

WELCOME

WEBINAR

Methodologies and Data-based Technology Solutions for Improved European Mobility

November 22, 2022
14:00 CET



These projects have received
funding from the European Union's
Horizon 2020 research and
innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.** **SoBigData**

RECIPROCITY

Meet the Speakers

Methodologies and Data-based Technology Solutions for Improved European Mobility



Thierry Chevallier
MobiDataLab



Ismini Stroumpou
SYN+AIR



Luca Pappalardo
SoBigData++



Marcell Romsics
RECIPROCITY



These projects have received
funding from the European Union's
Horizon 2020 research and
innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.** **SoBigData** **RECIPROCITY**

Agenda

Methodologies and Data-based Technology Solutions for Improved European Mobility

- 14:00 Welcome and Introduction
- 14:05 MobiDataLab's Transport Cloud
- 14:20 SYN+AIR Technology Solution
- 14:35 SoBigData++ Technology Solution
- 14:50 RECIPROCITY Project Methodology
- 15:05 General Q&A Session and Farewell

Danijel Pavlica (F6S)
Thierry Chevallier (AKKODIS)
Ismini Stroumpou (Sparsity)
Luca Pappalardo (ISTI-CNR)
Marcell Romsics (Zone Cluster)
All Participants



These projects have received funding from the European Union's Horizon 2020 research and innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.** **SoBigData** **RECIPROCITY**

MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

MobiDataLab

Labs for prototyping future mobility data sharing solutions

Webinar Methodologies and Data-Based solutions for improving Mobility

22/11/2022

mobidatalab.eu



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879)

MOBIDATALAB

Funded within the H2020 context Mobility for Growth 2018-2020 Call of the European Union

A European wide set of partners with different skills and backgrounds working commonly on shared project objectives.



Project Information

MobiDataLab

Grant agreement ID: 101006879

DOI

[10.3030/101006879](https://doi.org/10.3030/101006879)

Start date

1 February 2021

End date

31 January 2024

Funded under

SOCIETAL CHALLENGES - Smart, Green And Integrated Transport



| Objectives



Open Knowledge Base

- Assess the mechanisms enhancing data sharing in the transport sector
- Know how to improve the findability, accessibility, interoperability and re-usability of mobility data



Transport cloud prototype

- Prototype a scalable cloud solution for sharing mobility data
 - Showcase the most effective means to access and exchange mobility data

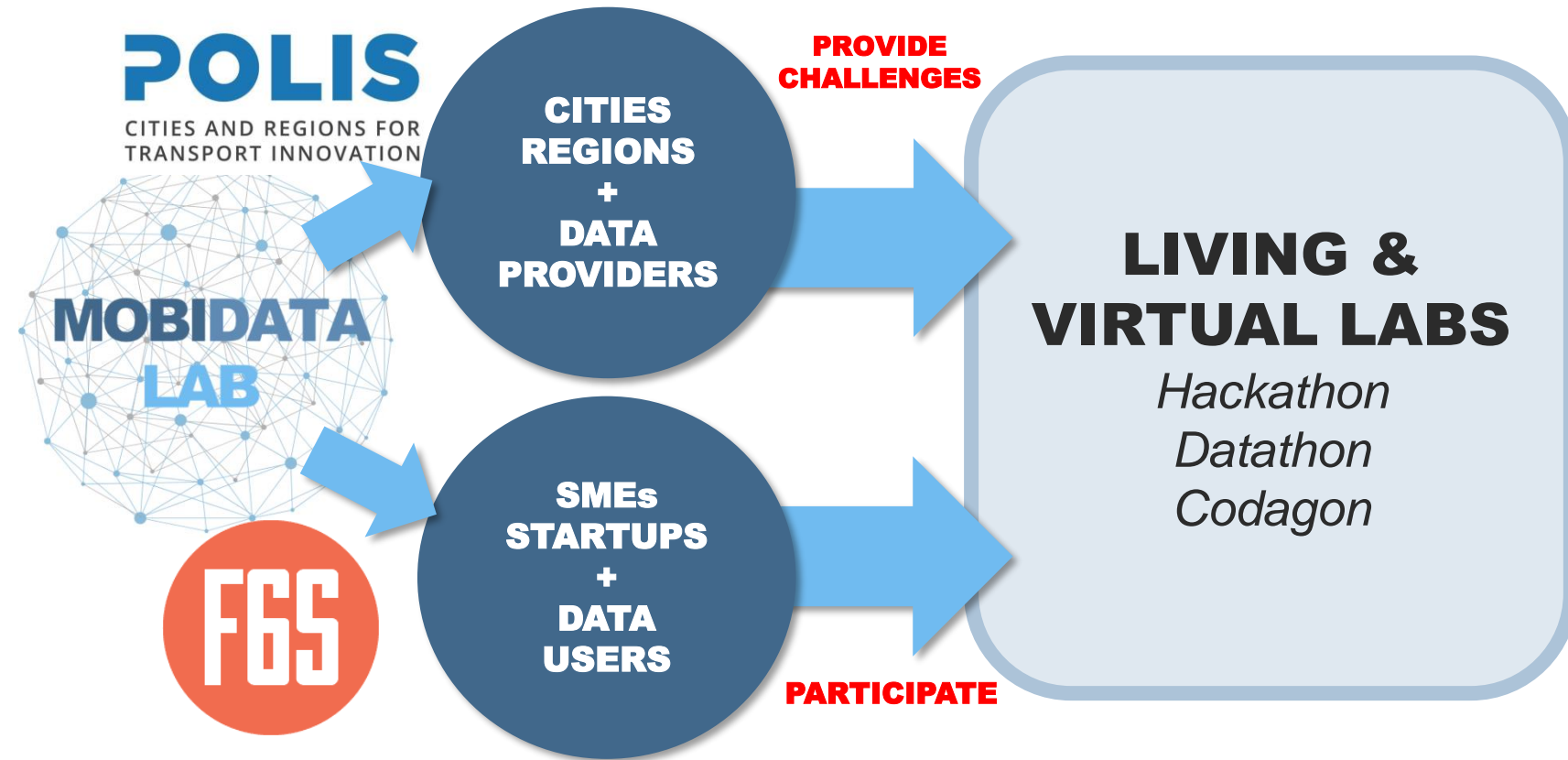


Living & Virtual Labs

- Gather representatives of mobility data providers and consumers
 - Conduct innovation sessions to co-create, explore, experiment, using (open) data as a tool

Living & Virtual Labs

OUR METHOD TO IMPROVE MOBILITY THROUGH DATA SHARING



Reference Group
of mobility stakeholders



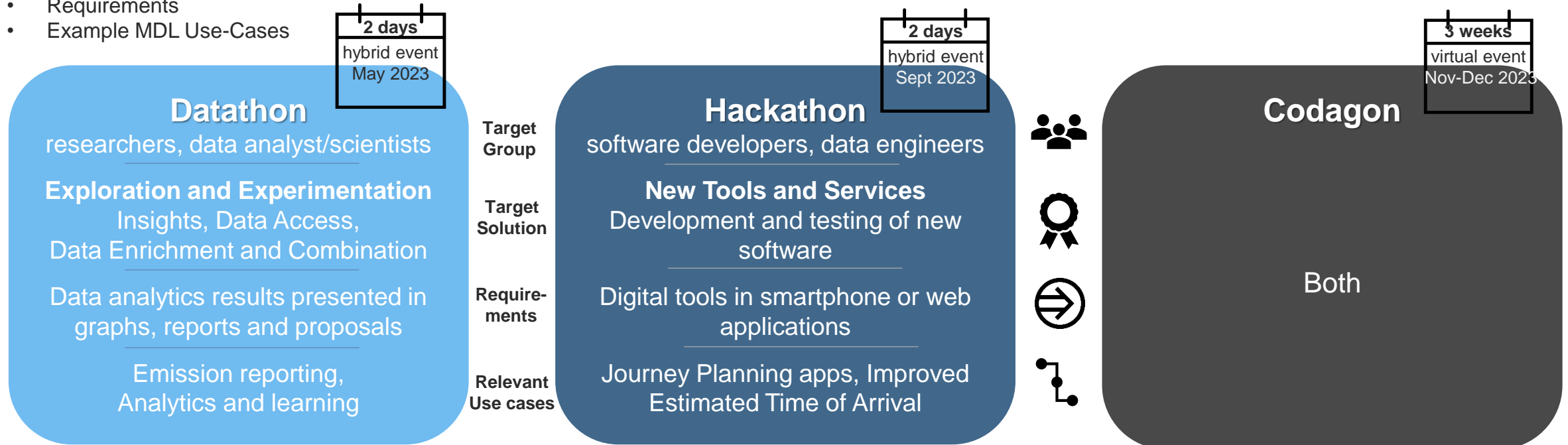
Living & Virtual Labs

DATATHON, HACKATHON, CODAGON

MDL Living Labs are envisioned to be open innovation environments aiming to design and develop new concepts and solutions, based on available **datasets** and **data providers' challenges**, that generate value or efficiencies to end-users/ citizens/ corporates and communities.

For the 3 MDL Living Labs the following characteristics were defined:

- Target participants
- Type of expected outcome
- Requirements
- Example MDL Use-Cases



| **Mobility challenge definition**

ONE-ON-ONE INTERVIEWS WITH THE REFERENCE GROUP MEMBERS

Reference Group: relevant mobility stakeholders (local/regional authorities, transport authorities, transport associations/clusters, transport operators and others).

Role:

- **Provide challenges** corresponding to concrete mobility problems that data analysis, visualisation and enrichment may help solve.
- **Participate in the Living Labs:** challenges will be proposed to a community of innovators and data users in hackathons/datathons during the year 2023.
- **Validate project results** and provide customised feedback

| Initial mobility challenges



**Optimisation of
transport flow
and ETA**



**Emission
reporting**



**Analytics and
learning**



**Journey
planning
services**



**Mobility as a
Service**



Use cases for operations

Use cases for research



**Inclusive
transport**



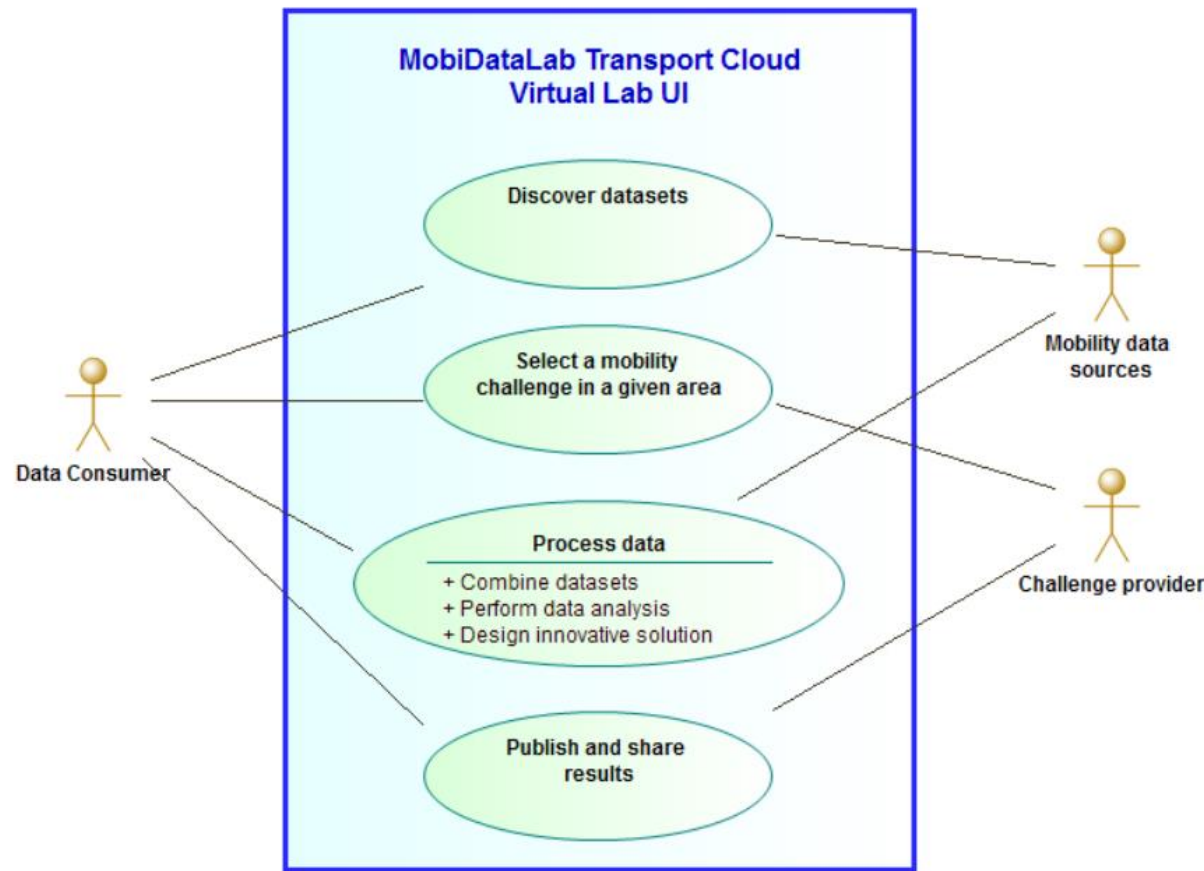
**Sustainable
transport**



**Linked Open
Data**

| Living & Virtual Labs

GENERIC USE CASE DIAGRAM FOR THE VIRTUAL LAB



System: the Virtual Lab User Interface (UI) allowing accessing the challenges proposed by the MobiDataLab stakeholders

Actors

- Primary: the MobiDataLab Data Consumer or participant to Virtual Lab sessions
- Secondary:
 - the mobility data sources made available through the data catalogue
 - the challenge provider: member of the MobiDataLab Reference Group of stakeholders, providing challenges corresponding to concrete problems they face

Use cases

- Discover datasets
- Select a mobility challenge in a given area: the data consumer can explore the challenges proposed by the referring municipalities and select one to take up
- Process data
 - Combine datasets using data access channels and data processors
 - Perform data analysis
 - Design innovative solution to the selected challenge.

| **Transport cloud requirements**

- Virtual Lab interface
- Metadata catalogues
- Service catalogue
- Data access through APIs (API gateway)
- Data access through desktop / SaaS softwares
- Processors, functions, libraries (SDK)
- Anonymisation methods

| Metadata catalogue solutions

Generic catalogues



Rome, Milan, Malaga, Trikala,
Timisoara, Bade Württemberg,
Flanders, etc. + certain NAPs



New York State
data.ny.gov



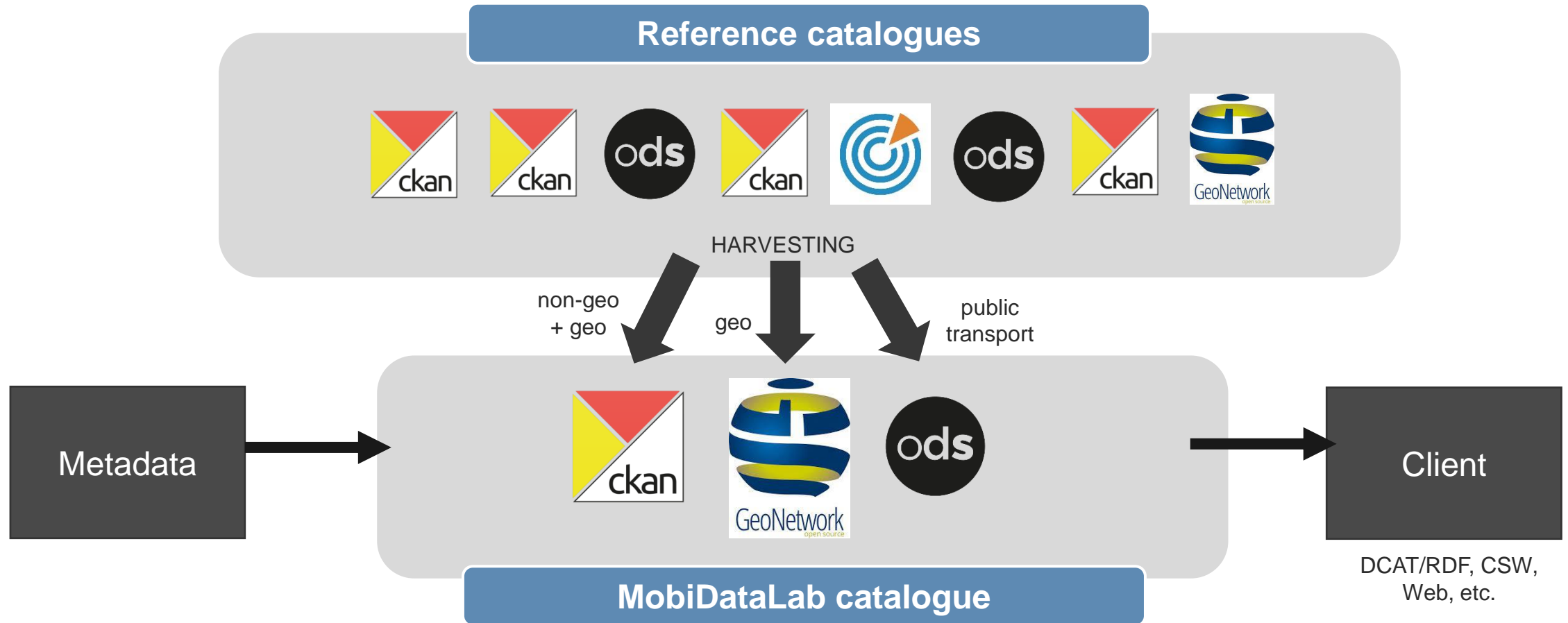
opendatasoft

Eindhoven, Brussels,
Toulouse, etc. + Kisio/Hove

Geo-referenced catalogues



Metadata harvesting flows

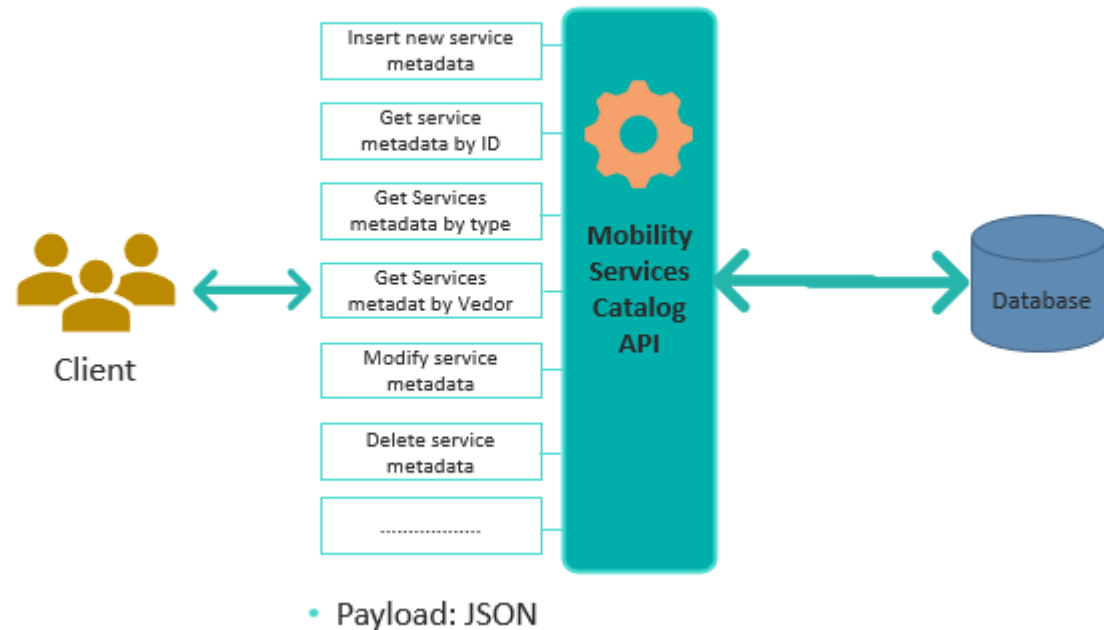


| Service catalogue

1. Mobility datasets (including data API/service)
2. Mobility services
 1. Routing
 2. Geocoding
 3. Traffic
 4.

HERE developer suite, Navitia suite, OpenStreetMap APIs, CARTO Maps, enRoute chouette, Fluctuo data flow, Google Maps, Mapbox, Apple Mapkit, GeoTab ignition, Inrix API, Okina, Otonomo, Populus, PTV, Remix data platform, SharedStreets, Tom Tom, Vianova, TIER mobility, Cubic transportation, TransiTool, etc.

Mobility Services Catalog REST App



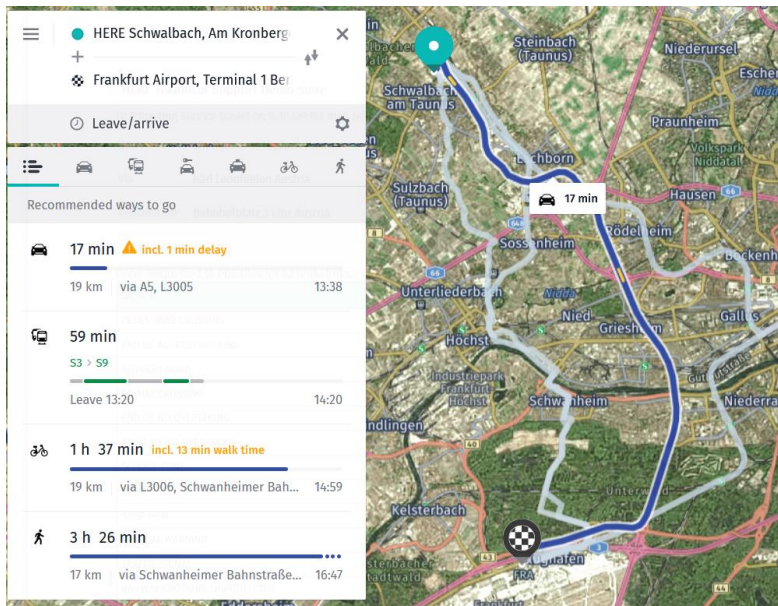
| Interoperability of data

- Mobility data
 - Public transport (TRANSMODEL, NeTEx, SIRI, OJP, GTFS, GTFS-RT)
 - New mobility (MDS, GBFS)
 - Road traffic (DATEX II)
- Metadata (DCAT, DCAT-AP, CSW)
- Data on the web (REST, XML / JSON, OpenAPI)
- Semantic web and Linked Open Data (RDF, SPARQL, OWL, JSON-LD)
- Spatial data (OGC WMS, WFS, API Features, OpenStreetMap, Shape, etc.)

