MoTiV Project
University Science Park
Univerzitná 8215/1, 010 26 Žilina, Slovakia

D4.6 – Open Dataset with Mobility/Behavioural Data

Project	MoTiV - Mobility and Time Value
Project no.	770145
Deliverable no.	D4.6
Type of deliverable	R (document)
Work package	WP4
Due date of delivery	31.07.2020
Actual date of delivery	30.09.2020
Author(s)	Cristian Consonni (EURECAT), Ludovico Boratto (EURECAT)
Contributors	Ghadir Pourhashem (UNIZA)
Responsible Partner	EURECAT
Document Version / Status	V1.4
Date	30.09.2020
Reviewed by	Ghadir Pourhashem (UNIZA)
Approved by	
Dissemination level	PU (Public)
Project website	www.motivproject.eu
Project Manager	Ghadir Pourhashem | UNIZA | ghadir.pourhashem@uniza.sk

Horizon 2020
European Union Funding for Research and Innovation
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 770145

Public report with the publication of an open version of the dataset used in the research project, for wider use of the scientific community, policymakers, and other relevant stakeholders. The dataset will be accessible via the project website and relevant Open Data repositories

This deliverable is associated to the MoTiV Task 4.6 described below.

Description of Task 4.6 "Data Curation of the Mobility/Behavioural Dataset"
Task 4.6 Data Curation of the Mobility/Behavioural Dataset (EUT) Towards the end of the project, the Task Leader will prepare an Open Dataset with mobility/behavioural variables for reuse and preservation. An online interface to the dataset will be established to allow online queries or to download customised dataset (e.g. from a specific country).

Document Revision History

Version	Date	Change description	Author
1.0	13/09/2020	First draft	Cristian Consonni (EUT), Ludovico Boratto (EUT)
1.1	22/09/2020	Reviewed draft	Ghadir Pourhashem (UNIZA)
1.2	28/09/2020	1st revision	Ludovico Boratto (EUT)
1.3	28/09/2020	2nd revision	Cristian Consonni (EUT)
1.4	28/09/2020	Quality check and final version	Ghadir Pourhashem, Olena Hundarenko (UNIZA)

Disclaimer

Table of Contents

List of Tables

List of Figures

About MoTiV

The Horizon 2020 project MoTiV (Mobility and Time Value) addresses the emerging perspectives on changing Value of Travel Time (VTT). Accordingly, it explores the dynamics of individual preferences, behaviours and lifestyles that influence travel and mobility choices. In other words, what does value of travel time mean for the end users, in relation to their travel experience?

The MoTiV project addresses VTT from the perspective of a single individual with a unique combination of personality, preferences, needs and expectations, in contrast with the traditional viewpoint of the economic dimension (time and cost savings). Its approach aims at achieving a broader and more interdisciplinary conceptualisation and understanding of VTT emphasising its "behavioural" component.

The main goal of the MoTiV project is to contribute to advancing research on VTT by introducing a conceptual framework for the estimation of VTT at an individual level, based on the value proposition of mobility. The conceptual framework will be validated through data collection and evaluation in at least 8 EU countries. The mobility and behavioural dataset will be collected using a mobile application developed by the project consortium, which will combine and integrate in an innovative way features from a multi-modal "journey planner" and an "activity/mobility diary". With this mobile app, end-users will be able to more easily track, understand, and re-evaluate travel decisions to make the most of their free time in accordance with personal preferences, lifestyle, interests, and budget. The target is to engage in the data collection process a minimum of 4.000 participants actively using the MoTiV app for at least two weeks. Besides validating the conceptual framework, the dataset will be made available to the scientific community as an Open Dataset to stimulate further research in this area.

The MoTiV project findings will produce scientific and policy outcomes, as well as potential business developments, including the development of new mobility services and the extension of existing applications, such as the ones offered by the business partners of the Consortium (i.e. *routeRANK journey planner*[1] and the *PiggyBaggy*[2] app for crowdsourced deliveries).

