as it determines quality of life, job attractiveness, value of properties. The design of public transit (bus, metro, train lines and frequencies) strongly impacts accessibility and is often the cause of spatial inequality between city centers and suburbs.

Accessibility can now be objectively quantified relying on big data, publicly available. For instance, GTFS is a data standard created by Google and now adopted by all the main transportation authorities around the world to publish their transit schedules. GTFS data can then be used by $3^{rd}$ party applications to offer several services, e.g., route planner apps. Other socio-economic and demographic data are also publicly available.

The goal of this internship is, first, to quantify and compare accessibility and spatial inequality in several European cities, solely relying on open data and open source software. Second, a method will be devised to automatically modify transit schedules in order to improve such metrics.

## Internship Activity

$1^{st}$ part: The student will deploy the open source big data application *CityChrone* [1] on our local computer cluster. CityChrone is able to ingest GTFS data [2], population information [3] and Open Street Maps and visualize accessibility scores. The student will critically observe the geographical distribution of accessibility in several cities, via advanced data visualization. He will devise several measure summarizing the spatial inequality and compare it for different European cities.

$2^{nd}$ part: The student will create and implement a method to ingest GTFS data, perform some modification (e.g., change the frequency of some lines), regenerate the GTFS data accordingly and calculate the new accessibility, after the modification. The student will conceive a systematic method to reduce spatial inequality via proper modifications. Such a method could be based on network metrics (degree of connectivity of nodes, betweenness centrality, etc.)

## Administrative information

Where: Télécom SudParis, Palaiseau campus

When: Starting from March 2022, duration: 6 months

Supervisor: Assoc. Prof. Andrea Araldo and Vincent Gauthier (Institut Polytechnique de Paris – Télécom SudParis)

## To apply

Candidates must be M2 level ($2^{nd}$ year of MSc or last year of cycle ingénieur) in Data Science, Computer Science, Transportation Engineering or Applied Mathematics.

The candidate must be proficient in Python and have excellent programming skills. She/he must also be able to perform some basic server administration and work with remote servers (ssh). She/he must be proficient with github or similar platforms.

In the first interview, the student will have to show that she/he is able to create a virtual machine with Ubuntu-server (no-gui) installed there, install CityChrone on such a machine and access its graphical interface remotely or from the host.

The candidate must send to andrea.araldo@telecom-sudparis.eu: (i) marks of all her/his BSc and MSc studies, (ii) Cv, (iii) motivation letter.

## Perspectives

Possibility to start a PhD, after the internship. The topic would be big data and reinforcement learning for the design of multi-modal transit.

## References

[1] Biazzo, I., Monechi, B., & Loreto, V. (2019). General scores for accessibility and inequality measures in urban areas. Royal Society Open Science, 6(8).

[2] https://data.iledefrance.fr/explore/dataset/horaires-prevus-sur-les-lignes-de-transport-en-commun-dile-de-france-gtfs/information/

[3] Facebook data for good: Population Density Maps, https://dataforgood.fb.com/tools/population-density-maps/