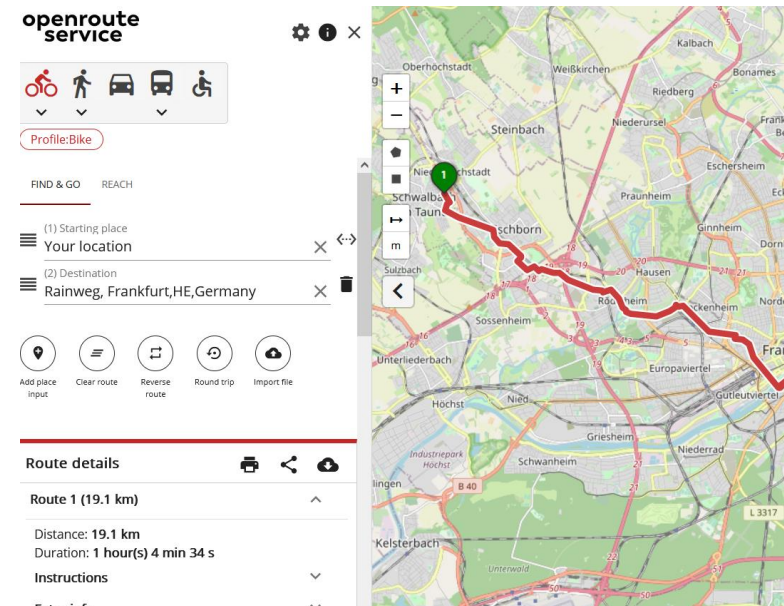
Interoperability of services

Location-based services and their interfaces play an important role to enable consumers to use mobility data for their individual use-cases. However, **services and their interfaces lack interoperability.**

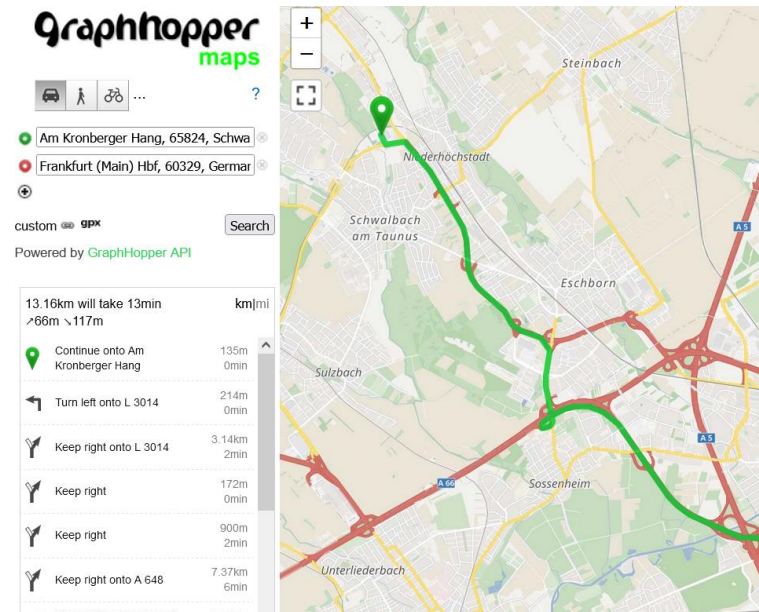
HERE routing



OpenRouteService



graphhopper



| Data access

DURING MOBILITY DATATHON AND HACKATHON



+ participants preferred tools

Desktop / SaaS tools

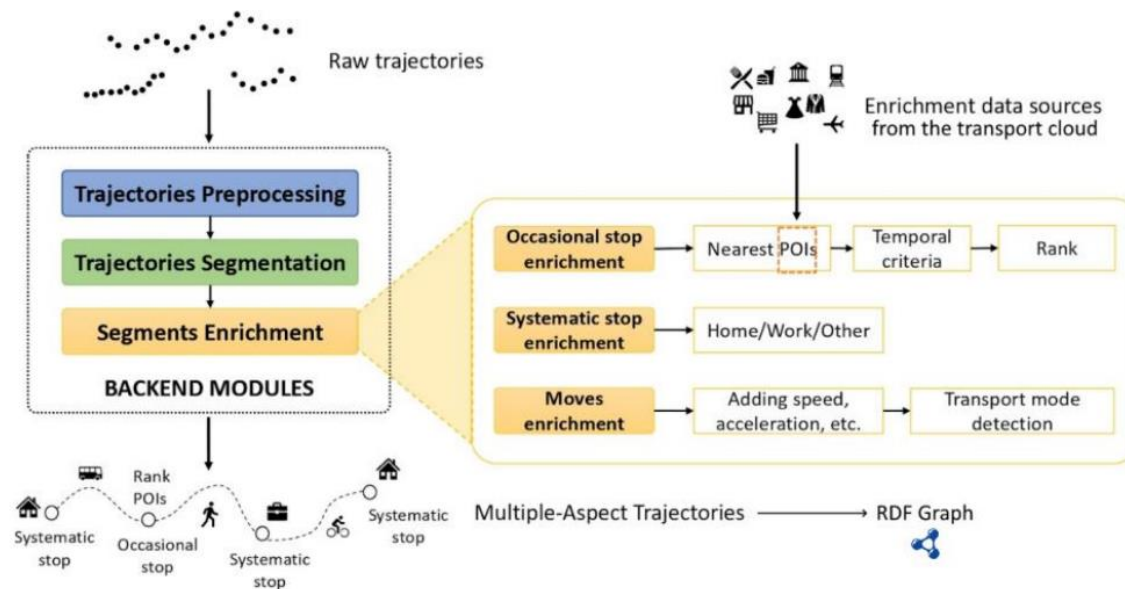


+ on-demand services on
the MDL catalogue

Service APIs

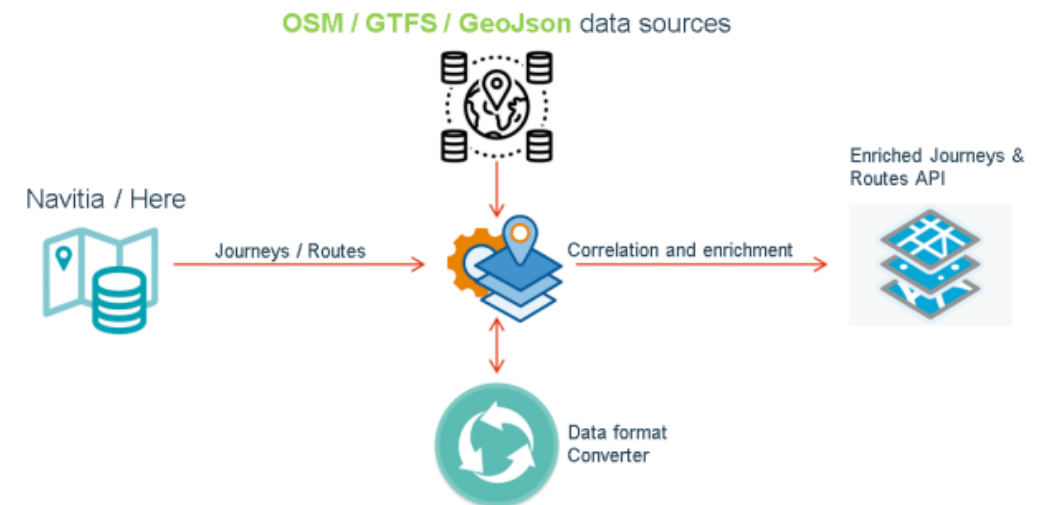
Data processors

SEMANTIC ENRICHMENT OF MOBILITY DATA



A multiple aspect trajectory expresses movement data that is heavily semantically enriched with dimensions (i.e., aspects) representing various types of semantic information (e.g., stops, moves, weather, traffic, events, and points of interest)

GEO ENRICHMENT OF MOBILITY DATA



The mobility data mashup API is a Rest API that consumes a target API and enriches it with additional attributes extracted from a source API and produces the same format as the target API.

Data anonymisation methods

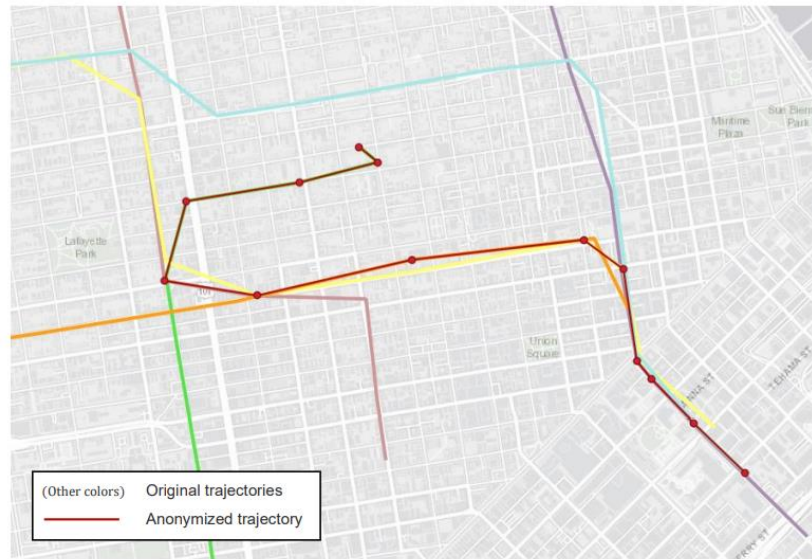
SWAP LOCATIONS

permuting the locations in trajectories among other trajectories



SWAP MOB

swapping segments of trajectories with other trajectories

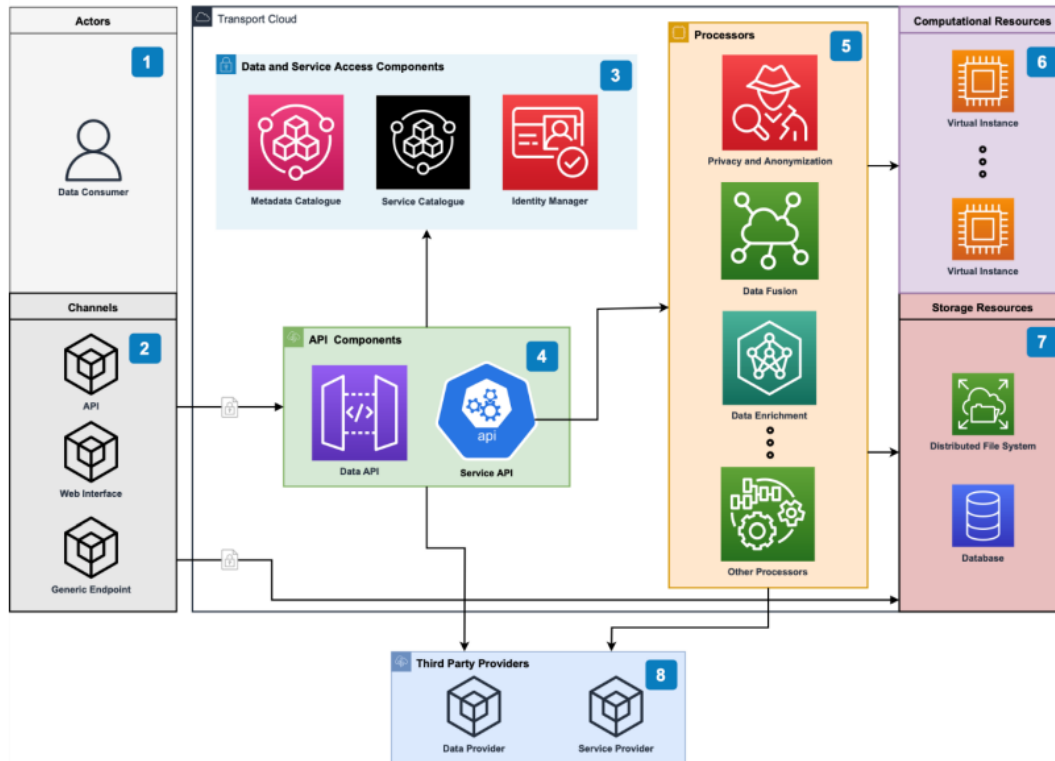


MICROAGGREGATION

partitioning the data set into disjoint clusters containing each at least k similar trajectories and aggregate by replacing them with the cluster centroid



The Transport Cloud



Transport Cloud architecture

CLOUD FEDERATION

The elicited architecture covers cloud federation, governance and use cases requirements by using established standards and components.

METADATA AND SERVICE CATALOGUES

A set of software catalogue systems reviewed on their capabilities are integrated into the Transport Cloud, easing data findability and accessibility.

PRIVACY AND ANONYMISATION

Anonymization mechanisms allow that personal mobility data can still be collected, analyzed and shared with privacy guarantees.

DATA ENRICHMENT

The Transport Cloud includes a toolbox for data set enrichment, containing so-called processors which give the ability to create novel data and services.

| **Key take-aways (at this stage!)**

- Improving mobility data sharing with the European level in mind is more complex, due to :
 - different languages (sounds trivial but is not! – think about metadata)
 - different contexts (geographical, societal, legal, commercial, etc.)
 - sometimes different softwares (e.g. journey planners, metadata catalogues)
- Still a long way to go for making mobility data FAIR
 - the lack of service interoperability is still problematic (e.g. journey planners and MaaS)
- Solutionism (i.e. providing a solution to users before their true problems have been identified) should be avoided ...
- ... but scalability, replicability and sustainability must be anticipated (including for business/cooperation models)

Thank you for listening

Thierry CHEVALLIER, AKKODIS
thierry.chevallier@akkodis.com



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879)

The content of this document reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein. The MobiDataLab consortium members shall have no liability for damages of any kind that may result from the use of these materials.

AKKA



AETHON
ENGINEERING

kisio



KU LEUVEN

F&S

POLIS

here



UNIVERSITAT ROVIRA I VIRGILI

SYN+AIR's Smart Contracts Framework— Methodologies and Data-based Technology Solutions for Improved European Mobility

Ismini Stroumpou/Sparsity Technologies (3rd party of UPC), Project Coordinator

Table of contents

- Problem statement
- SYN+AIR's Smart Contracts Framework (solution)
- Next steps & conclusions

Problem Statement



Joana wants to travel from Barcelona to Brussels and wants to have a door2door seamless journey

Ideal situation is:

1. Single ticket at a reasonable price
2. Real time information
3. Low waiting time at the interchanges
4. Re accommodation in case of disruption



This is the objective of SYN+AIR: to allow TSPs to collaborate in a simple and efficient way by using **SYN+AIR's Smart Contracts Framework**



Transport Service Providers (TSPs) need to collaborate/agree on:

1. Share data e.g., schedules, traveler's ID, assistance-PRM etc.
2. Pricing scheme / Revenue sharing
3. Responsibility sharing

SYN+AIR's Smart Contracts Framework

What is the Smart Contracts Framework?

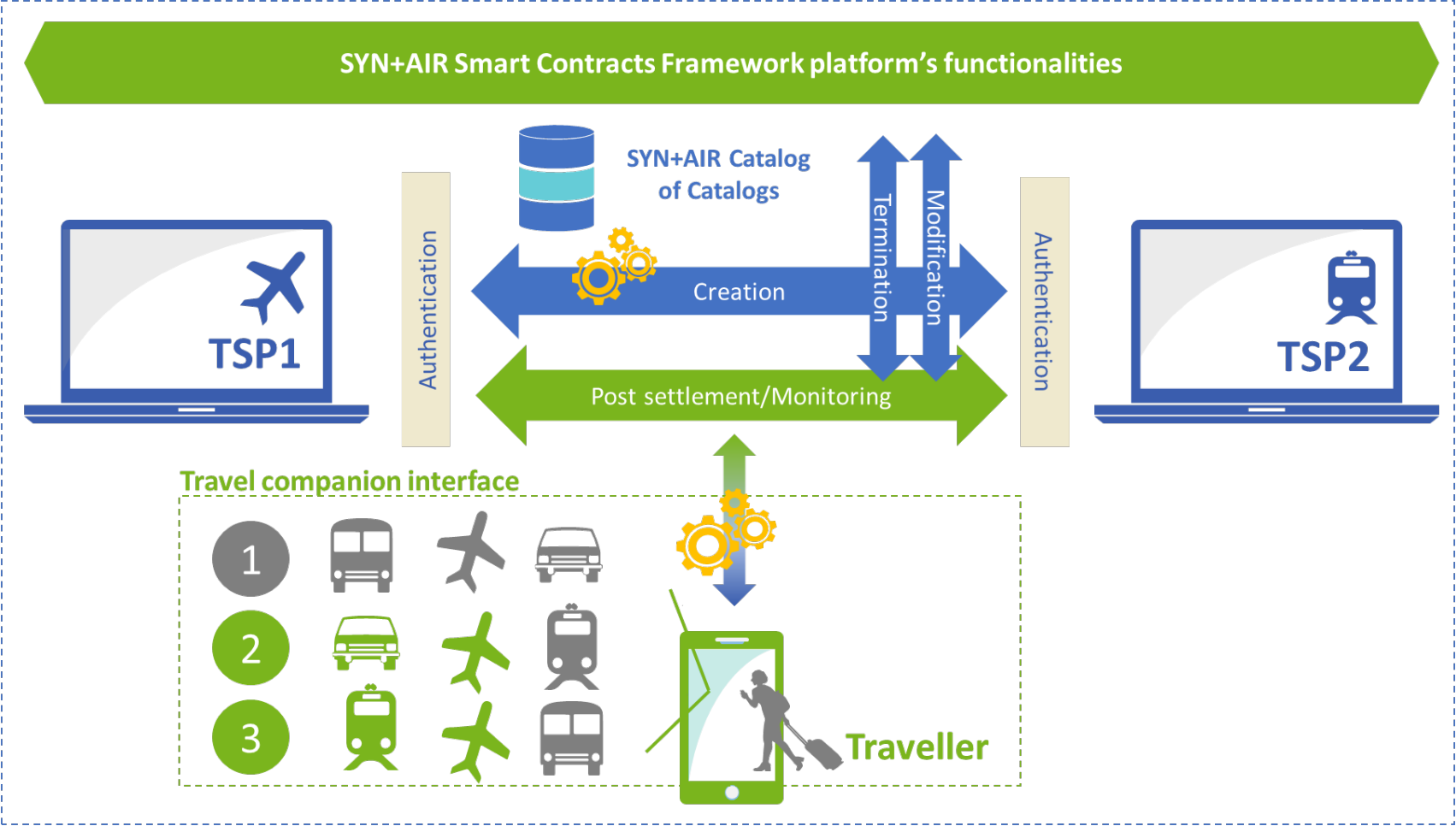
A **business process** that defines data exchange rules among TSPs that share the common goal of getting the passenger to the destination through a multimodal chain of trips

- provides a **centralized hub** for **contracts management** via a web platform (named SYN+AIR platform) that allows TSPs to create, modify, terminate, and validate signed contracts (both data **sharing agreements** and **smart contracts**).
- the storage of TSPs' data and, all the data sharing processes are executed **outside** of the SYN+AIR platform.