Partners

[1] https://www.routerank.com

[2] http://piggybaggy.com

List of Abbreviations and Acronyms

ICT	Information and Communications Technology
IFA	Influence Factor Analysis
VOT	Value of Time
VTT	Value of Travel Time

MoTiV Consortium Partners and Acronyms

Acronym	**Full name**
UNIZA	Žilinská univerzita v Žiline
CoRe	CoReorient Oy
ECF	European Cyclists' Federation ASBL
EUT	Fundació Eurecat
INESC ID	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa
routeRANK	routeRANK Ltd
TIS	Consultores em Transportes Inovação e Sistemas S.A.

Executive summary

Mobility is a system involving several stakeholders. Therefore, it is relevant to characterise mobility behaviour and preferences in a detailed way, to enable nuanced decisions. Current paradigms rely mostly on time saving, proposing to users' solutions that include the shortest path. Even though the value of travel time can be extended beyond travel duration, no dataset to characterize mobility and value of travel time from different perspectives exists. This creates a gap between novel mobility paradigms and the characterisation of user mobility. To enable the mining of user mobility under these new paradigms, we present the MoTiV (Mobility and Time Value) dataset, which contains data about travellers and their journeys, collected from a mobile application, called Woorti. Each trip contains multi-faceted information: from the transport mode, through its evaluation, to the positive/negative experience factors. We also present a use case, which compares corresponding legs with different transport modes, studying experience factors that negatively impact users. We conclude by discussing other application domains and research opportunities enabled by the dataset.

Structure of the document

This deliverable is structured as follows: Section Introduction introduces the motivation behind the work contained in this deliverable and outlines our contribution. Section Related Datasets presents related datasets and outlines the differences with respect to the one proposed in this work. In Section Context, we contextualize our work with respect to the Woorti app. In Section MoTiV Dataset Description, we present the details of our dataset, and in Section Case Study: Experience Factors Impacting Negatively Use of Public Transport, Cycling and Walking versus Private Cars, we provide a concrete use case to extract knowledge on the factors that negatively impact the use of different means of transport. Section Research Opportunities provides the community additional scenarios where our dataset can be employed. Finally, in Section Conclusions and Future Work, we conclude our report and highlight future initiatives to make this dataset grow.

Introduction

The extraction of actionable knowledge from user mobility has been a central perspective mainly for mobility stakeholders who, in turn, used this knowledge to adapt the services to the end users. This process, known as behavioural-data mining, aims at extracting patterns from the users' behaviour, in order to get to know them (Manca, Boratto and Carta, 2018; Boratto, Carta, Kaltenbrunner, and Manca, 2018). Under this paradigm, end users have mostly been passive actors.

In the last 20 years, not only mobility has radically changed – we travel much more than we used to – but, with the advent of Web platforms such as journey planners (Sourlas and Nathanail, 2019), users have also become *active actors* in providing their mobility preferences. Users can sort trips based on different types of preferences (e.g., length, duration, emissions) and decide to complete a trip by combining services of different operators. Hence, novel mobility paradigms emerged, in which users' *value of travel time* is defined as the combination of different factors; e.g., a trip by bike is surely not the shortest among the various means of transport, but might be valuable from multiple perspectives, such as emissions, costs, and fitness for the users (Devarasetty et al., 2012).

Related work. Data coming from online platforms were previously used to mine user behaviour, to consider aspects such as their willingness to pay for services (Zografos, Androutsopoulos, and Apospori, 2012) and challenges in behaviour change (Schrammel et al., 2015).

Manca and collaborators (Manca et al, 2017) presented a survey on mobility patterns considering social media data. González (González et al., 2008), instead, consider mobile phone data, while Calabrese et al. (Calabrese et al., 2013) extract mobility patterns from urban sensing data. Goulias (Goulias, 2018) surveyed the existing travel behaviour models.

Data analysis can produce insights that serve as input for other purposes, such as the improvement of transport services by considering user needs (Sierpiński and Staniek, 2017), the promotion of changes of the user habits (Schrammel et al., 2015), and the improvement of journey planners and transport portals (Esztergár-Kiss, 2016; Vargas, Weffers, and da Rocha, 2011).

Other studies go beyond data analysis, e.g., to extract topic models from geo-location data (Hasan and Ukkusuri, 2014), to forecast the evolution of preferences over time thanks to a Markov model (Zarwi, Vij, and Walker, 2017), or to provide a personalised journey planning (Jakob et al., 2014).