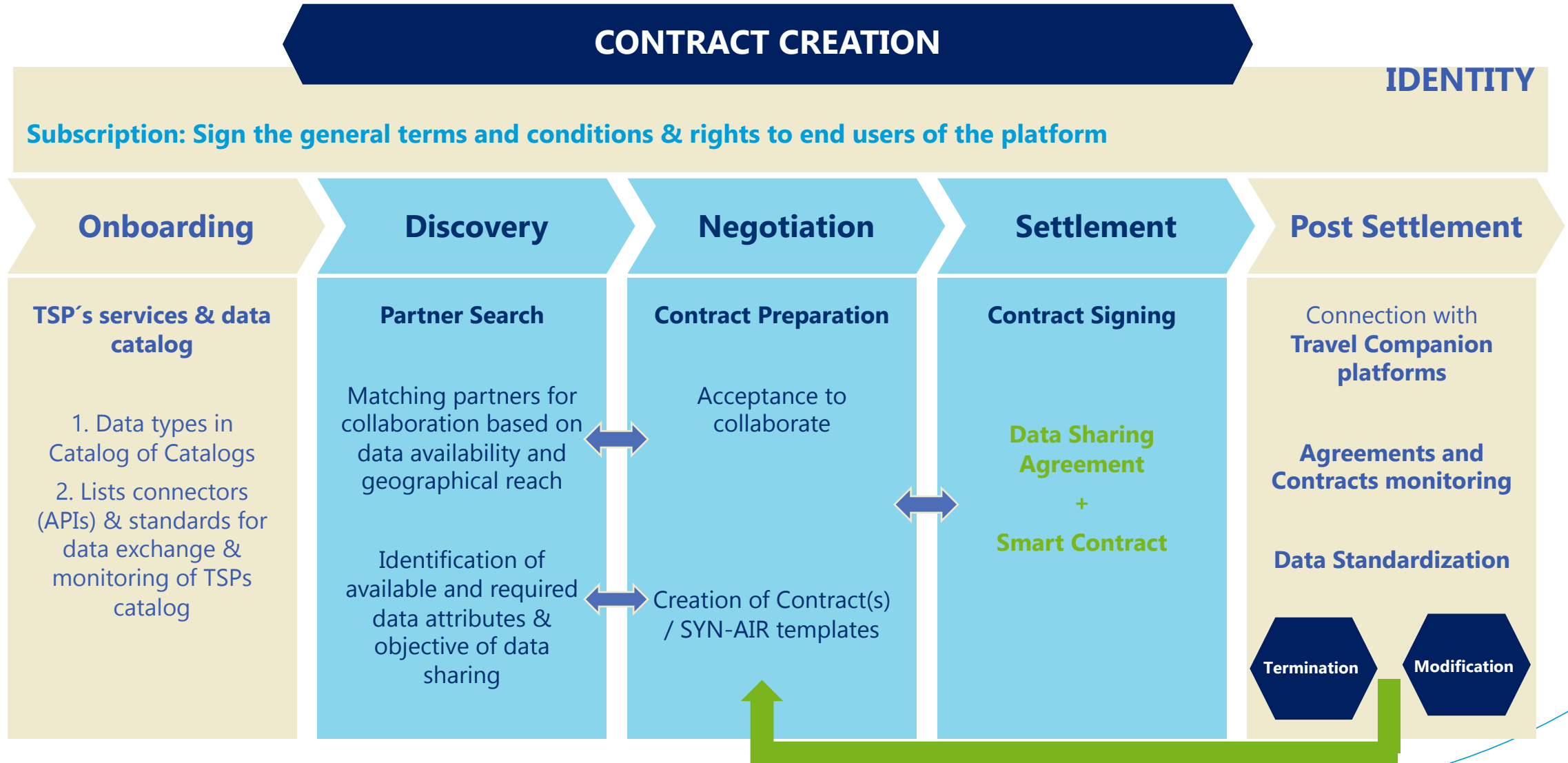
Proposed solution-“Backend”

Components/Actors/Functionalities

Central point to manage the contracts between the TSPs, facilitated by a web platform



Proposed solution (Steps that a TSP need to follow)



1

General terms and conditions

1. under what conditions the SYN+AIR platform is provided;
2. the **rights and obligations of TSPs** when using the SYN+AIR platform;
3. **the (by default) rights and obligations of TSPs when processing the Data** being shared by the means of **Data Sharing Agreements**;
4. **the (essential) rights and obligations of TSPs** deriving from **Smart Contracts**.

2

User agreements

Designation of authorized users (3 types) of SYN+AIR platform

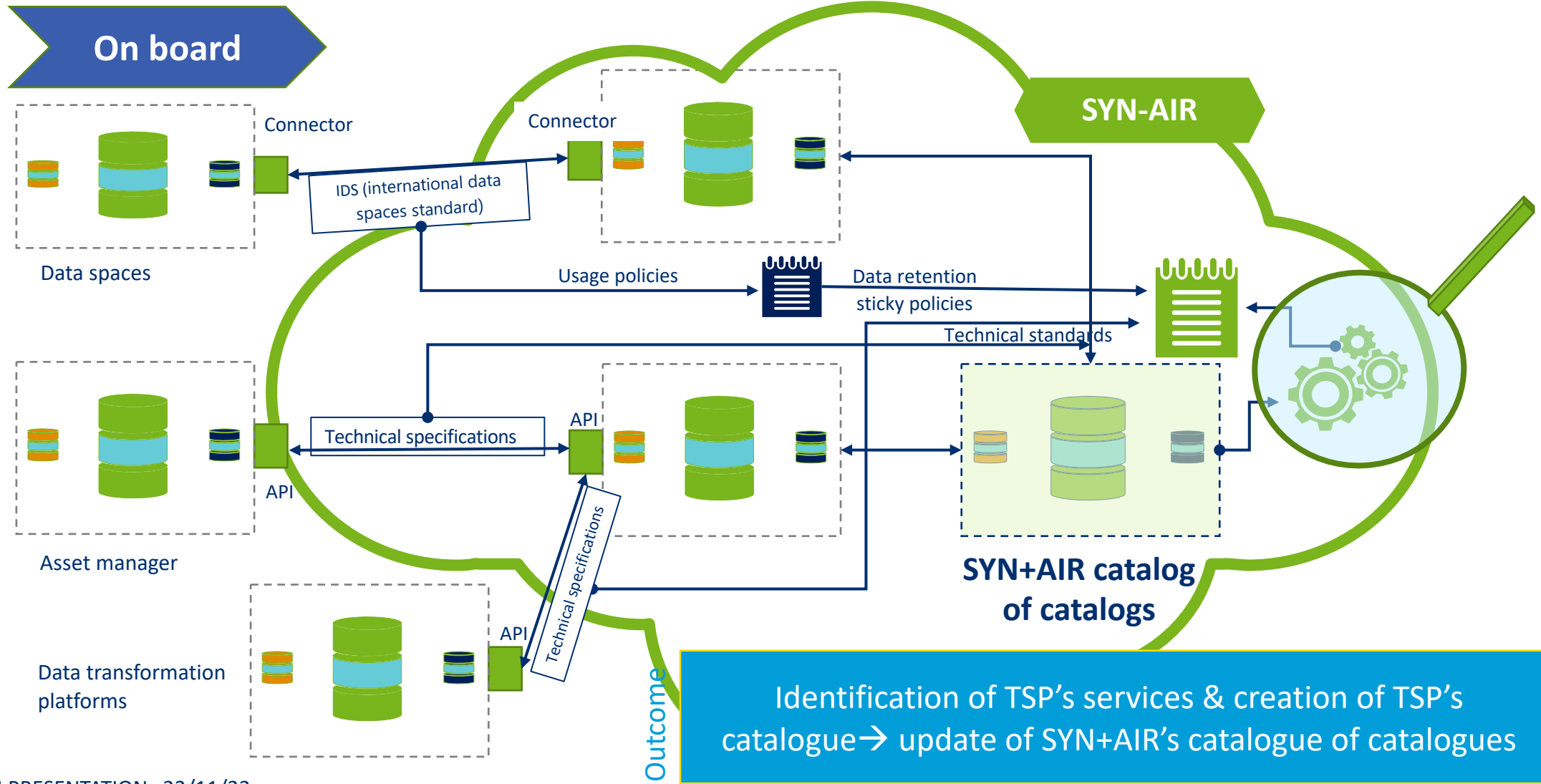
1. Duly authorized and entitled to act on behalf of the legal entity (i.e., director or the CEO) → sign the Data Sharing Agreements and the Smart Contracts;
2. Administrative Representative: to negotiate the terms of the DSA and of the SC on behalf of the TSP;
3. **Technical Representative**: to create and maintain the TSPs' data catalogue(s) in the SYN+AIR platform.

Outcome

Accepted terms & conditions, Assign authorized users of platform

Central management of contracts

De Centralized management of data storage and sharing among TSPs



Discovery-Creation of data matching mechanism

1. Identification of **data generated by each stakeholder per mode and the data that are required to be exchanged to have a seamless journey**
2. Creation of a model composed by different **Data Flow Diagrams** representing trip chains (aka alternatives that a traveller might have for going from A to B (door-to-door))



Air Transport as the main leg:
[Train → Airplane → Train],
[Metro → Airplane → Bus],
[Bus → Airplane → Uber/Taxi] and
[Uber/Taxi → Airplane → Metro]



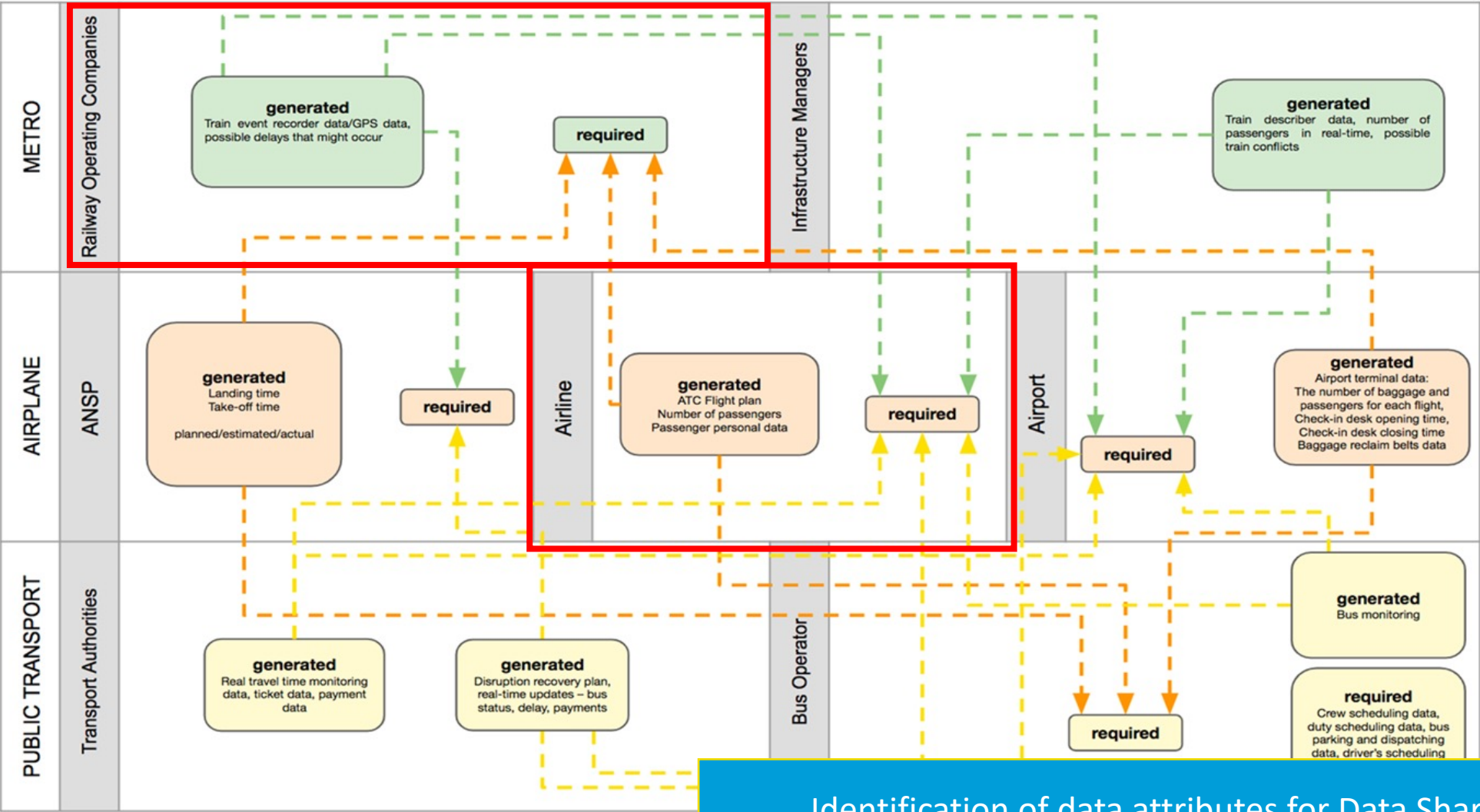
Rail Transport as the main leg:
[Uber/Taxi → Train → Bus],
[Bus → Train → Uber/Taxi] and
[Metro → Train → Metro]



Maritime Transport as the main leg:
[Train → Ferry → Uber/Taxi],
[Uber/Taxi → Ferry → Train] and
[Bus → Ferry → Bus]

Discovery

Matching-example



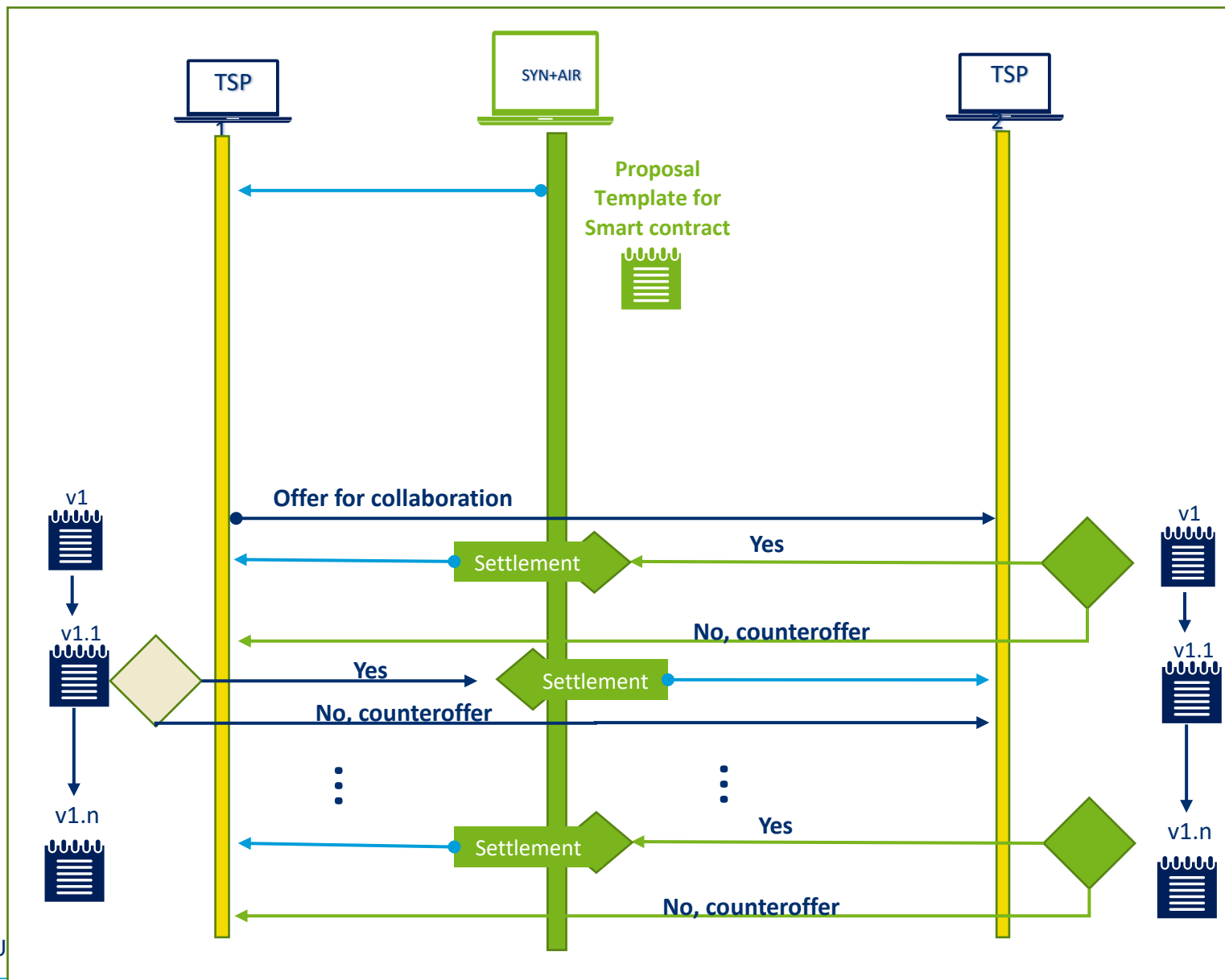
Outcome

Identification of data attributes for Data Sharing Agreements, Creation of negotiation templates

Discovery objectives of data sharing agreements per pair

TSPs pairs	Synchronization of timetables	Optimization of resources	Single Ticketing	Disruption recovery plans and real time updates	Reduce waiting time, reduce number of cars on the access roads	Increase car sharing
ANSPs + Railway operator				X		
Airline + Railway operator	X		X	X		
Airport + Railway operator		X		X	X	
ANSPs + Bus operator				X		
Airline + Bus operators	X		X	X	X	
Airline + Public Transport Authority			X	X		
ANSPs + Public Transport Authority				X		
Airline + Taxi or DRT				X		X
Airport + Taxi or DRT				X	X	X
Railway operator + MaaS operator			X	X	X	
Bus operator + Railway operator	X		X	X		
Ferry + Railway operator	X		X	X		
Ferry + MaaS operator			X	X	X	
Ferry + Public Transport Authority			X	X		
Port Authority + MaaS operator		X		X		
Ferry + Bus			X	X	X	
Ferry + Taxi or DRT			X	X		X
Port Authority + Railway operator		X		X		
Port Authority + Bus operator		X		X	X	

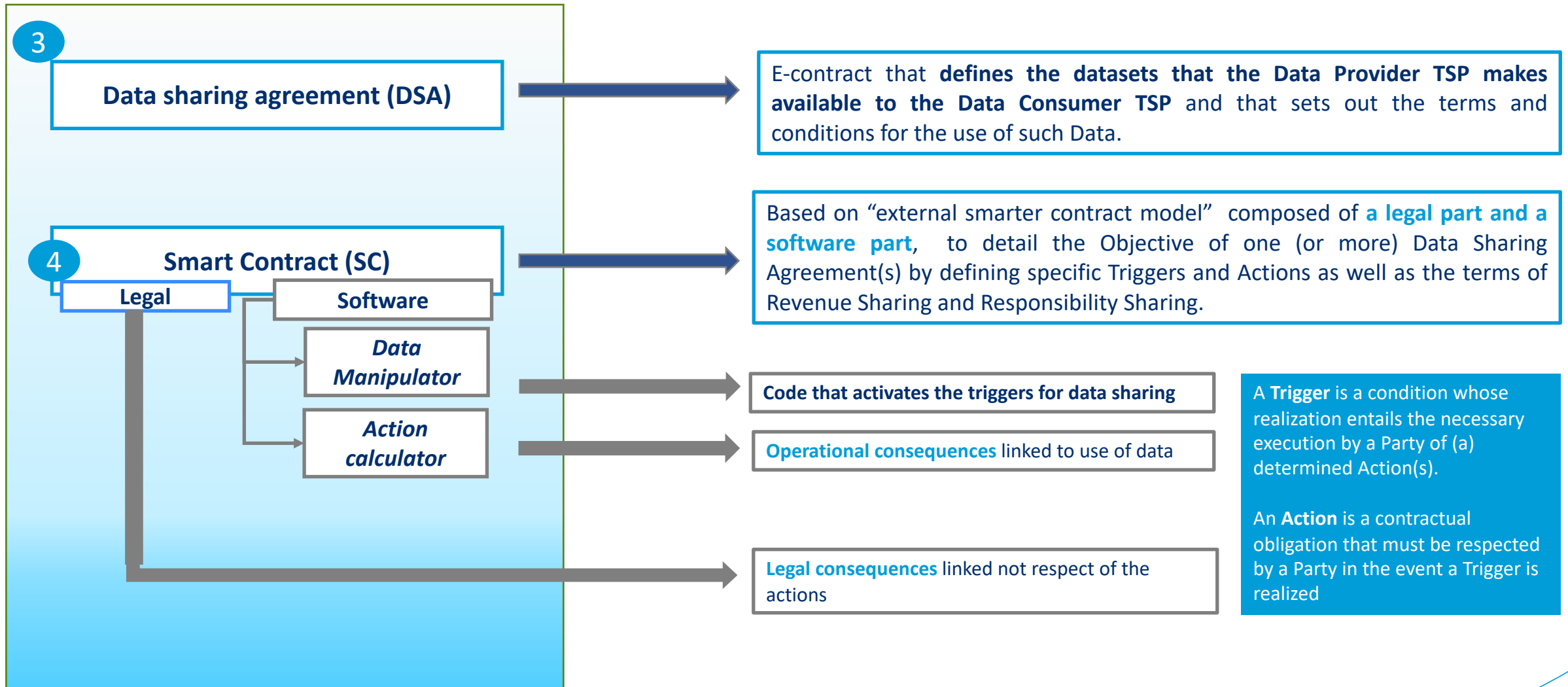
Negotiation process



Outcome

Data sharing agreement(s)
& smart contract

Negotiation/Settlement-legal layer



Negotiation/Example of Smart contract template

Legal part

SC Activation datetime: 01-09-2022 00:00:00

SC Termination datetime: 01-09-2023 00:00:00

Obligations

Until modification or termination of this SC, both parties have the following obligations.

Obligations of both parties:

- Deliver **actions** dictated by the software part which, in summary, are to:
 - Inform other party and Travel Companion (TC)
 - Calculate and take operational decisions when asked
 - Reimburse passengers fully, partially or with voucher depending on the situation

Obligations of TSP1: Coordinate trips with TSP2's flights

Obligations of TSP2: ...

Revenue sharing: The single-ticket's price will cost XX% more/less than the sum of the prices of the separated tickets. Each party will receive their corresponding share of the price using a percentage/pre-determined/other approach for price estimation.

Action – Responsibility sharing:

action 1 – legal consequence

...

action_32 - legal consequence

Signature

Legal Representative of TSP1, Legal Representative of TSP2



Software part

Data Manipulator: As in all templates

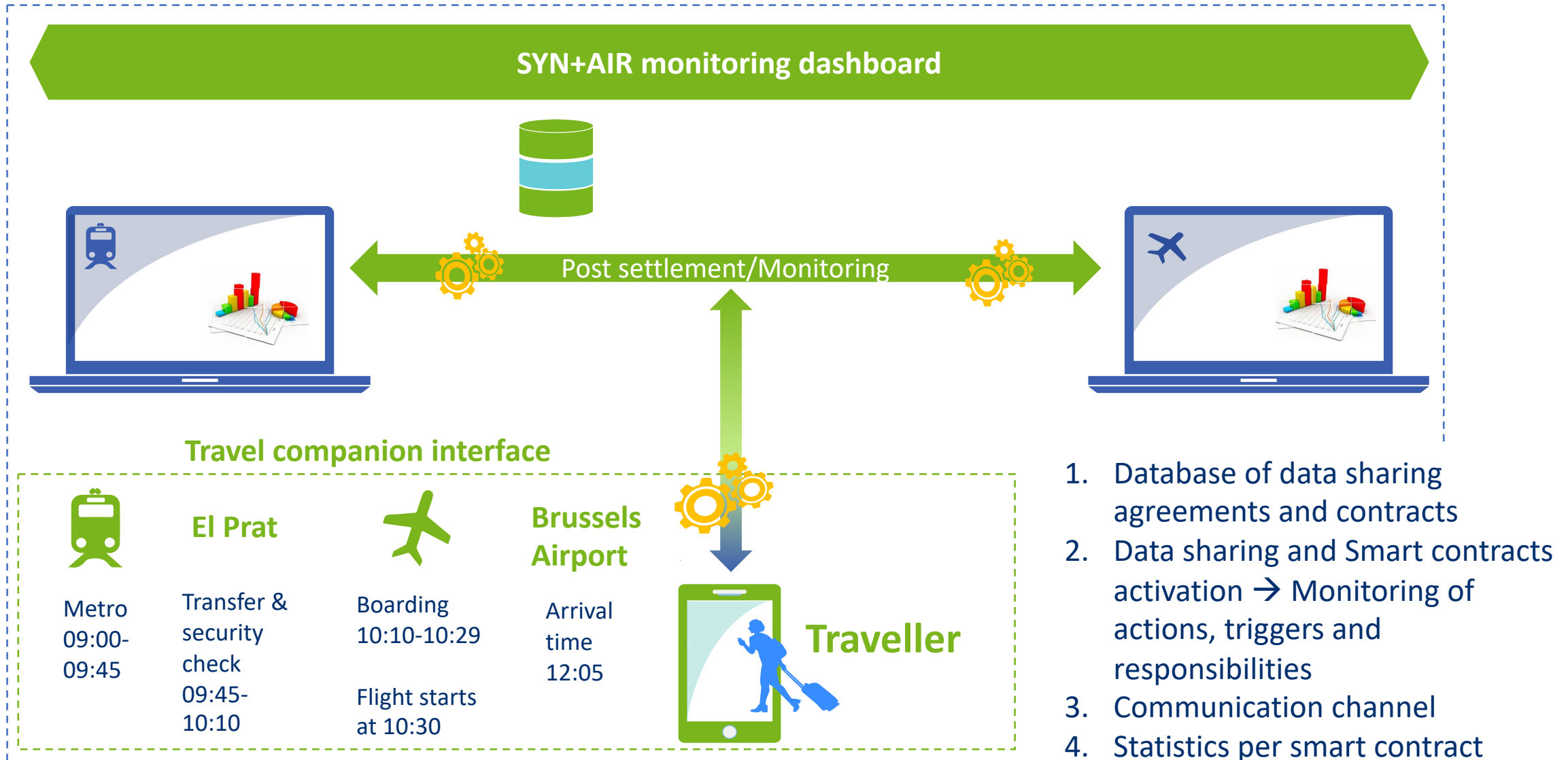
<p>ΑΙΘΙΟΪΛ ΙΝΪΙ ΪΙΝΪ ΙΝΝ ΙΪΘΙ ΝΪΛΙΚΛΥ ΙΛΙ ΝΪΪΪΚΟΚΛΥ</p> <p>C ΑΝΙΙ ΟΝ ΪΚΘ ΪΪ ΕΕΕΝ ΑΕΕΝ ΟΝ ΙΝΛΛΙΪΘ ΪΪ ΛΙΪΙΝΝΙΝΘ ΙΪΘΙ</p> <p>C ΑΝΙΙ ΙΝΝ ΪΪ ΙΝΛΟΝΪΘ ΟΝ ΪΙ ΛΚΝΘΙΛΚΛΥ ΙΝΝ ΟΝΚΥΙΥΝΝ ΙΛΙ ΛΝΟΚΙΟ ΪΪ ΑΪΟΚΝΛ ΑΪΛΙΟΛΙΟΝΝ</p> <p>C ΑΝΙΙ ΟΝ ΛΝΟΚΙΟ ΪΪ ΟΘΝ ΝΪΝΟΚΙΝΕ ΝΝΙΝΙΘΟΝΝ ΙΪΝΟΘ ΪΪΚΝ ΙΝΝΝΙΝΝΛΙΚΛΥ ΙΪΟΚΝΛ</p> <p>C ΑΝΙΙ ΟΝ ΛΝΟΚΙΟ ΪΪ ΝΙΝΝΙΛΙΥΝ ΟΚΙ ΪΪ ΕΝΙΘΙΛ ΑΝΕΝΙΛΚΝΛ</p>
--

Action Calculator: Calculates required actions by the parties. It is grouped by scenarios and phases.
Actual code below

[illegible]

¹ The Data Manipulator will always be listening for triggers like this

Post settlement/monitoring



Next steps & conclusions

Next Steps & conclusions

The SCF is a co-created **concept** with TSPs and mobility stakeholders.
Each step performed by TSPs to sign a data sharing and/or a smart contract will pave the way towards a better travel experience

The next steps are:

- 1) **To develop and test SYN+AIR's platform**
- 2) Test and assess the platform by "**translating**" a real data sharing contract e.g. between a airline and a railway operator and upgraded to also to create a smart contract to accompany the data sharing
- 3) Demonstrate to TSPs that data sharing can unlock new opportunities
- 4) To engage and use results from other EU initiative such as MobiDataLab, Data4PT, Catena-X, NAPCORE etc.

To achieve seamless D2D transport collaboration is mandatory at all levels

Identify where MobiDataLab findings can be used

CONTRACT CREATION

IDENTITY

Subscription: Sign the general terms and conditions & rights to end users of the platform

Onboarding

TSP's services & data catalog

Knowledge Base of the MobiDataLab

- 1. Data types in catalog of Catalogs
- 2. Lists connectors (APIs) & standards for data exchange & monitoring of TSPs catalog

Discovery

Partner Search

Matching partners for collaboration based on data availability and geographical reach

Identification of available and required data attributes & objective of data sharing

Negotiation

Contract Preparation

Acceptance to collaborate

Creation of Contract(s) / SYN-AIR template

Settlement

Contract Signing

Data Sharing Agreement
+
Smart Contract

Data Sharing business and revenue models of the MobiDataLab

Post Settlement

Connection with **Travel Companion platforms**

Agreements and Contracts monitoring

Data Standardization

Termination

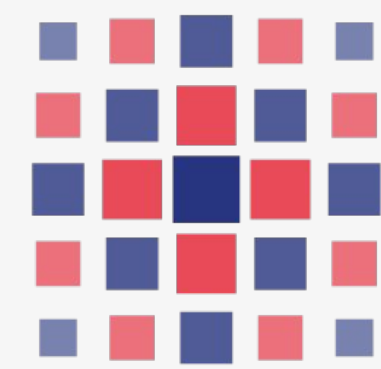
Modification



Q & A

For more information you can always download our deliverables from the project website <http://syn-air.eu/>

THANK YOU FOR
YOUR ATTENTION



SOBIGDATA
RESEARCH INFRASTRUCTURE

European Research Infrastructure for Big Data and Social Mining

Luca Pappalardo

ISTI-CNR, Pisa


www.sobigdata.eu

Why SoBigData RI

Responds to the rising demand for cross-disciplinary data-driven research and innovation

- ▶ Democratising the benefits of data science and Big Data within an ethical framework that harmonizes individual rights and collective interest
- ▶ Focus on **Social Mining** to understand and model complex social phenomena
- ▶ Open Data ecosystem: adequate means for accessing big social data together with algorithms for extracting knowledge from them



- 
- *Social mining is a toolbox aimed at understanding and modelling complex social phenomena and their evolution.*
 - *Combines **data-driven and model-driven analytic approaches** (network science, data mining, and AI): “what is” and “what if.”*

Exploratories

Vertical **contexts** fostering tangible progress towards grand societal challenges



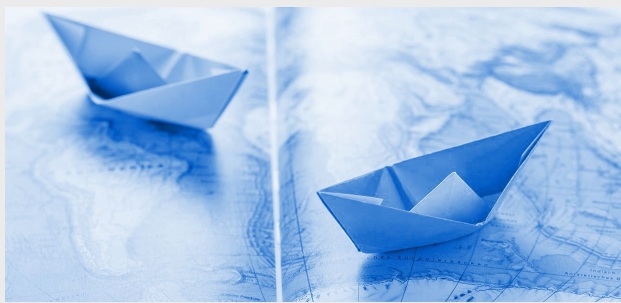
Societal Debates
and
Misinformation
Analysis



Demography,
Economy
and Finance 2.0



Sustainable
Cities for Citizens



Migration
Studies



Social Impact of AI
and explainable
machine learning



Sports
Data Science



Network
Medicine



Disaster response
and recovery



Societal and
industrial impact of
Next-Generation
Internet and beyond
5G networks



Sustainability and
management of
future networks and
Cyber- Physical
systems



Pervasive
intelligence
modeling for future
society



Sustainable Cities for Citizens



► Models for flows generation

- A Deep Gravity model for mobility flows generation. Nature communications
Nature communications 2021

► Mobility Analytics for official statistics

- Evaluation of home detection algorithms on mobile phone data using individual-level ground truth
EPJ data science, 10. 2021

► Mobility and COVID-19

- Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures
Scientific reports, 2021

► Urban impact of AI

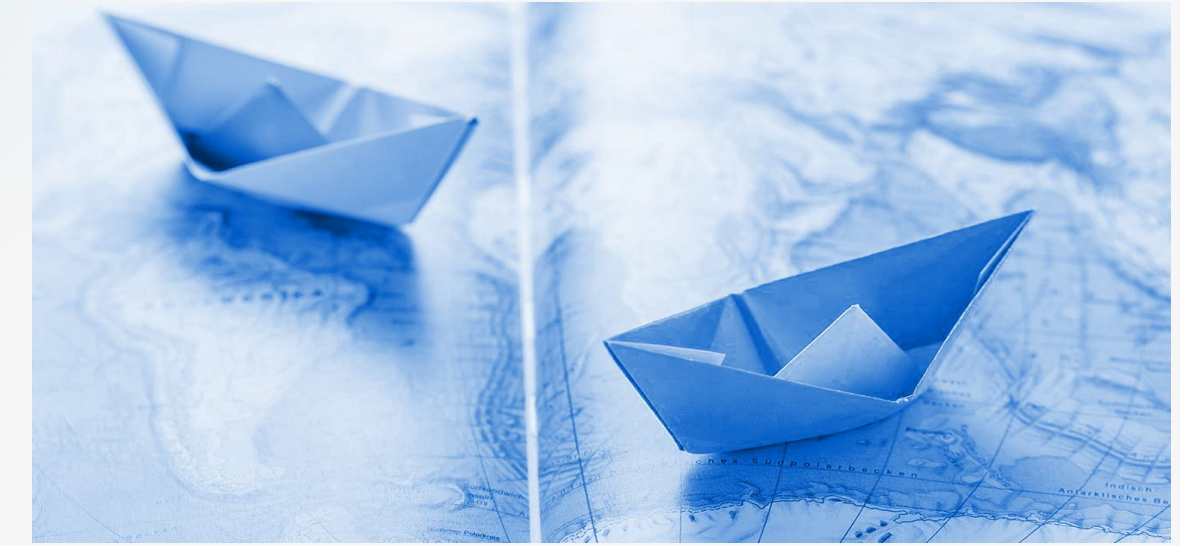
- Connected Vehicle Simulation Framework for Parking Occupancy Prediction
SIGSPATIAL 2022
- How Routing Strategies Impact Urban Emissions
SIGSPATIAL 2022

► Estimation of CO2 emissions and pollution

- Gross polluters and vehicle emissions reduction
Nature Sustainability 2022



Migration Studies



► Detecting migration from social media

- Digital Footprints of International Migration on Twitter
Intelligent Data Analysis, 274-286. 2020.
- Human migration: the big data perspective
International Journal of Data Science and Analytics, 1-20. 2020.
- Home and destination attachment: study of cultural integration on Twitter
Scientific reports, 2021

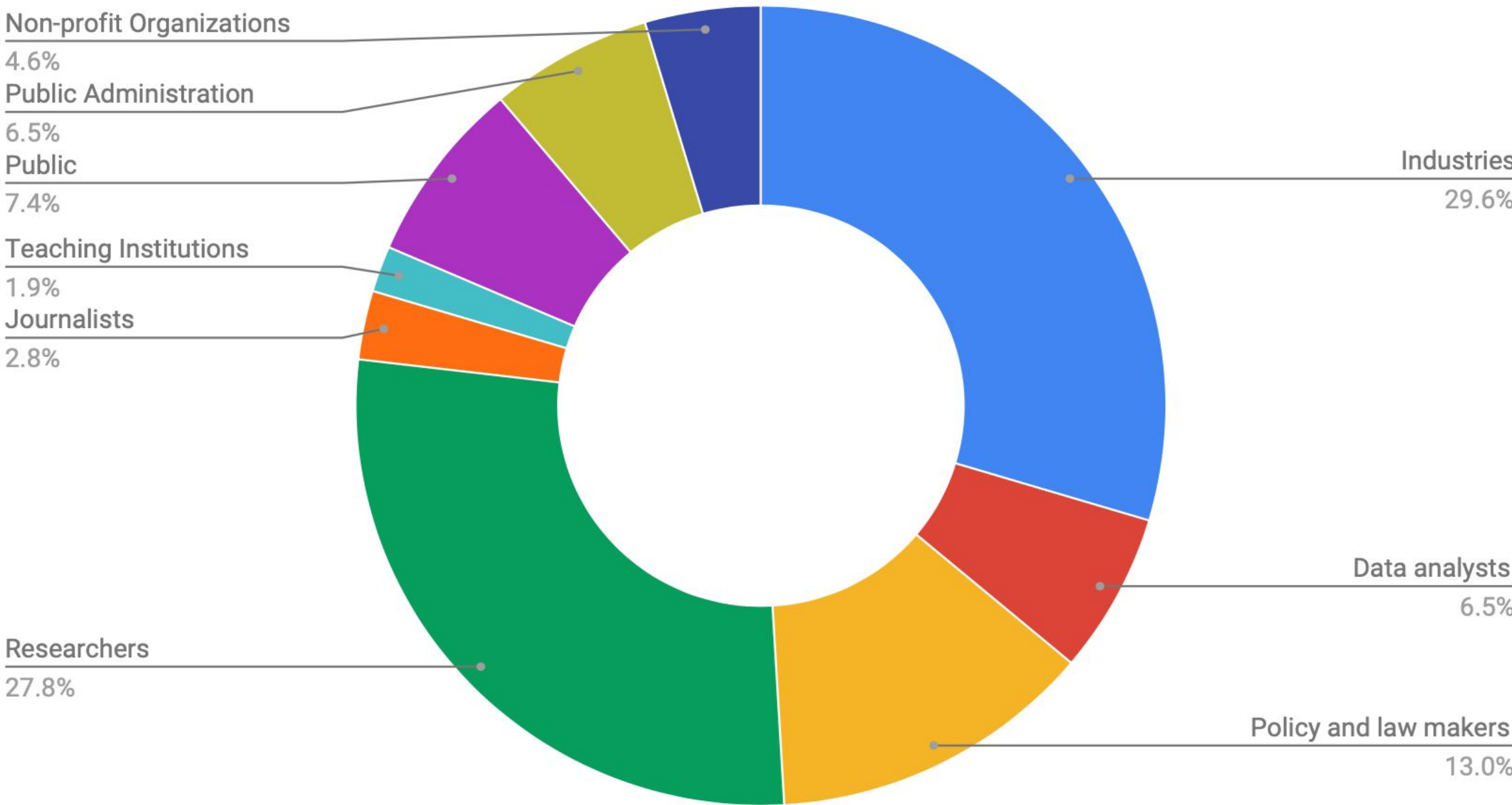
► Measuring international migration and friendships

- Internal migration and mobile communication patterns among pairs with strong ties
EPJ Data Science 2021
- Turnover in close friendships.
Sci Rep 2022
- Dataset of Multi-aspect Integrated Migration Indicators
Data paper under revision

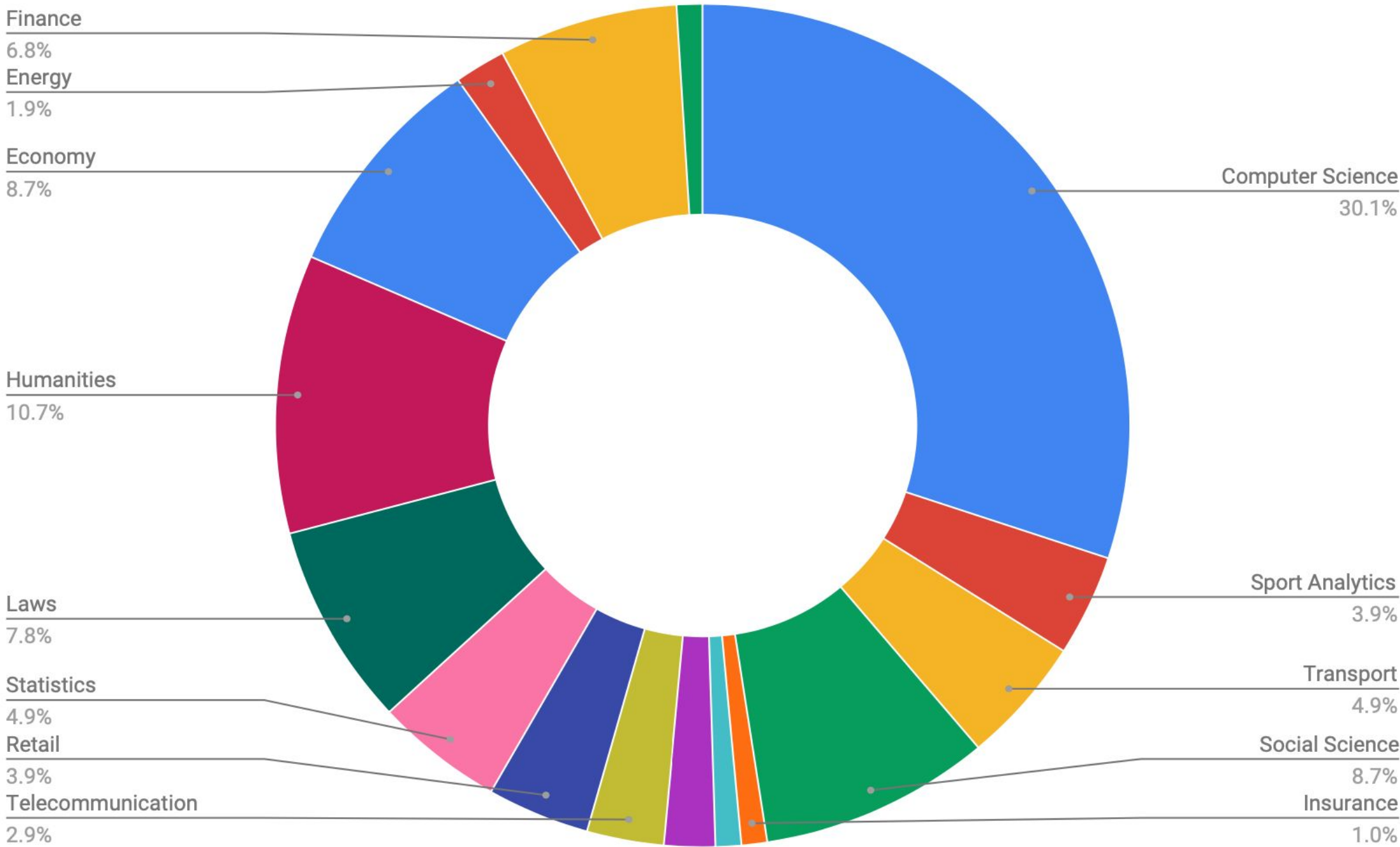


Multi-disciplinary panorama

Type of Stakeholder

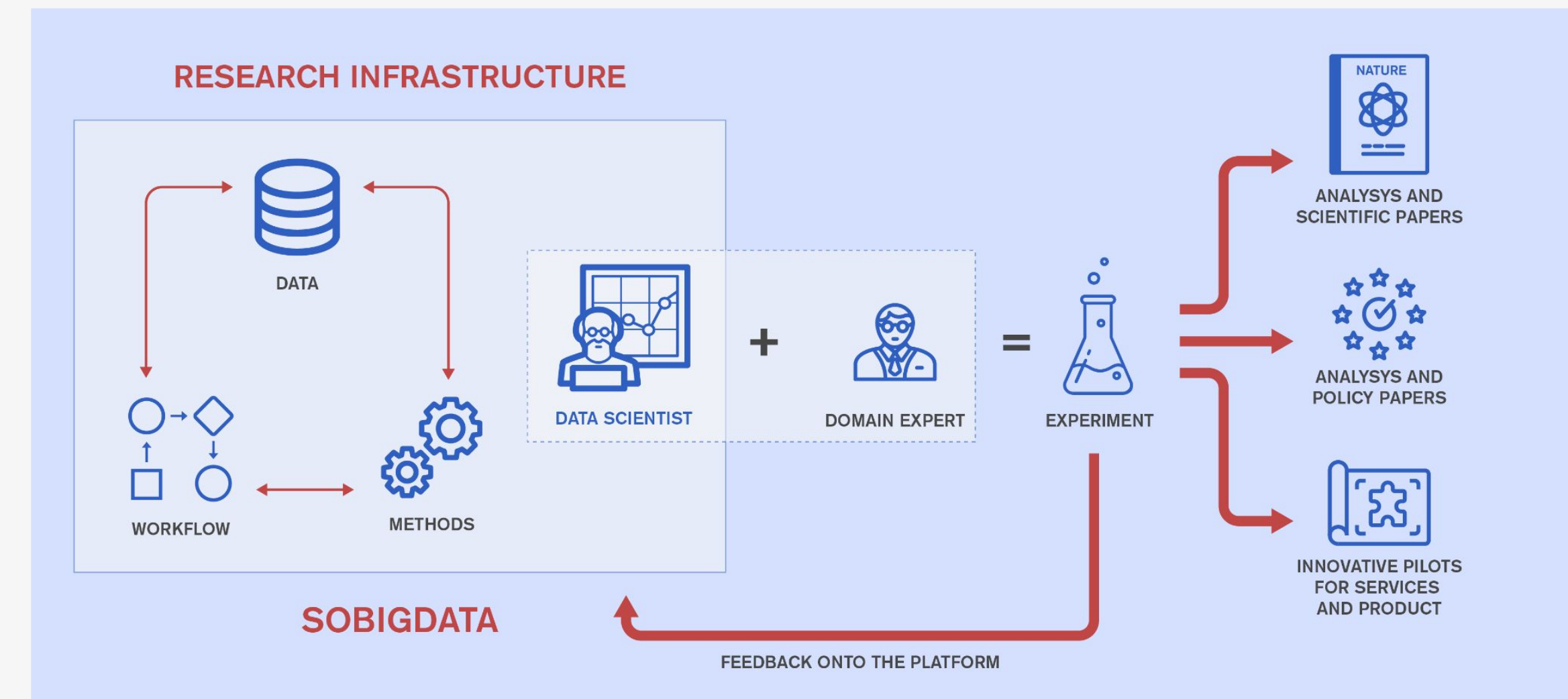


Application field



Responsible Data Science principles

- to act as a Wind Tunnel for innovative and responsible data science science: a place to experiment, to train, to share ideas and to learn.
- to operationalize the ethical values within Big Data Analytics.
- **Responsible open Data Science is FAIR + FACT (ethics by design)**
- to transfer this view to industry and PA which are not always aware of the responsible data science pipeline and how to exploit its full potential in a sustainable way.