As previously mentioned, the knowledge coming from the user analytics can also be used as a form of actionable knowledge, e.g., to improve transport service according to the user needs (Sierpiński and Staniek, 2017), to promote changes in the user habits (Schrammel et al., 2015) such as the adoption of greener and healthier solutions (Gabrielli et al., 2014), or improve the usability and services provided by journey planners and transport portals (Esztergár-Kiss, 2016; Vargas, Weffers, and da Rocha, 2011).

Open issues. While we previously highlighted that user mobility has introduced novel paradigms related to value of travel time, our analysis of the literature showed that existing studies capture user behaviour from a single perspective. When moving from academic research to industrial applications and platforms, a one-to-one relationship between value of travel time and shortest path exists. Indeed, when considering services such as Google Maps and Waze, trip options are ranked by shortest path. We believe that this might be due - at least from the research side - to the lack of knowledge on *what is actually valuable for the users in their mobility choices*. While several existing datasets capture mobility by considering a single perspective (e.g., one transport mode, or only trip coordinates), *no dataset, capturing both explicit preferences in terms of value for the users when making their mobility*

choices and implicit information coming from the trip (e.g., coordinates), exists. This, in turn, has reflections on the industrial applications, which do not capture explicit feedback on additional factors, possibly relevant for the user. Hence, value of travel time remains a concept trapped inside the mobility community, which cannot be fully exploited by the knowledge extraction and Web mining community, to create actionable knowledge also for Web platforms.

Our contribution. To overcome this issue, in this dataset we present *MoTiV (Mobility and Time Value).* To show how the extraction of knowledge on users' value of travel time can concretely impact transport stakeholders, we also present a use case that characterises what are the negative factors associated to a trip, when this is performed by means of public transport or by cycling, w.r.t. when the same trip is performed by car.

Related Datasets

In this section, we present datasets related to the one presented in this resource paper and conclude by highlighting the difference between our dataset and the existing ones. Readers should note that mobility is such a broad concept that we decided to focus on those which provide information about user behaviour and preferences in mobility.

Social and trip data. Microsoft GeoLife is a social network that allows users to share their experience, both through GPS data and with pictures, and connect with other users. The publicly available GPS trajectory dataset[3] was collected by 182 users, in a period of over two years. As the authors of this dataset highlight, no explicit info about the means of transport is available, and no evaluation of the trips is offered.

Check-in data. Check-in data, coming from platforms such as Twitter, Foursquare, or Gowalla, is usually used to consider user preferences related to their mobility. Indeed, knowing where the users go and with which frequency, allows to characterise users' mobility and their preferences (especially if check-ins can be paired with reviews). Examples of datasets belonging to this class can be found here.[4] While this class of datasets is widely employed in various personalisation algorithms, the concept of a trip is entirely missing, thus losing all the information about the means of transport, or the relevant factors for the users.

Trip-only data. Another class of datasets collect trip information. A big selection of 250 datasets available for research purposes can be found here.[5] Trip datasets usually do not associate the trips with a user, not allowing a characterisation of the individual user mobility. In addition to this, they are usually associated with a single transport of mode (e.g., taxis).

---

[3] https://www.microsoft.com/en-us/research/project/geolife-building-social-networks-using-human-location-history/#!downloads

[4] Foursquare dataset, https://sites.google.com/site/yangdingqi/home/foursquare-dataset

[5] Data World, https://data.world/datasets/trips

Contributions of the MoTiV dataset. Mobility datasets usually *focus on a narrow perspective* and usually the full user-mobility picture is lost. The goal of the MoTiV dataset, presented in this deliverable, is to capture multiple perspectives behind user mobility. This is done by:

- having trips with different transport modes (something not available in classic trip datasets);
- collecting trip coordinates (this is usually not available in check-in datasets);
- collecting reviews and evaluations of each trips;
- working at a lower granularity, by not only considering the trip as a whole, but also the activities that compose it. This allows to consider the different experience factors that characterise user mobility at different stages of the trip.

Context

In the context of the MoTiV, the value of travel time is analysed from a traveller's perspective, assuming that time and cost savings are not always the main criteria influencing route and mode choice. Depending on the traveller's transport attitude and context, other criteria such as environmental impact, comfort, or even weather conditions may influence the perceived value of a trip. In particular, MoTiV adopts the perspective that travel time can be worthwhile, i.e., it can be allocated for activities that the user finds useful, enjoyable, or productive.