FIRST AID FOR DATA SCIENTIST

This course rises within the EU SoBigData project: we developed this online course in order to make sure that all users are familiar with the basic elements about: ethics, data protection, and intellectual property law.

Access to the platform

Username

Password

[Log in](#)

[Forgotten your username or password?](#)

[New account](#)

An “operational” Ethical and Legal framework

designed to function as a factory for sensing, understanding ethico-legal issues and data literacy in digital society:

guidelines and white papers related to Big Data and AI issues and **ethics based service** .

Operational ethical and legal board performs first-aid clearance of the experiments.

Services (Open Science tools)

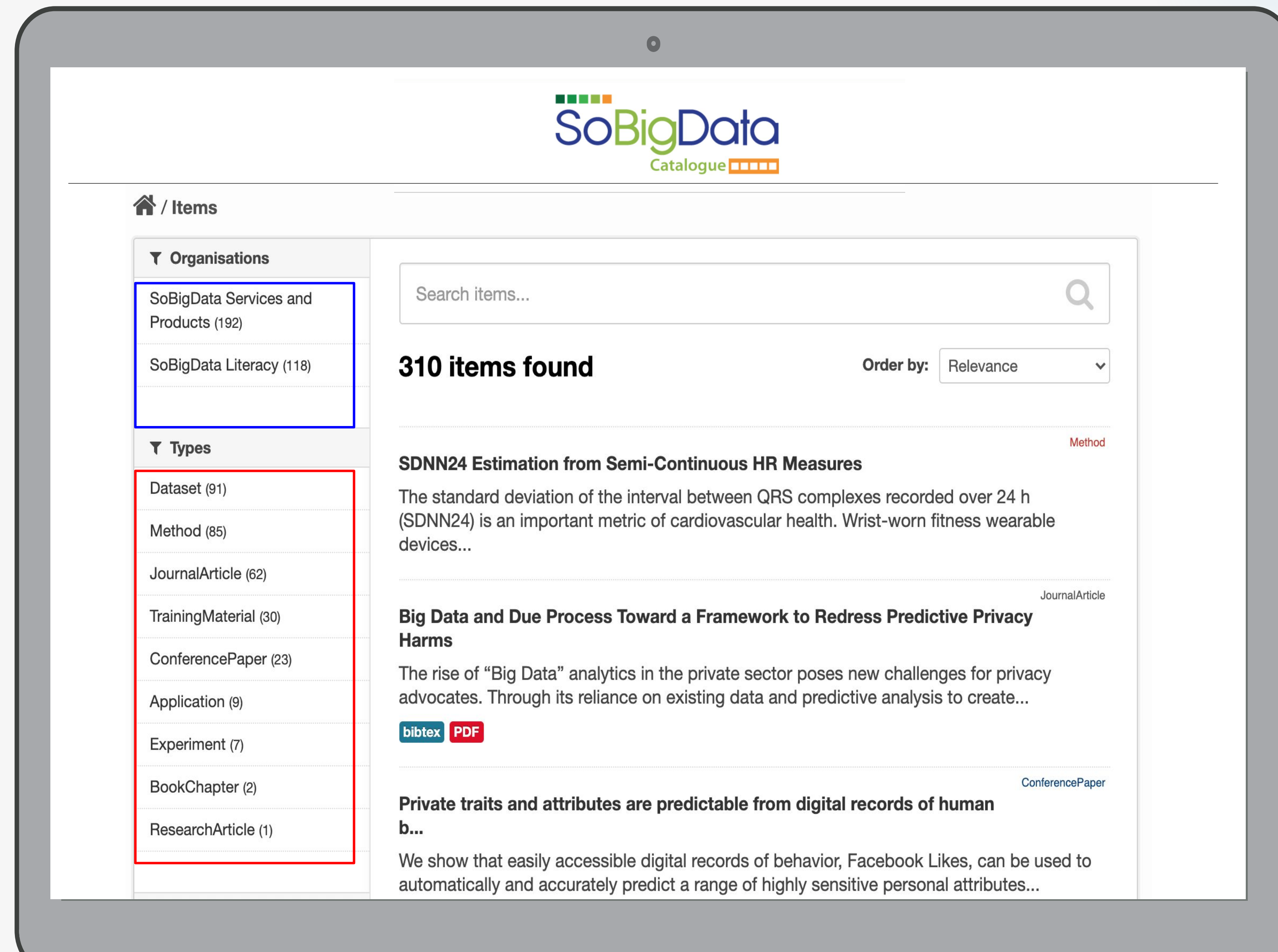
- ▶ **Virtual Research Environments:** Dedicated areas where users may share information and experiences focused on a specific topic. It has a dedicated shared workspace and may be equipped with dedicated computational resources
- ▶ **Workspace:** A cloud storage services which may be shared among users
- ▶ **Catalogue:** The Catalogue is a tool for finding and exploring Datasets, Methods, Applications, Experiments, and publications through different navigable views.
- ▶ **SoBigData Interactive Programming environment:** the users develop their methods on a Jupyter Notebook using cutting edge libraries developed by experts.
- ▶ **Methods Engine:** a cloud service to execute methods integrated in the RI exploiting the interconnection of servers (completely transparent to the final users).
- ▶ **Training Area:** A set of online modules for the responsible data scientist (work in progress)



Services (People and Facilities)

- ▶ **Transnational Access:** a mobility program to host researchers in our nodes
- ▶ **HPC Access:** access to high performance computing facilities owned or connected to one of our nodes
- ▶ **High level expertise:** Time and effort of the multi-disciplinary researchers distributed in our nodes.
 - ▶ Researcher
 - ▶ Legal and Ethical evaluations
 - ▶ Technology transfer (Training)





- The Catalogue is a tool for finding and exploring Datasets, Methods, Applications, Experiments, and publications through different navigable views.
- FAIR and FACT principles are implemented through the metadata, identity management and accountability



- ☒ **Default Standard - 2GB RAM / 2 cores**
The Default notebook server includes Python, R, Julia, Octave and Java kernels and a number of community libraries preinstalled for Python. 3 flavors are available: Standard, Small and Medium
- ☐ **Default Small - 4GB RAM / 4 cores**
- ☐ **Default Medium - 8GB RAM / 4 cores**
- ☐ **Master in Big Data Analytics & Social Mining notebook server - 2GB RAM / 2 cores**
The notebook server is preconfigured with Python libraries for Data Science. It's meant for the students of the Master course in Big Data Analytics & Social Mining from the University of Pisa

Start

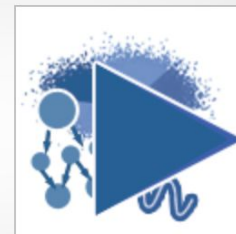
Access to the Data Space

The data space contains the set of input and output data sets of the users. It is possible to upload and share tables. Data sources can be chosen from those hosted by the infrastructure. Outputs of the computations can be even saved in this space.



Execute an Experiment

This section allows to execute or prepare an online experiment or method. The section is endowed with a list of algorithms for executing models for various application domains.



Check the Computations

This section allows to check the status of the computation. A list of processes launched by the user is shown along with meta-information. By clicking on the completed jobs it is possible to visualize the data set contents.



	Societal Debates	Demography, Economy & Finance 2.0	Sustainable Cities for Citizens	Migration Studies	Sport Data Science	Social Impacts of AI and Explainable Machine Learning
Text and Social Media Mining (TSM)						
Complex Network Analysis (CNA)						
Human Mobility Analytics (HMA)						
Web Analytics (WA)						
Visual Analytics (VA)						
Privacy Enhancing Technology (PET)						

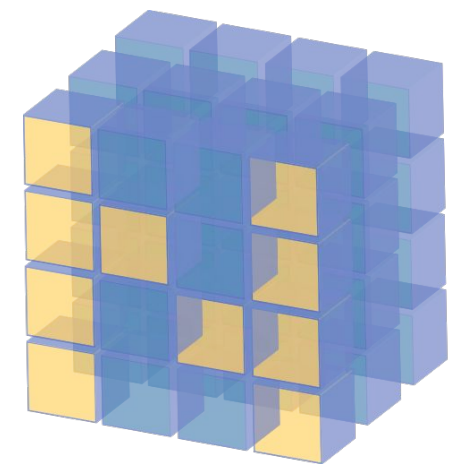
- *SoBigData Interactive Programming environment: the users develop their methods using cutting edge libraries developed by experts.*
- *SoBigData Method Engine a cloud service to execute methods integrated in the RI exploiting the interconnection of servers (completely transparent to the final users).*





scikit mobility

pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



NumPy



SciPy

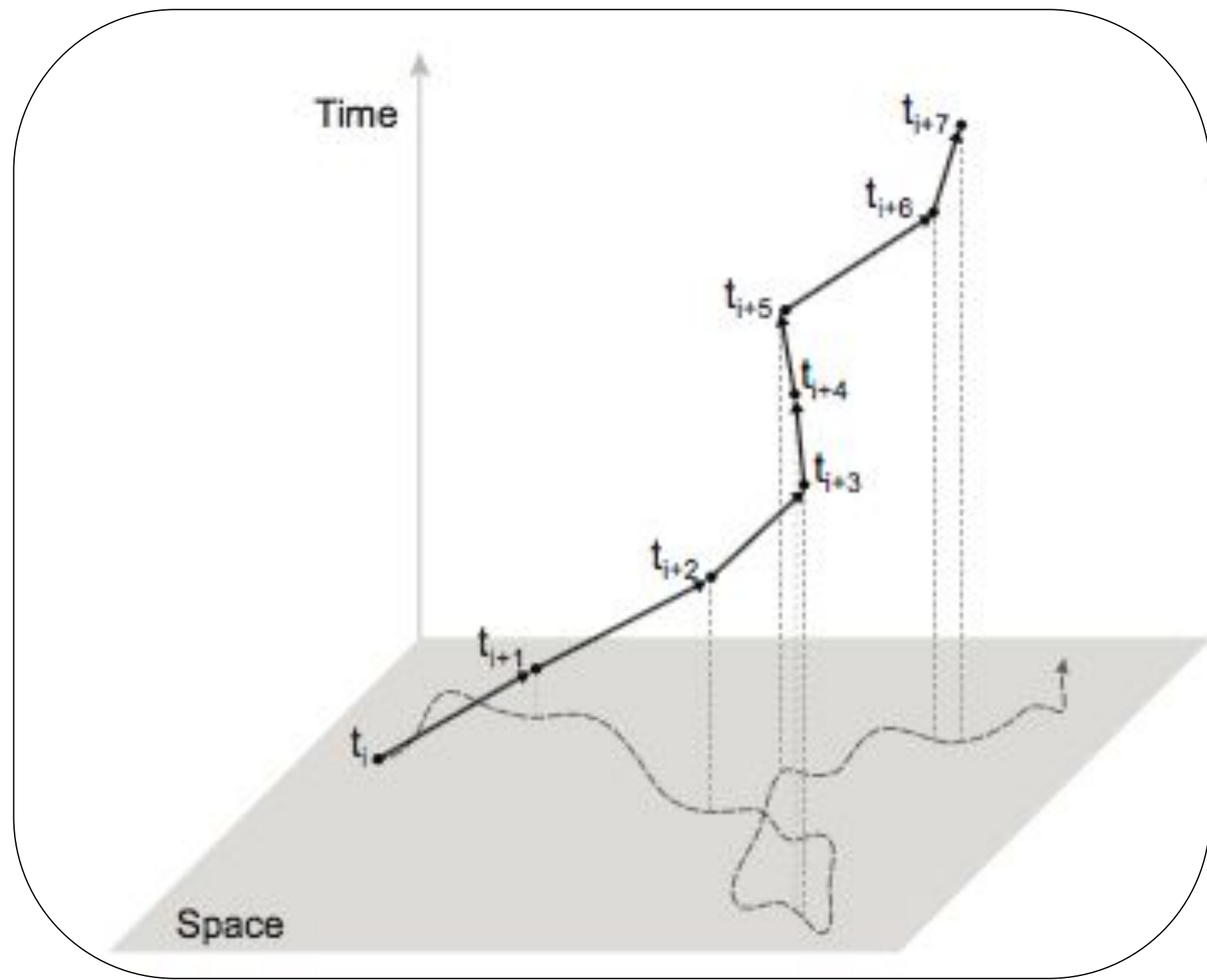


GeoPandas



folium

GeoPy

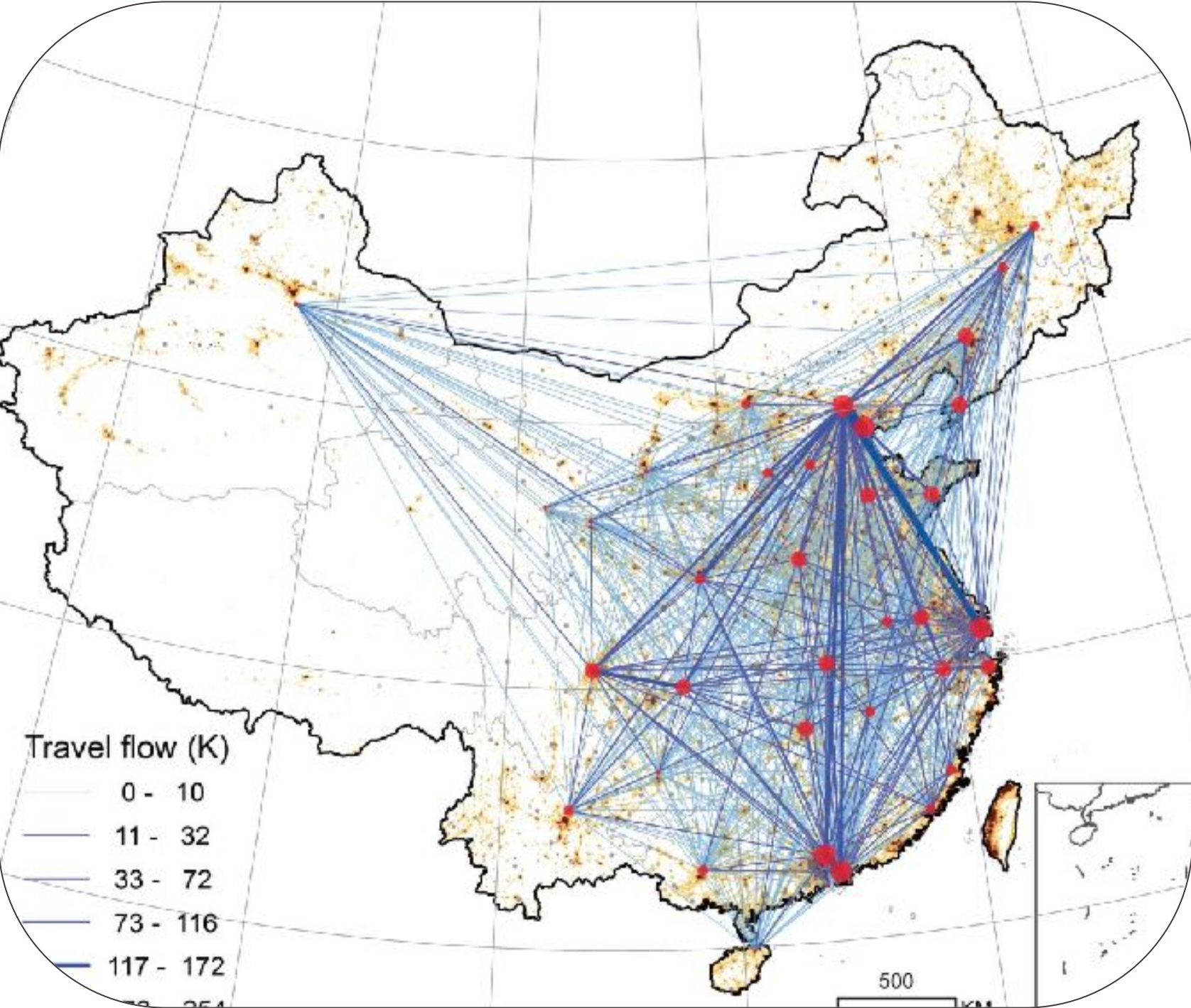


Trajectory

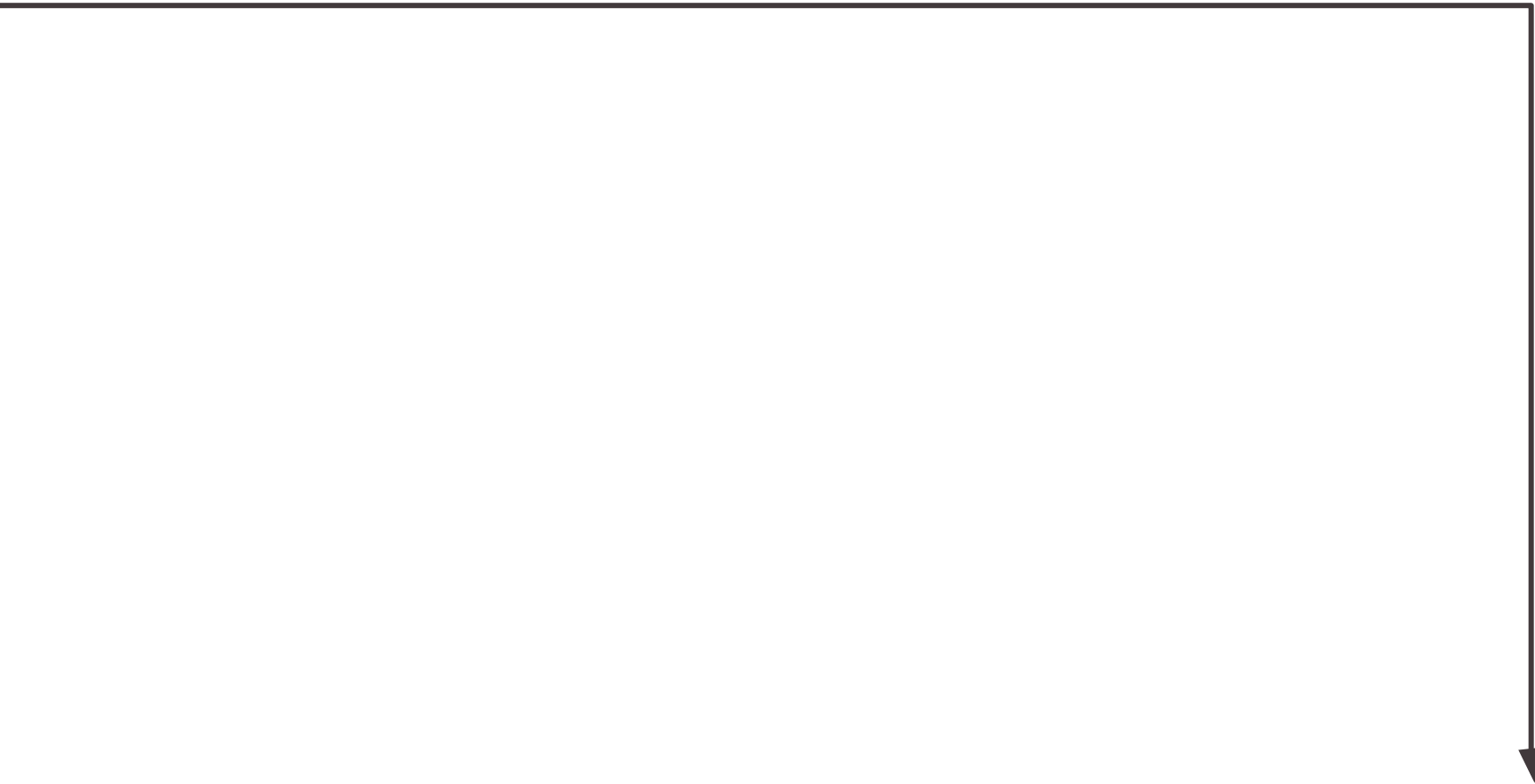


TrajDataFrame

	uid	datetime	lat	lng
0	1	2019-01-01 08:00:00.000000	46.397461	11.829286
1	1	2019-01-01 08:58:07.838255	46.479001	11.807456
2	1	2019-01-01 10:23:02.268305	46.479001	11.807456
3	1	2019-01-01 17:21:08.304837	46.506604	11.782809
4	1	2019-01-01 19:34:35.687601	46.515593	11.783268



Flows



Tessellation

FlowDataFrame

	tile_id	population	geometry
51	36001	304564	POLYGON ((-73.933672 42.76071, -73.809603 42.7...
23	36003	48787	POLYGON ((-78.308611 42.086675, -78.3088390000...
18	36005	1397366	POLYGON ((-73.783519 40.881033, -73.74806 40.8...
57	36007	199346	POLYGON ((-75.850388 42.327731, -75.8437919999...
20	36009	79819	POLYGON ((-79.060708000000001 42.347917, -79.06...

	flow	origin	destination
0	121606	36001	36001
1	5	36001	36005
2	29	36001	36007
3	11	36001	36017
4	30	36001	36019

```
from skmob import TrajDataFrame

filename = 'data/geolife.gz'
tdf = TrajDataFrame.from_file(filename)      # load data

tdf.head(3)
```

	lat	lng	datetime	uid
0	39.984094	116.319236	2008-10-23 05:53:05	1
1	39.984198	116.319322	2008-10-23 05:53:06	1
2	39.984224	116.319402	2008-10-23 05:53:11	1


```
from skmob.preprocessing import detection, clustering

# detect stops and clusters
stdf = detection.stay_locations(tdf, spatial_r_km=0.2)
cstdf = clustering.cluster(stdf, cluster_radius_km=0.1)

cstdf.head(3)
```

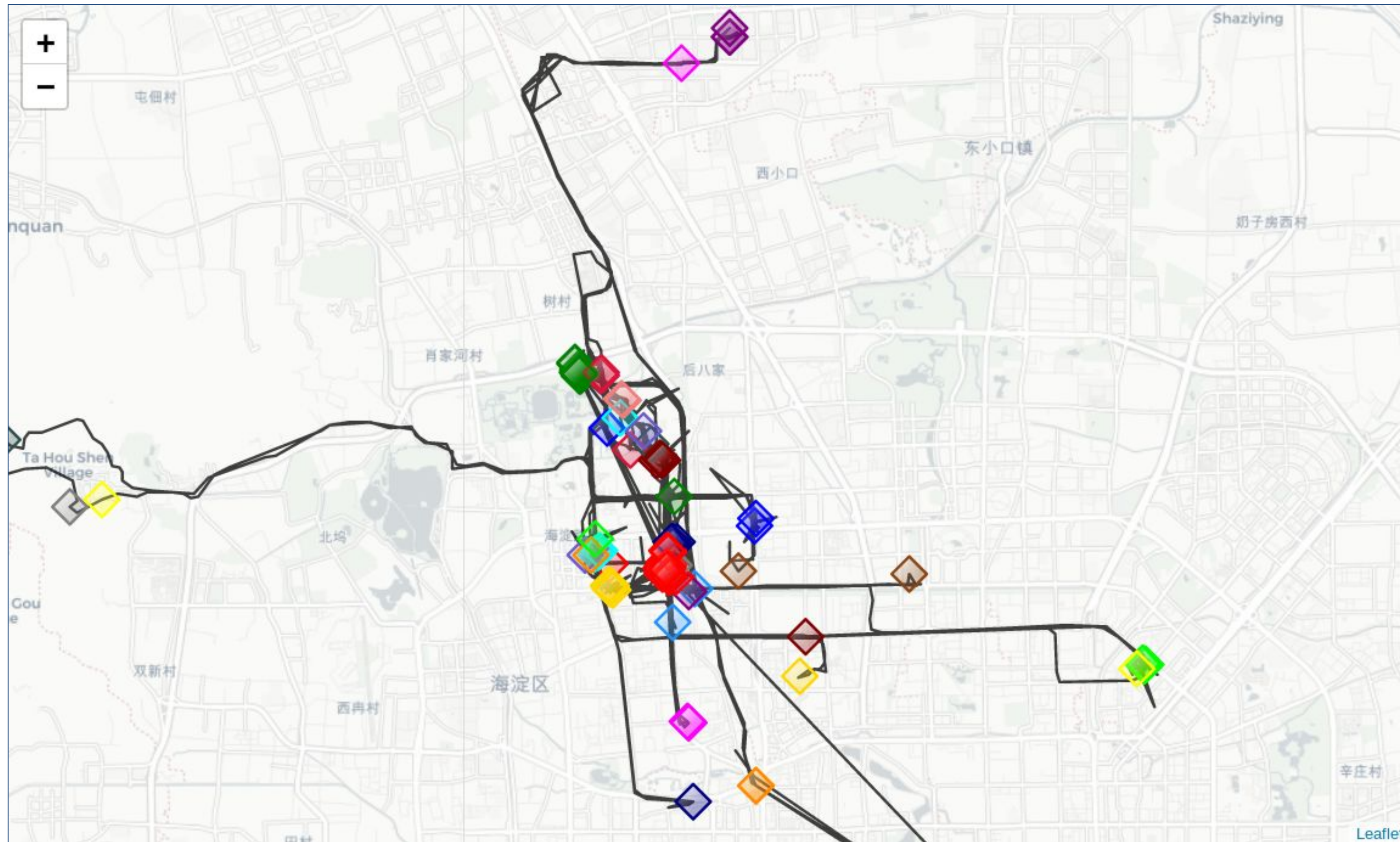
	lat	lng	uid	datetime	leaving_datetime	cluster
0	39.978253	116.327275	1	2008-10-23 14:01:05	2008-10-23 18:32:53	0
1	40.013819	116.306532	1	2008-10-23 19:10:09	2008-10-24 07:46:02	1
2	39.978987	116.326686	1	2008-10-24 08:10:39	2008-10-24 09:48:57	0

date of departure

identifier of cluster


```
map_f = tdf.plot_trajectory()  
cstdf.plot_stops(map_f=map_f)
```

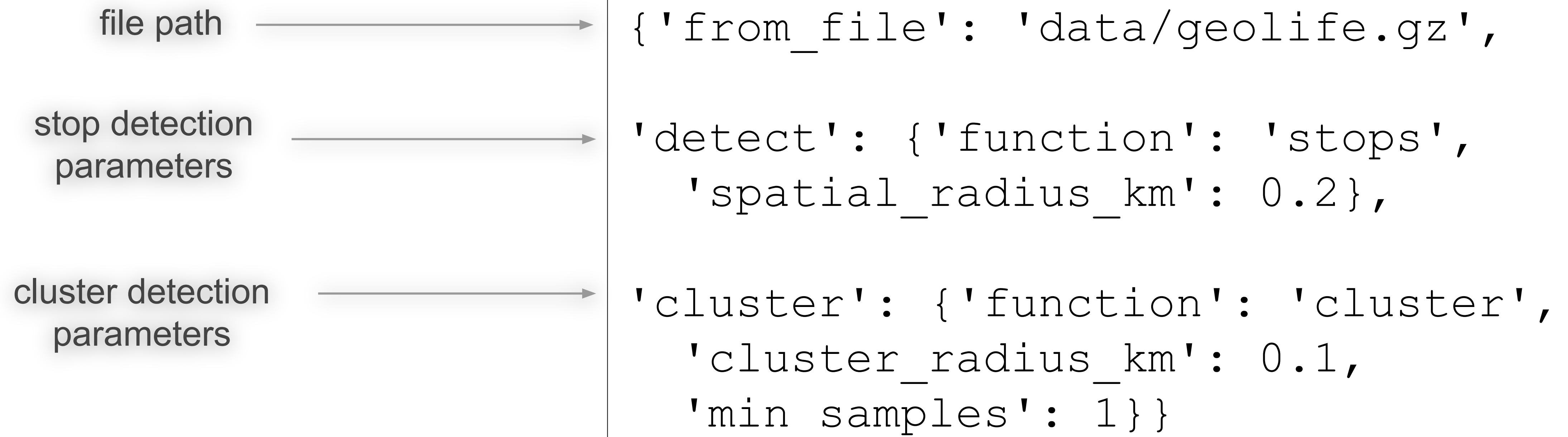
```
# plot trajectory  
# plot clusters
```

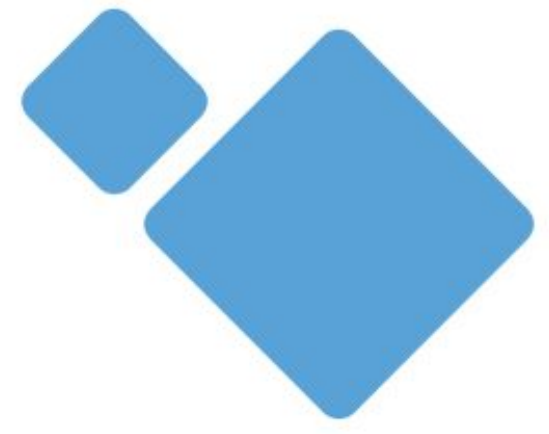


```
from skmob.preprocessing import detection, clustering

# detect stops and clusters
stdf = detection.stay_locations(tdf, spatial_r_km=0.2)
cstdf = clustering.cluster(stdf, cluster_radius_km=0.1)

cstdf.parameters
```





skmob.measures

Statistics and mobility measures

statistics regarding
groups of individuals



Collective Measures

`uncorrelated_location_entropy` (traj[, ...])

`mean_square_displacement` (traj[, days, ...])

`visits_per_location` (trajs)

`homes_per_location` (traj[, start_night, ...])

`visits_per_time_unit` (traj[, time_unit])

`origin_destination_matrix` (traj[, ...])

statistics regarding
single individuals



Individual Measures

<code>radius_of_gyration</code>	<code>(traj[, show_progress])</code>
<code>k_radius_of_gyration</code>	<code>(traj[, k, show_progress])</code>
<code>random_entropy</code>	<code>(traj[, show_progress])</code>
<code>uncorrelated_entropy</code>	<code>(traj[, normalize, ...])</code>
<code>real_entropy</code>	<code>(traj[, show_progress])</code>
<code>jump_lengths</code>	<code>(traj[, show_progress, merge])</code>
<code>maximum_distance</code>	<code>(traj[, show_progress])</code>
<code>distance_straight_line</code>	<code>(traj[, show_progress])</code>
<code>waiting_times</code>	<code>(traj[, show_progress, merge])</code>
<code>number_of_locations</code>	<code>(traj[, show_progress])</code>
<code>home_location</code>	<code>(traj[, start_night, ...])</code>
<code>max_distance_from_home</code>	<code>(traj[, start_night, ...])</code>
<code>number_of_visits</code>	<code>(traj[, show_progress])</code>
<code>location_frequency</code>	<code>(traj[, normalize, ...])</code>
<code>individual_mobility_network</code>	<code>(traj[, ...])</code>
<code>recency_rank</code>	<code>(traj[, show_progress])</code>
<code>frequency_rank</code>	<code>(traj[, show_progress])</code>

```
from skmob.measures.individual import radius_of_gyration,  
k_radius_of_gyration
```

```
# compute radius and k-radius
```

```
rg_df = radius_of_gyration(tdf)
```

```
krg_df = k_radius_of_gyration(tdf, k=3)
```

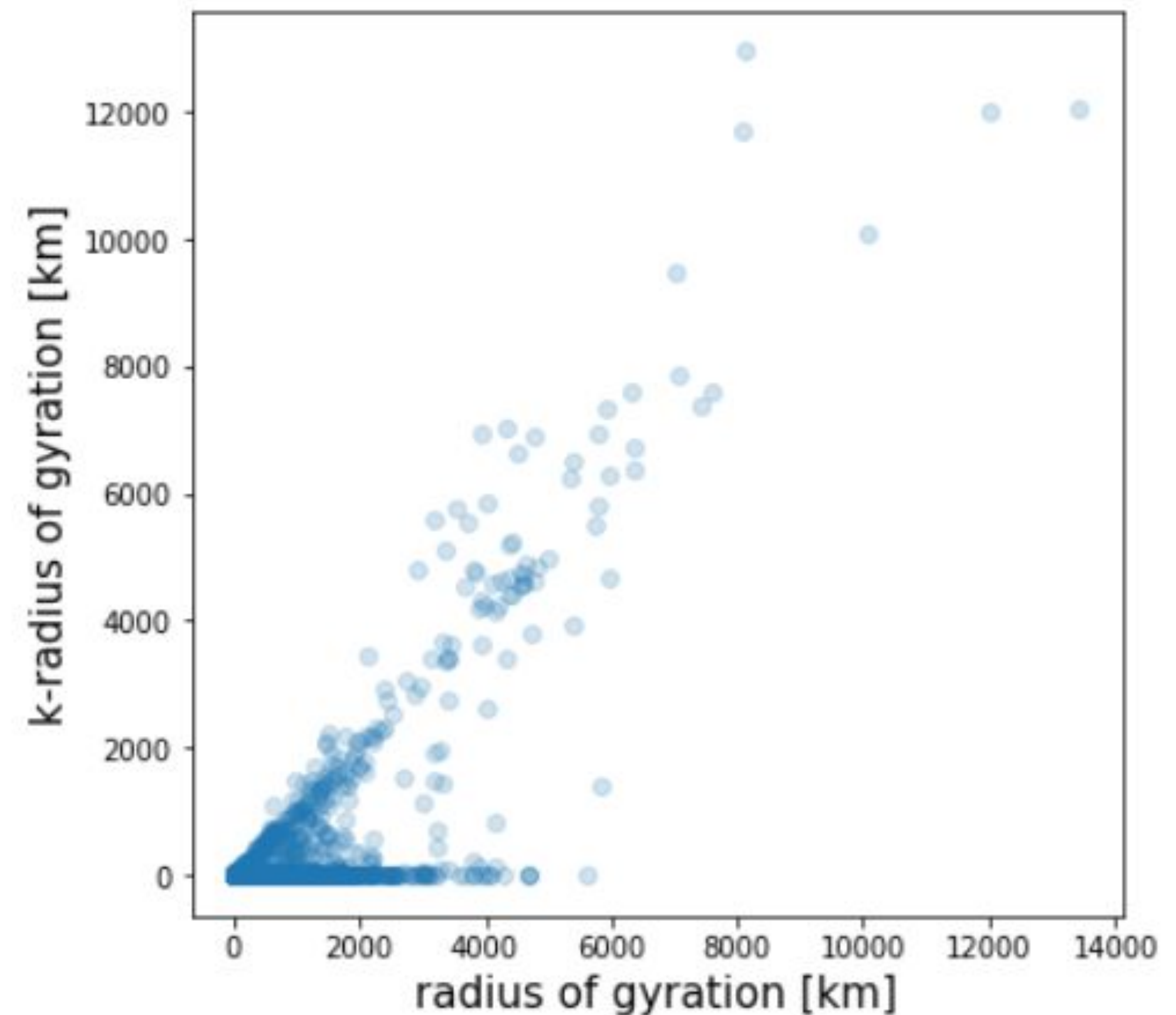
uid radius_of_gyration		
0	0	1564.436792
1	1	2467.773523
2	2	1439.649774
3	3	1752.604191
4	4	5380.503250

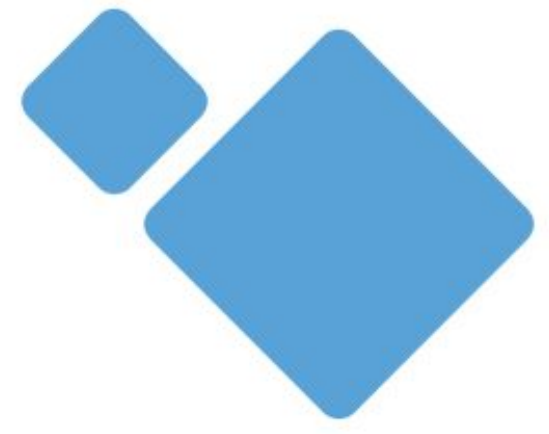
uid 3k_radius_of_gyration		
0	0	7.730516
1	1	3.620671
2	2	6.366549
3	3	10.543072
4	4	3910.808802

merge the two dataframes

```
rg_krg_df = pd.merge(rg_df, krg_df, on='uid')
```

	uid	radius_of_gyration	3k_radius_of_gyration
0	0	1564.436792	7.730516
1	1	2467.773523	3.620671
2	2	1439.649774	6.366549
3	3	1752.604191	10.543072
4	4	5380.503250	3910.808802





skmob.models

Generative models

Collective models

Gravity, Radiation

Individual models

EPRs

```
.fit()           # fit model's parameters from data  
  
.generate()    # generate synthetic trajs/fluxes
```

```
from skmob.models import Gravity

# load a spatial tessellation
tsl = GeoDataFrame.from_file("NY_counties.geojson")

# instantiate a singly-constrained gravity model
gravity = Gravity(gravity_type='singly constrained')

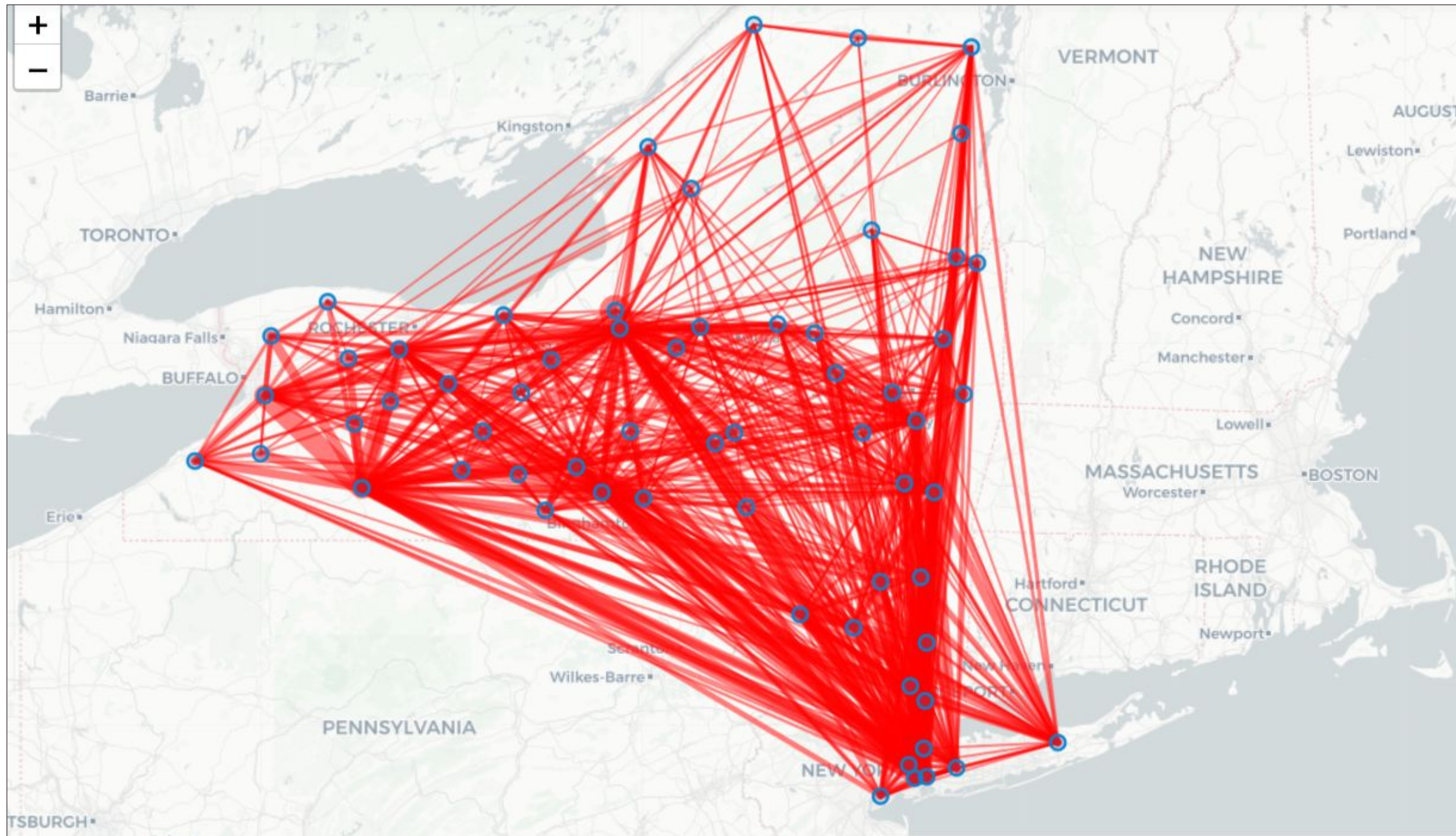
# load flow data and fit the model
fdf = FlowDataFrame.from_file("NY_flows.csv",
tessellation=tsl)

gravity.fit(fdf, relevance_column='population')
print(gravity)
```

```
Gravity(name="Gravity model", deterrence_func_type="power_law",
deterrence_func_args=[-1.763643891368663], origin_exp=1.0,
destination_exp=0.5906256813013072, gravity_type="singly constrained")
```



```
g_fdf = gravity.generate(tsl)  
g_fdf.plot_flows(min_flow=300) # plot the generated flows
```




```

from skmob.models.epr import Ditras
from skmob.models.markov_diary_generator import MarkovDiaryGenerator

# instantiate model
...
ditras = Ditras(...)

tessellation = gpd.GeoDataFrame.from_file("data/NY_counties_2011.geojson")
start_time = pd.to_datetime('2019/01/01 08:00:00')
end_time = pd.to_datetime('2019/02/01 08:00:00')

# generate the trajectory of one agent
tdf = ditras.generate(start_time, end_time, tessellation, n_agents=1)

tdf.head(3)

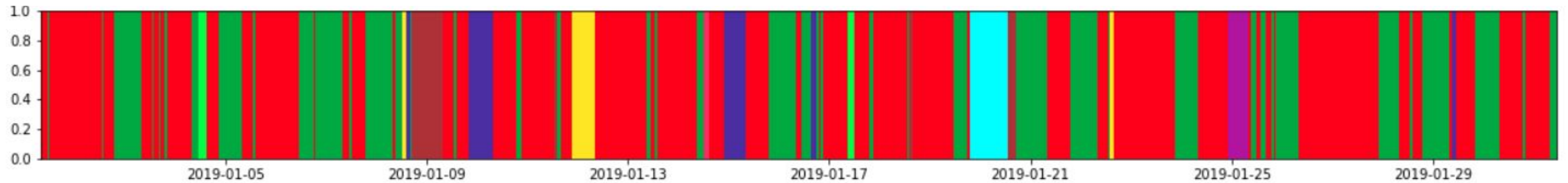
```

	uid	datetime	lat	lng
0	1	2019-01-01 08:00:00.000000	42.587412	-74.314766
1	1	2019-01-01 08:43:10.663092	42.587412	-74.314766
2	1	2019-01-01 09:05:19.433161	42.228184	-75.924827