MoTiV shifts perspective from considering travel time as spent – or, worse, wasted - to time that can be characterised by other activities. Furthermore, this characterisation is not limited by defining time as productive or unproductive time, because it is not necessarily related to its evaluation in terms of cost. Worthwhile time is independent of what can be monetised. The definition of worthwhile time encompasses multiple dimensions of travel time value from the perspective of the traveller.

MoTiV Dataset Description

The dataset presented in this deliverable has been collected through the *Woorti* app,[6] developed within the scope of the MoTiV project. The application supports both Android and iOS devices, and it is available in 11 languages. Data are inserted in the app directly by users that can register their trips at any time. Furthermore, data collection is also facilitated by dedicated data collection campaigns, coordinated by MoTiV campaign managers; these campaigns have targeted 8 European countries: Belgium, Finland, France, Italy, Norway, Portugal, Slovakia, Spain. Overall, the data considered in this report covers a period of 8 months, from May 1st, 2019 to 15 December2019 as described in the Deliverable 4.3 "Data Collection Campaign Report"

The use of the Woorti app consists of three main phases:

1. *Onboarding*: upon installing the application and registering a new account, the user is introduced to the functionalities of the app. During this process, the user enters their travel preferences as well as, optionally, some basic demographic information.
2. *Trip recording*: the user can start a new trip and the Woorti app automatically collects data in background.

[6] The name of the app is a play on the words *"worth it"* referring to worthwhile travel time.

3. *Trip validation*: when a trip is finished, the user can review the data, validate it, and insert other data regarding the trip (trip purpose, mood, etc). A 'validated' trip required that the user reviewed an automatically detected trip and confirmed (or corrected) the individual legs composing a trip as well as their mode (also automatically detected based on sensor patterns.

Within the application, a user can access their data through several screens and visualise and edit their profile information and trips. Furthermore, the application features a dashboard that presents to the user multiple statistics related to their validated trips, both at an individual level and by comparison with the Woorti community.

User Details

Trip data are contained in the table `user_details.csv`, whose fields are reported in Table 1.

User data collected during the onboarding phase.

Table 1: user details collected by the Woorti app. Only the fields in bold (user_id, registration_timestamp, gender, and age) are required.

Field	Description and admissible values
user_id	user identifier. a string of 29 characters
registration_timestamp	trip start date, formatted as '%Y-%m-%d %H:%M:%S.%f'[7]
gender	gender of the user based on three predefined choices: *Male, Female, and Other,*
age	user's age in a range. based on eight predefined age ranges: *16-19, 20-24, 25-29, 30-39, 40-49, 50-64, 65-74, 75+*
language	language used for the Woorti app expressed as a 3-letter ISO 639-2/B code.
city	user's city of residence.
country	user's country of residence expressed in ISO 3166-1 alpha-3 code.
education_level	user's education. Possible values are: *Basic (up to 10th grade), High School (12th grade),* and *University*
marital_status	user's marital status. Based on five predefined choices: *Single, Registered relationship, Married, Divorced, Widowed*
number_of_people	number of people in the household. It can be a number from 1 to 4 or 5+
labour_status	users's labour status. Based on five predefined categories: *Student, Employed full-time, Employed part-time, Pensioner, Unemployed*

[7] For further details, see Python's `strftime()` function documentation: https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

years_of_residence	years of residence in the household and possible values are: *Less than 1*, *1 to 5*, and *More than 5*

User's Generic Worthwhileness Elements

MoTiV characterises users' preferences and experiences along three dimensions of fitness, enjoyment, and productivity, defined as follows (in parenthesis, we report a description of each dimension, visualised by users during data collection):

- *Fitness* measures how much the user values the fact that when traveling they can exercise («When you walk, cycle, or even run on your travels, you are getting exercise and keeping in shape»);
- *Enjoyment* is related to how the travel can be used for fun or relaxing activities («Relaxing or having fun: taking time to listen to music, rest or meditate; engaging in social media; observing the surroundings»);
- *Productivity* captures how much the user values the possibility of using travel time to complete some tasks, either personal or work-related. («Using travel time to get things done, not only for work or study, but also personal things like managing home or family stuff»).