→ TrajDataFrame

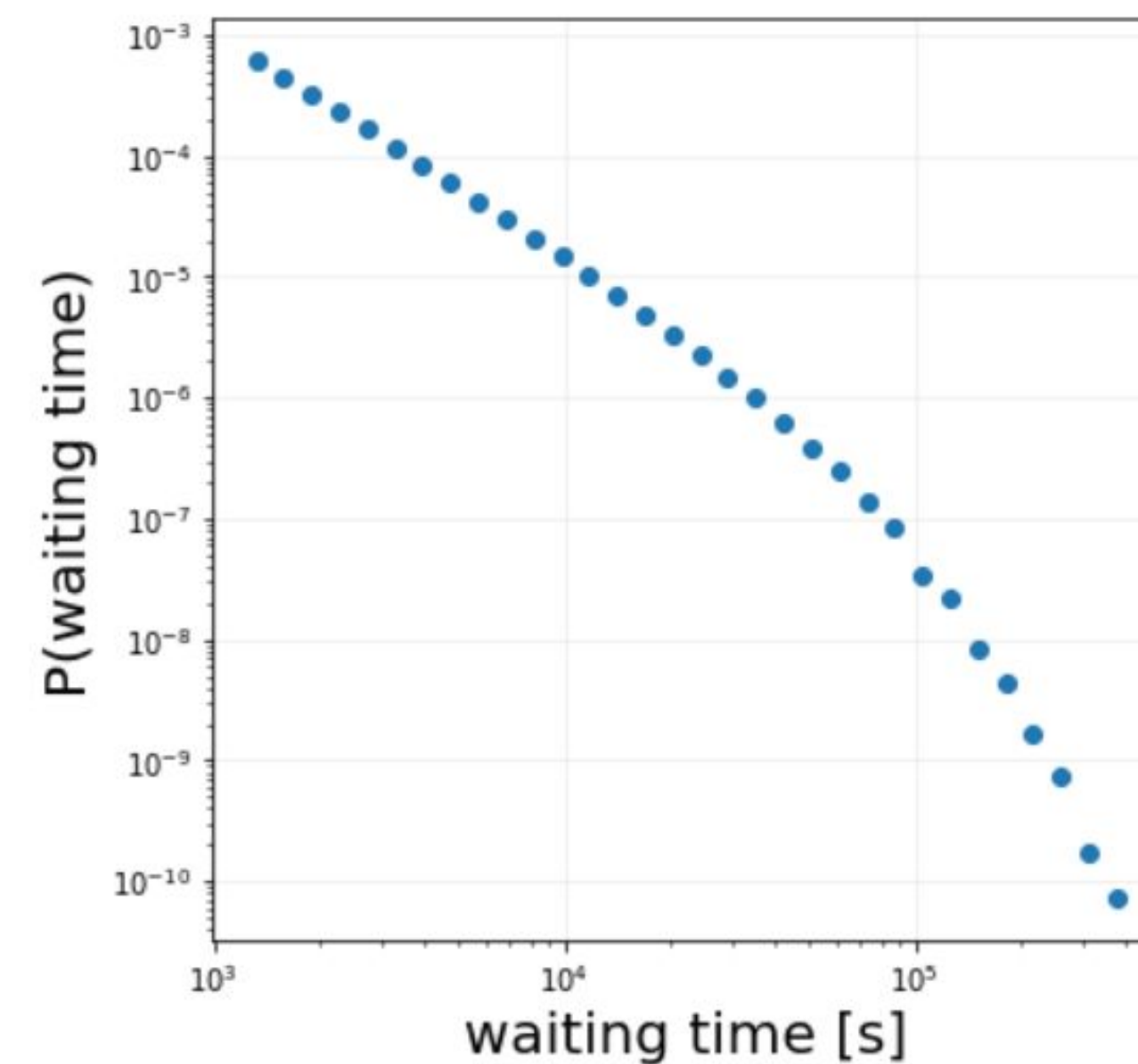
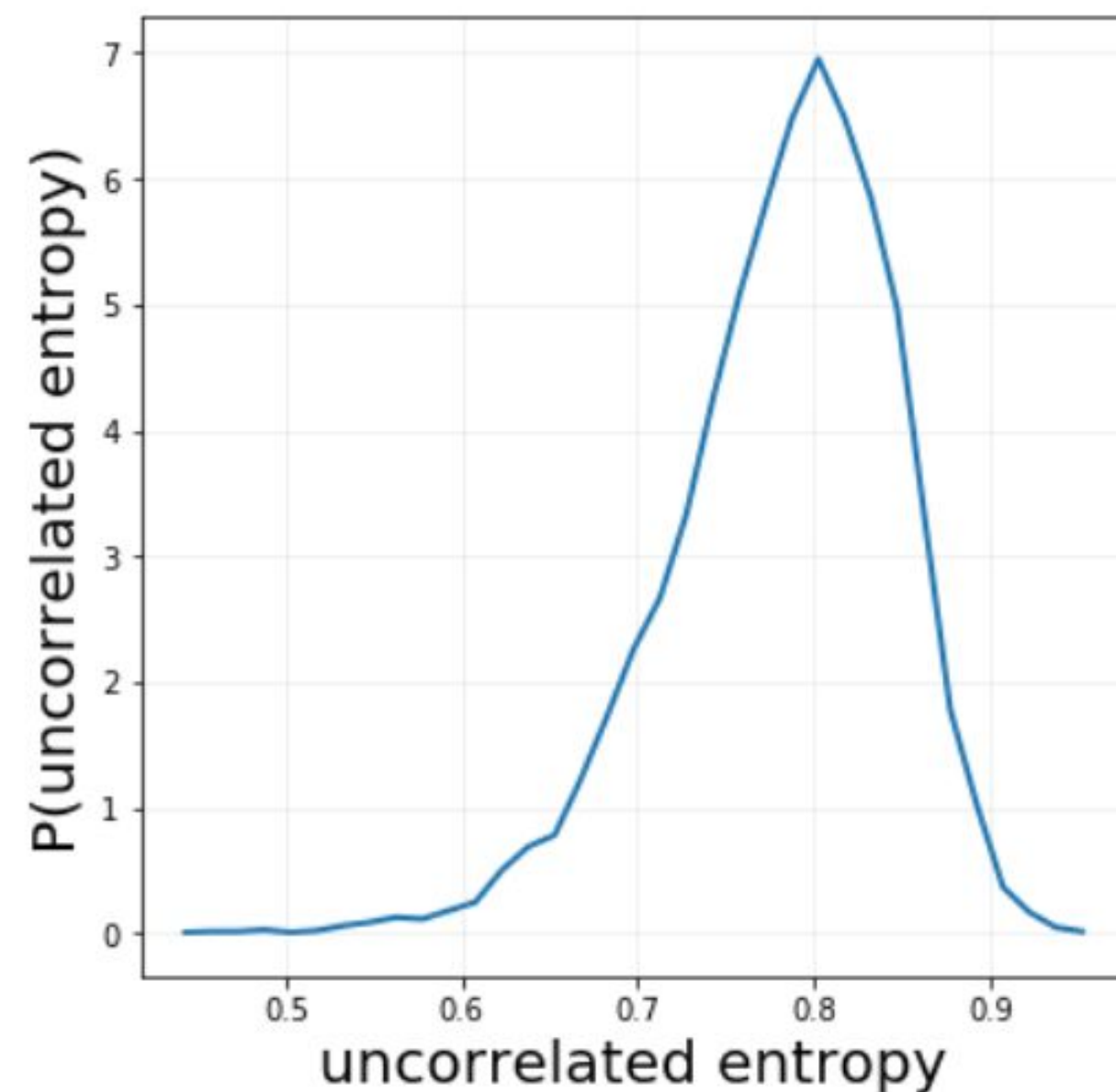
```
# plot diary
```

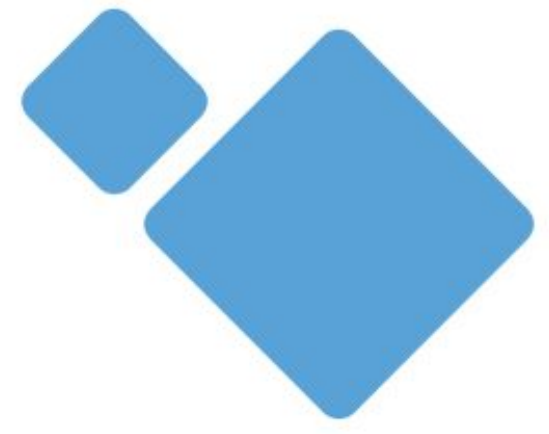
```
clustering.cluster(detection.stops(tdf)).plot_diary(user=1)
```



```
ue_df = uncorrelated_entropy(tdf, normalize=True) # compute unc entropy
```

```
wts = waiting_times(tdf, merge=True) # compute waiting times
```





skmob.**privacy**

Privacy risk assessment

A Privacy Perspective

Mobility info is sensitive

locations visited can be used to infer religion, health issues, habits, preferences




```
from skmob.privacy import attacks
```

```
filename = 'data/privacy_toy.csv'
```

```
tdf = TrajDataFrame.from_file(filename)
```

```
loc_att = attacks.LocationAttack() # instantiate attack
```

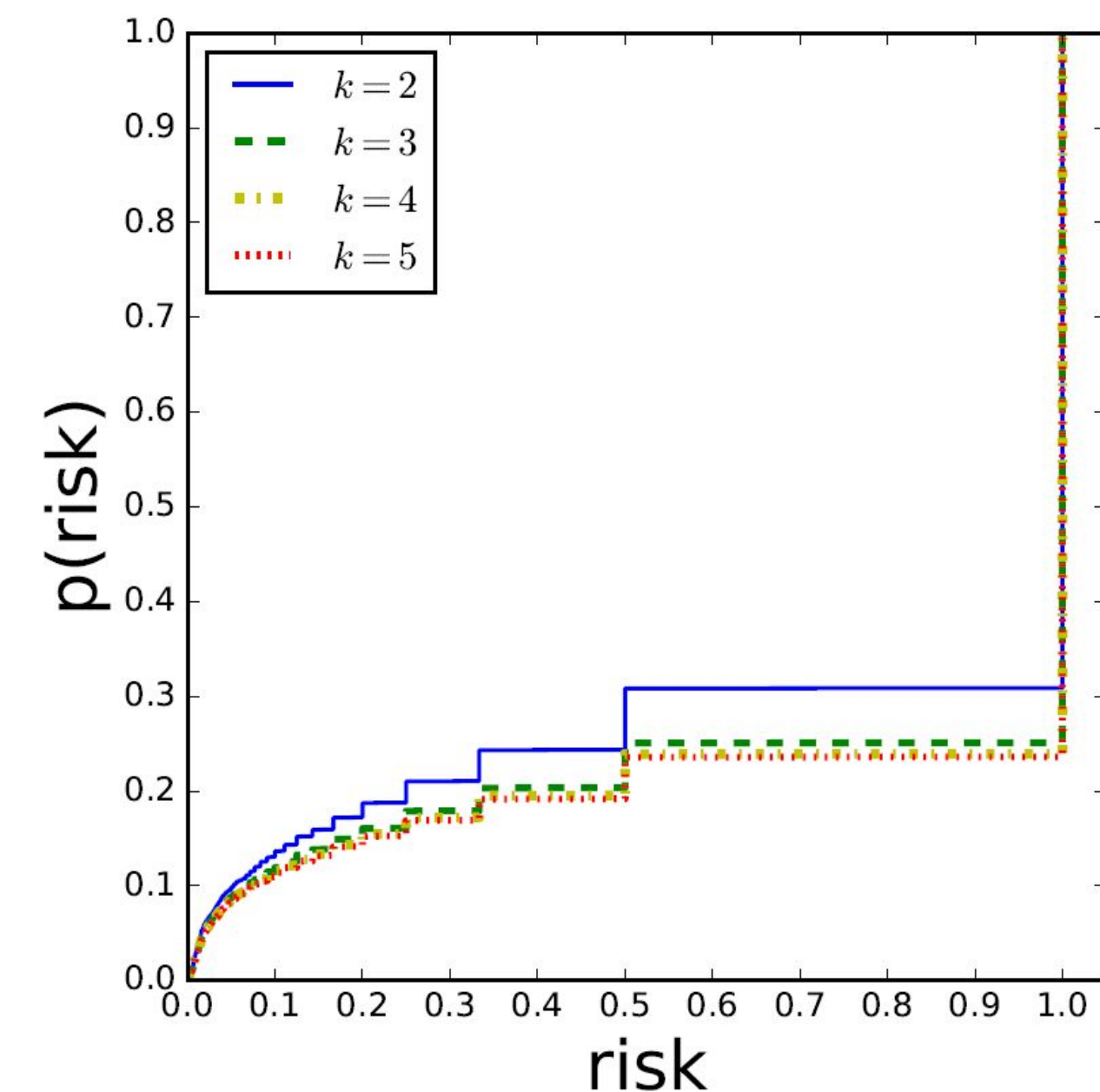
```
risk_df = loc_att.assess_risk(tdf) # assess privacy risk
```

```
risk_df.head(3)
```

user
identifier

re-identification
risk

	uid	risk
0	1	0.333333
1	2	0.500000
2	3	0.333333



New features

- ▶ More (and more efficient) preprocessing algorithms (e.g., home location detection, trajectory splitting, compression)
- ▶ New spatial tessellation functions
- ▶ More collective and individual models
- ▶ Segregation models and routing algorithms



[PyPI page](#)

[Home page](#)

Author: skmob Developers

License: new BSD

Summary: A toolbox for analyzing and processing mobility data.

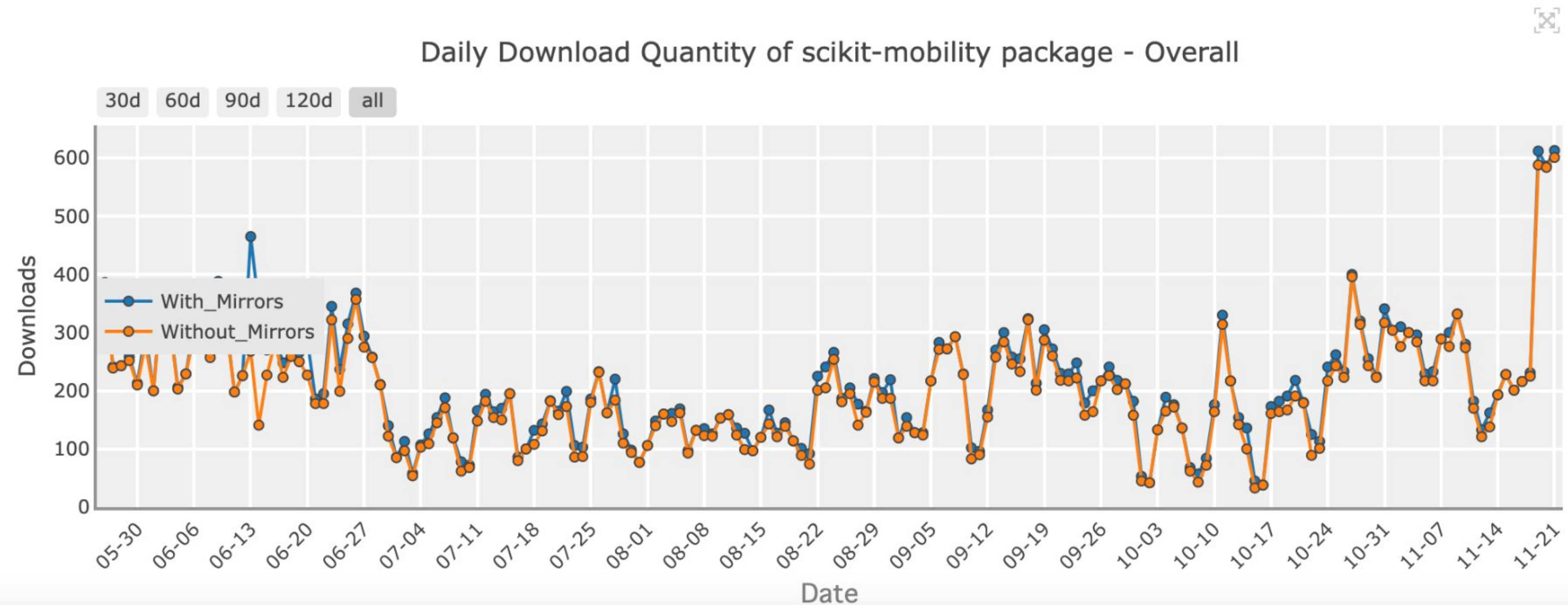
Latest version: 1.3.1

Required dependencies: [folium](#) | [geojson](#) | [geopandas](#) | [h3](#) | [pandas](#) | [pooch](#) | [powerlaw](#) | [python-igraph](#) | [requests](#) | [scikit-learn](#) | [statsmodels](#) | [tqdm](#)

Downloads last day: 601

Downloads last week: 2,643

Downloads last month: 8,311



Thanks - and stay tuned!

Impact on scientific communication		Users community		Project community	
97+	Journal papers	10,500+	VA users	135+	Researchers including PhDs
155+	Conference papers	94+	TA users	Available resources	
2	Book chapters	900+	Trainees	101	Unique big datasets
1	Monograph	5500+	Audience	123	Social mining methods and apps
120		Pilot projects developed with companies		1mln	Peak access to applications



www.sobigdata.eu

info@sobigdata.eu



RECIPROCITY



The RECIPROCITY project

Accelerating replication for smarter and cleaner mobility

Marcell Romsics

“

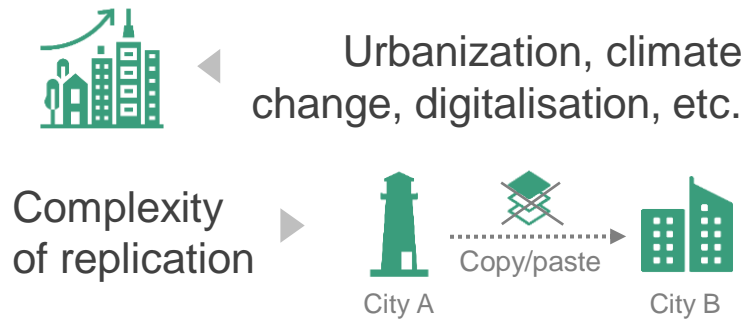


Our mission

RECIPROCITY aims at transforming European cities into climate-resilient and connected, multimodal nodes for smart and clean mobility through an innovative four-stage replication approach

Why? Responding to European mobility challenges

CHALLENGES

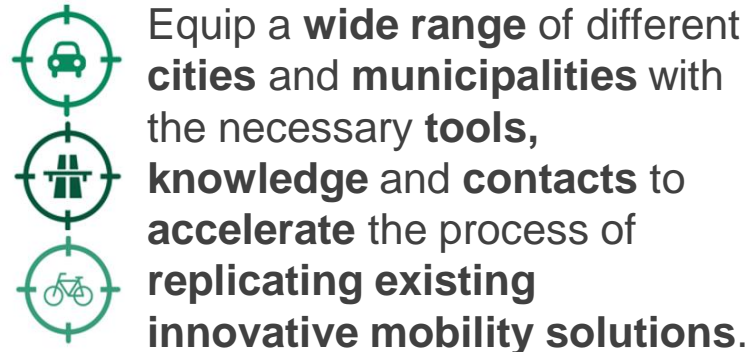


Support **Europe's efforts** to transform cities into climate-resilient, connected multimodal nodes for **smart and clean mobility** through **replication**.