User preferences and experiences are encoded in two main sets of values, called worthwhileness[8] values:

- *generic worthwhileness values:* they are a triplet of values (F, E, P) for *fitness*, *enjoyment*, and *productivity*, respectively. They measure how much the user values these dimensions in general when traveling;
- *specific worthwhileness values:* they are triplets of values (F, E, P) that the user is asked to assign for each specific mode of transport chosen in the onboarding phase. The transport modes that the user selects during the onboarding phase are called *preferred transport modes*. Specific worthwhileness values are the measure of how much the user values *fitness*, *enjoyment*, and *productivity* when using that specific transport mode.

During the onboarding phase, the user was asked to provide both the generic and specific worthwhileness values on a scale from 1 to 100. When evaluating trips, the user is asked to provide an evaluation for each dimension of *fitness*, *enjoyment*, and *productivity* using a scale from low to high (`low`, `medium`, `high`). For consistency with the evaluation values, the onboarding values can be scaled to the same three classes: `low`, for values in [0-33]; `medium`, [34-66]; and `high`, [67-100].

The table `user_generic_worthwhileness_values.csv` contains the generic worthwhileness values, its columns are:

- `userid`, the user identifier;
- `fit`, the value for fitness [0-100];
- `prod`, the value for productivity [0-100];

[8] Although this diction of the word is less widespread than the more common variant *"worthiness,"* it is used throughout the project, so we keep it for consistency with the project itself.

- `enjoy`, the value for enjoyment [0-100];

User's Specific Worthwhileness Elements

The table `user_specific_worthwhileness_values.csv` contains the specific worthwhileness values, its columns are:

- `userid`, the user identifier;
- `motid`, the mode of transport identifier;
- `fit`, the value for fitness [0-100];
- `prod`, the value for productivity [0-100];
- `enjoy`, the value for enjoyment [0-100];

The mapping between mode of transport ids and text is contained in `mots.csv`.

Trips info

Trip data are contained in the table `trips.csv`, whose fields are reported in Table 2.

Table 2: Trip information.

Field	Description
tripid	trip identifier
userid	user identifier
start_date	trip start date, formatted as '%Y-%m-%d %H:%M:%S.%f'
end_date	trip end date, formatted as '%Y-%m-%d %H:%M:%S.%f'
average_speed	average speed during trip
max_speed	max speed during trip
distance	leg distance in meters
duration	leg duration in seconds
mood_rating	evaluation of trip mood on a scale from 1 to 5
did_you_have_to_arrive	did the user had to arrive at a predefined time?
how_often	how often does the user make this trip?
use_trip_more_for	use_trip_more_for
manual_start	has the trip recording been started manually?
manual_end	has the trip recording been ended manually?

validation_date	trip validation date, formatted as '%Y-%m-%d %H:%M:%S.%f'[9]
os	operating system
os_version	operating system version
model	Model

Legs info

Leg data are contained in the table `legs.csv`, whose fields are reported in Table 3.

Table 3: Leg information.

Field	Description
legid	leg identifier
class	leg type, either `Leg` or `WaitingEvent`
userid	user identifier
tripid	trip identifier
motid	mode of transport identifier
start_date	leg start date, formatted as "%Y-%m-%d %H:%M:%S.%f"[10]
end_date	leg end date, formatted as '%Y-%m-%d %H:%M:%S.%f'[11]
true_distance	leg distance in meters
leg_distance	leg distance in meters
leg_duration	leg duration in seconds
worthwhileness_rating	worthwhileness rating, -1 if not set
transport_category	transport category
campaign	data collection campaign
weekday	day of the week when the leg was performed
weekday_class	classification of the day, either `Working_day` or `Weekend`

[9] For further details, see Python's `strftime()` function documentation: https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

[10] For further details, see Python's `strftime()` function documentation: https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

[11] For further details, see Python's `strftime()` function documentation: https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

Leg coordinates are contained in the table `legs_coordinates.csv`, whose fields are reported in Table 4. Coordinates are anonymised depending on the fact that the point is located in an urban, suburban or rural area. To classify the points, we used the "functional urban areas by country" classification provided by the Organisation for Economic Co-operation and Development (OECD).[12]

Table 4: Leg coordinates and point classification.