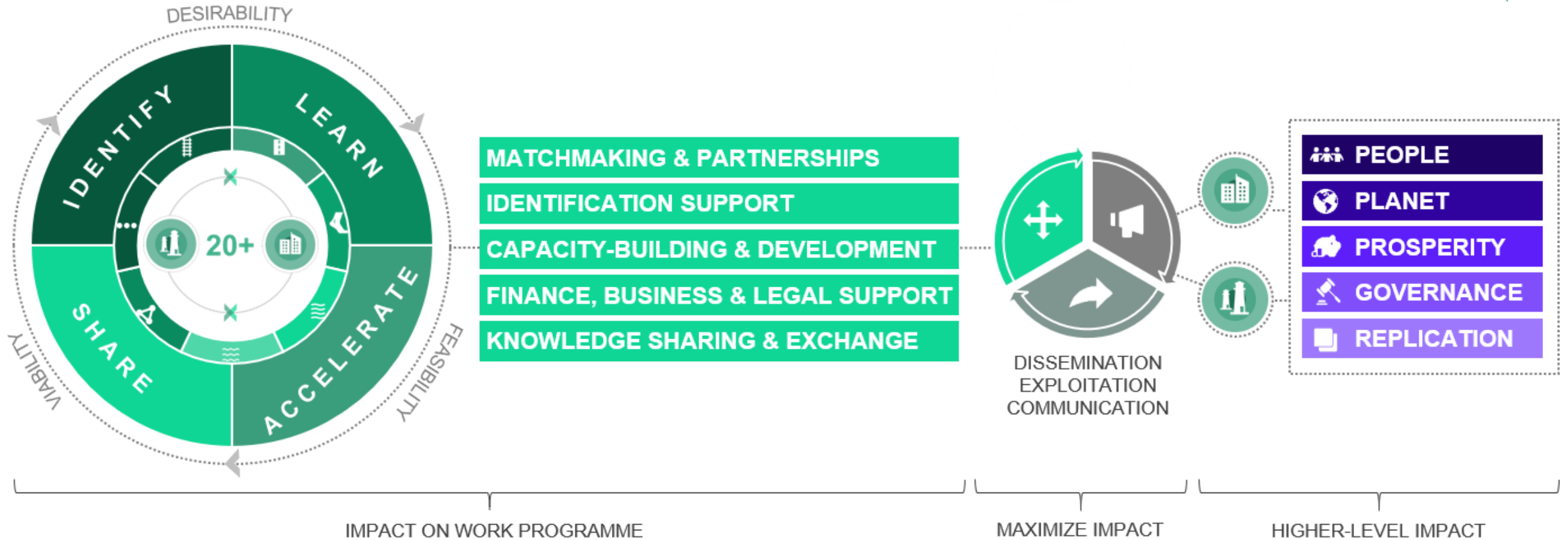
KEY QUESTION

How can the results of mobility pilot projects be used to implement smart city strategies and community initiatives more efficiently, faster and on a larger geographic scale?

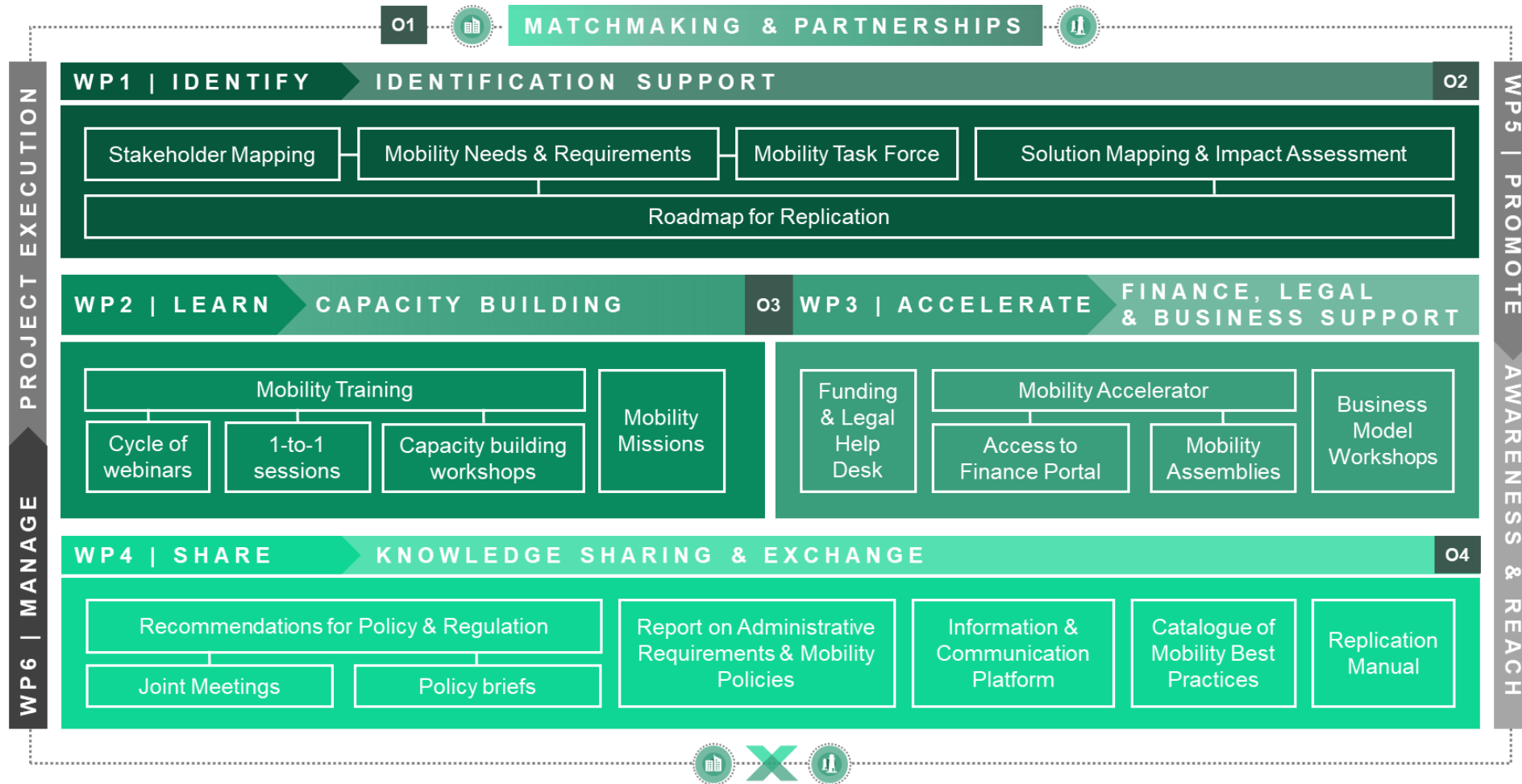
SOLUTION



How? The RECIPROCITY concept



How? Innovative four-stage replication framework



Who? Targeting stakeholders from the quadruple helix



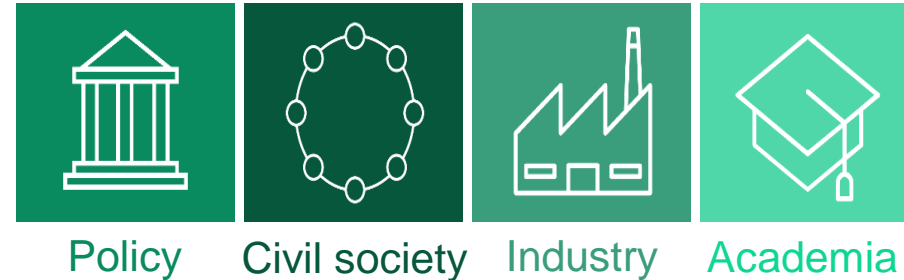
ECOSYSTEM

30+ cities & municipalities from Europe & H2020 countries:

- More than 25 Mio. people
- 16 European countries
- Coverage of 72.5 km²
- 10+ cities going through rapid social & economic change
- Rural, peri-urban & urban cities and areas
- Lighthouse & follower cities



STAKEHOLDERS



1000+ quadruple helix stakeholders



Project partners and overview



10 Partners



9 Countries



32 Months



1.5M€

RECIPROCITY

Greenovate!
EUROPE



Innovation
acceleration
expertise

Mobility
sector
knowledge

Experience with
urban mobility
planning



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006576

RECIPROCITY benefits



Better, smarter and cleaner mobility for all



Reduced greenhouse gas emissions and better air quality



Prosperity for all regions



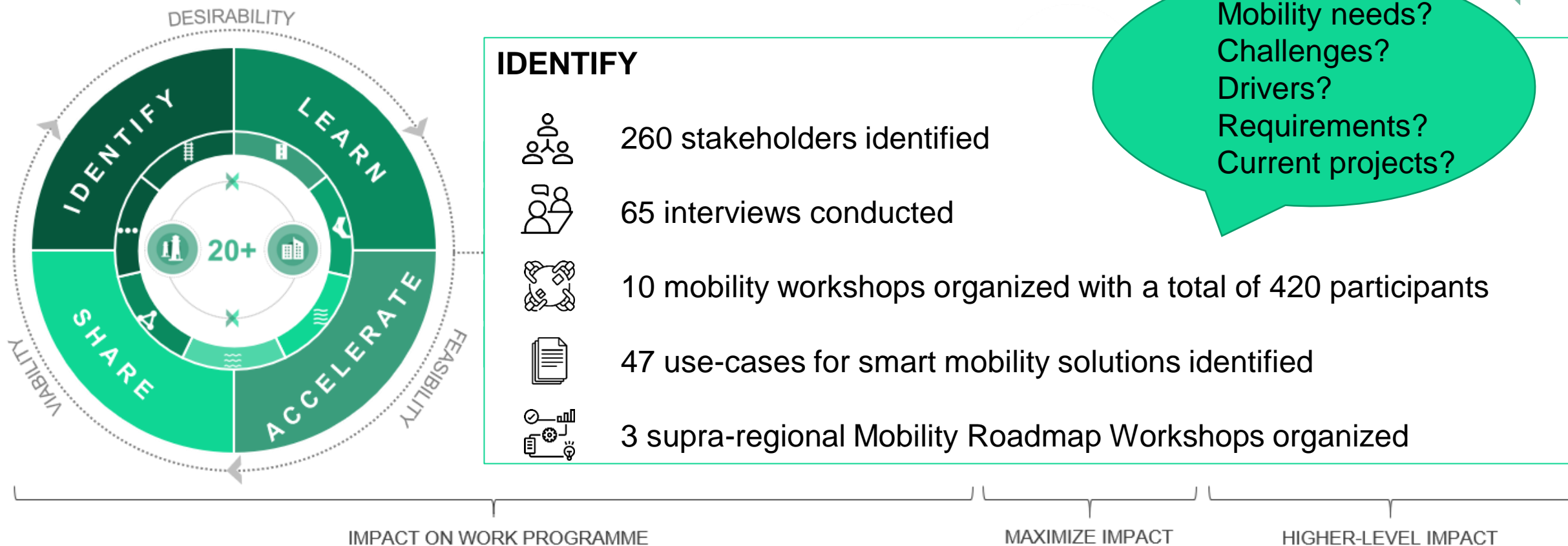
Improved local governance and transnational collaboration



Wide-scale replication despite different backgrounds



The RECIPROCITY concept



Single Traffic Model

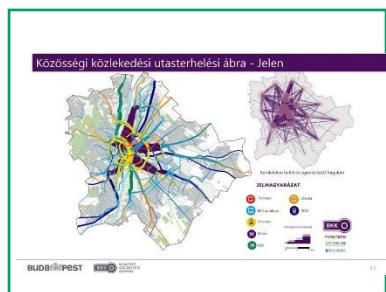
(Use Case #43)

Description / usage numbers

This is a model that includes public and continuously maintained data on total transport in the capital. These up-to-date traffic count information and statistics help the transport development projects taking place in Budapest and provide uniform, representative data for them. This Single Traffic Model is now a great help in preparing investments, during preliminary feasibility studies.

The Single Traffic Model is based on the all-transport model concept and provides a unified modeling basis for transport development planning: thanks to the software-independent design of the model, it can be accessed and used by any design company, continuously developed and improved through new developments related to various design tasks.

Pictures, diagrams etc. *(if available and useful)*



Source (picture): bkk.hu

RECIPROCITY



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006576

Responsible RECIPROCITY project partner: ZONE Cluster

External partners involved: Centre for Budapest Transport

City/village/region & country:

X	urban	X	peri-urban	X	rural
	tourism	X	everyday		commuter
technological readiness (TRL) [1-9]					9
simplicity to replicate [1=hard,... 5=easy]					3

Facts & figures *(if available and not confidential; add more if useful), e.g.*

- Costs for implementation & for operation (yearly):
- External expertise needed:
- Duration of implementation (from concept to finished project): 18 Months
- Website (only in hungarian):

<https://bkk.hu/downloads/747/9PA0CvHFqmgjdE8NJDAPVQ==>

document not confidential

BKK FUTÁR

(Use Case #42)

Description / usage numbers

The BKK FUTÁR system has two user groups: users of public transport and those who maintain it.

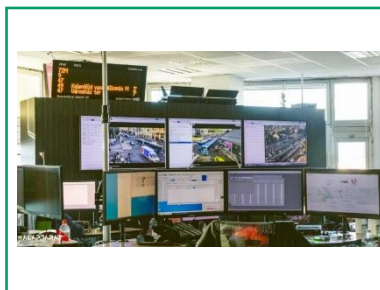
An interface available to users on the Internet, with virtually any online connection, consisting of a map and a trip planner. This allows public transit users to get up-to-date travel information.

In addition, the Centre for Budapest Transport, which maintains the public transport of Budapest, can monitor various vehicles and intervene in the operation of the system and network when necessary.

Pictures, diagrams etc. *(if available and useful)*



Source (picture): bkk.hu



Responsible RECIPROCITY project partner: ZONE Cluster

External partners involved: Centre for Budapest Transport

City/village/region & country:

X	urban	X	peri-urban		rural
X	tourism	X	everyday	X	commuter
technological readiness (TRL) [1-9]					9
simplicity to replicate [1=hard,... 5=easy]					3

Facts & figures *(if available and not confidential; add more if useful), e.g.*

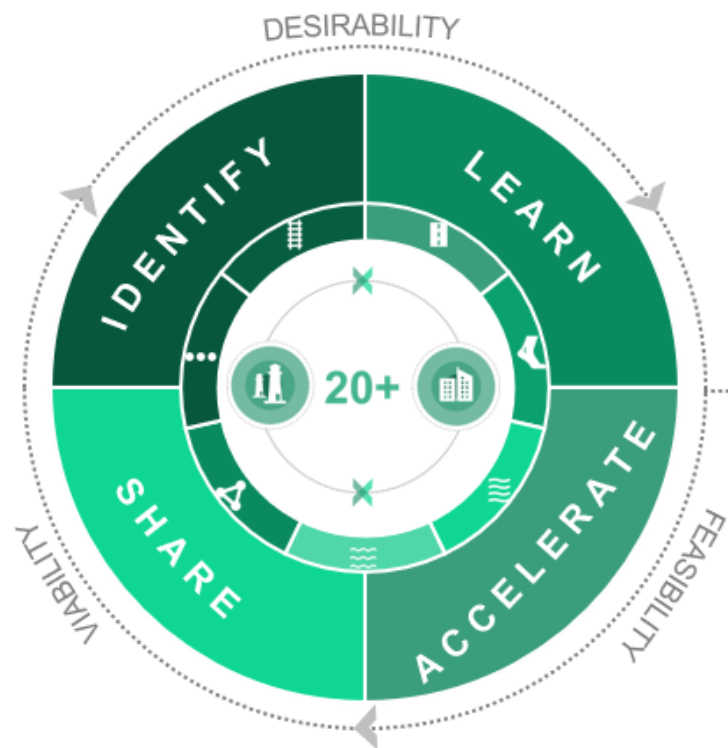
- Costs for implementation & for operation (yearly):
- External expertise needed:
- Duration of implementation (from concept to finished project): 12 Months
- Website: <https://futar.bkk.hu/>

“



What comes next?

The RECIPROCITY concept



LEARN

Two Mobility Missions - Travel Vouchers can be applied for

- [Linz, 28.-31.03.2023 \(co-located with Assembly\)](#)
- 24.-26.05.2023, Istanbul

Mobility Training Programm

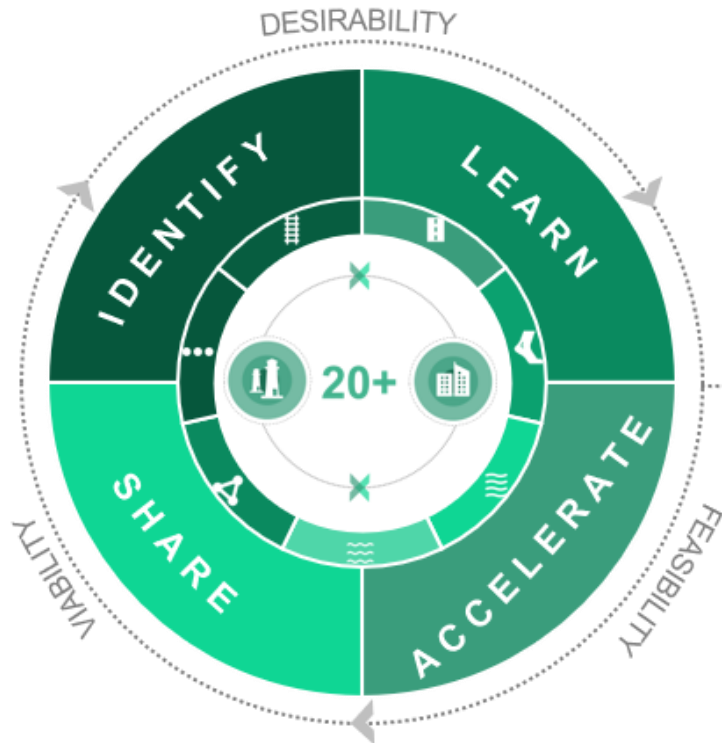
- Webinars
- Capacity Building Workshops
- 1:1-Sessions – next [matchmaking event on Horizon Europe Calls : 23/11, 28/11, 02/12 2022](#)

IMPACT ON WORK PROGRAMME

MAXIMIZE IMPACT

HIGHER-LEVEL IMPACT

The RECIPROCITY concept



ACCELERATE

Three Mobility Assemblies

- [30.11.-01.12.2022 Annual POLIS Conference, Brussels](#)
- [28.-31.03.2023, Futureforum Linz, March 2023 \(co-located with Mission\)](#)

Funding & Legal Helpdesk

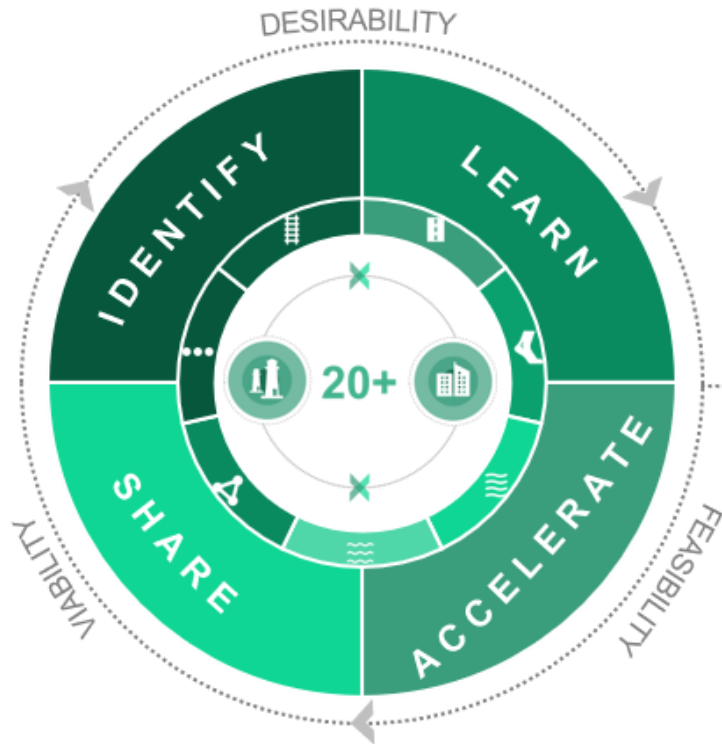
- One-stop shop for financial and legal questions – send us enquires

IMPACT ON WORK PROGRAMME

MAXIMIZE IMPACT

HIGHER-LEVEL IMPACT

The RECIPROCITY concept



SHARE

Online Information & Communication Platform

- Knowledge Center:
 - repository for all the recommendations, reports and outputs of the project (eg. Replication manual, generated business model patterns);
 - access to financial and legal expertise;
- Matchmaking tool.

IMPACT ON WORK PROGRAMME

MAXIMIZE IMPACT

HIGHER-LEVEL IMPACT

Thanks for your attention. Any questions?



Marcell Romsics

ZONE Cluster

marcell.romsics@zonecluster.eu

Connect with RECIPROCITY!



[RECIPROCITY_EU](#)



[RECIPROCITY project](#)



www.reciprocity-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101006576



Q&A

Methodologies and Data-based Technology Solutions for Improved European Mobility



These projects have received
funding from the European Union's
Horizon 2020 research and
innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.** **SoBigData**

RECIPROCITY

Interested? Connect!

Methodologies and Data-based Technology Solutions for Improved European Mobility



Thierry Chevallier

thierry.chevallier@akkodis.com



Ismini Stroumpou

ismini@sparsity-technologies.com



Luca Pappalardo

luca.pappalardo@isti.cnr.it



Marcell Romsics

marcell.romsics@zonecluster.eu



These projects have received
funding from the European Union's
Horizon 2020 research and
innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.**

SoBigData

RECIPROCITY

Interested? Connect!

Methodologies and Data-based Technology Solutions for Improved European Mobility



Diana Guardado

dianag@f6s.com



Danijel Pavlica

danijel@f6s.com



These projects have received funding from the European Union's Horizon 2020 research and innovation programme

organised by **MOBIDATALAB**
Labs for prototyping future mobility data sharing solutions in the cloud

supported by **synair.** **SoBigData**

RECIPROCITY

MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

The logo for synair, featuring the word "synair" in a blue sans-serif font. Above the "i" is a stylized orange and blue swoosh that ends in a small airplane icon.

The logo for SoBigData, with "So" in blue and "BigData" in green. Above "So" are four small vertical bars of increasing height, and below "BigData" are four small green plus signs.

The logo for RECIPROCITY, with the word in white capital letters inside a green rectangular box. A small green plus sign is located to the right of the box.

THANK YOU FOR JOINING!

mobidatalab.eu



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879)