Field	Description
legid	leg identifier
start_lat	latitude of the starting point of the leg
start_lon	longitude of the starting point of the leg
end_lat	latitude of the ending point of the leg
end_lon	longitude of the ending point of the leg
start_name	name of the city - if available - where the starting point of the leg is located
start_country	name of the country - if available - where the starting point of the leg is located
start_class	classification of the starting point of the leg: urban, sub-urban, rural
end_name	name of the city - if available - where the starting point of the leg is located
end_country	name of the country - if available - where the starting point of the leg is located
end_class	classification of the starting point of the leg: urban, sub-urban, rural

Points where anonymised using the following criterion, applied both latitude and longitude (Table 5):

- for *urban* areas, points are rounded to the third decimal place;
- for *sub-urban* areas, the third decimal place is rounded to the nearest 0.5
- for *rural* areas, the second decimal place is rounded to the nearest 0.5

Table 5: Example of coordinate anonymisation.

Coordinate	urban	sub-urban	rural
0.173257	0.173	0.175	0.150
0.177131	0.177	0.175	0.200
0.178131	0.178	0.180	0.200
0.174323	0.174	0.175	0.150

[12] Functional urban areas by country, https://www.oecd.org/regional/regional-statistics/functional-urban-areas.htm

Activities

The table `activities.csv` contains information about activities performed during trips, its columns are:

- `legid`, leg identifier
- `activity`, activity name, possible values are *Accompanying*, *Browsing*, *Cycling*, *Driving*, *Eating*, *Listening*, *PersonalCare*, *ReadingDevice*, *ReadingPaper*, *Relaxing*, *Talking*, *Thinking*, *Walking*, *Watching*, and *Other*.

Experience Factors

The table `experience_factors.csv` contains information about experience factors affecting trips, its columns are:

- `legid`, leg identifier;
- `factor`, experience factor name;
- `type`, factor categorisation;
- `minus`, a boolean value, the factor was rated as negative;
- `plus`, a boolean value, the factor was rated as positive.

Note that there are both cases where a factor appears, but it received neither plus nor minus, or others, when it received both. In the analysis, we decided to eliminate those. The rationale was that the factors with neither choice were present, because they were selected and then deselected, while for the case of both plus and minus are because the interface allowed that, but only a minority of the users realised that and it was not very clear from the interface.

Trip purposes

The table `purposes.csv` contains information about trips purposes, its columns are:

- `tripid`, the trip identifier;
- `purpose`, the trip purpose, possible values are *Business_Trip, Everyday_Shopping, Home, Leisure_Hobby, Personal_Tasks_Errands, Pick_Up_Drop_Off, Work,* and *Other*.

Weather data

Weather data have been collected through the API provided by the OpenWeatherMap service.[13] The API was queried regularly to obtain weather data for a set of 66 cities of interest in the scope of the project. Weather information was collected for the times of 09:00, 12:00 and 18:00 for each day from July 8th, 2019 to December 18th, 2019. The dataset contains two tables related to weather data: `weather_raw.csv` and `weather_legs.csv`.

The table `weather_raw.csv` contains the data parse obtained from the OpenWeather API, the document of each field is available on the OpenWeatherMap website.[14] The table

[13] https://openweathermap.org/, OpenWeatherMap is an online service owned by OpenWeather Ltd

[14] OpenWeatherMap historical weather API guide, https://openweathermap.org/history

`weather_legs.csv` contains the association between trips legs and the corresponding weather for the time and place, the available fields are presented in Table 6. For further details about temperature, cloud, precipitation and wind categorisation, we refer to Deliverable D5.2 (Impact Factor Analysis)

Table 6: Weather information.

Field	Description
weatherid	weather identifier
Legid	leg identifier
request_date	weather request timestamp, formatted as '%Y-%m-%d %H:%M:%S.%f'[15]
centroid_x	longitude of the centroid of the leg
centroid_y	latitude of the centroid of the leg
weather_scenario	classification of weather scenarios
apparent_temperature	apparent temperature
net_radiation	net radiation received by the terrain in the location of the centroid
temperature_category	categorisation based on apparent temperature
temperature_description	description of the temperature based on apparent temperature
cloud_category	categorisation based on cloud cover
cloud_main	description based on cloud cover
precipitation_category	categorisation based on precipitation level
precipitation_main	description based on precipitation level
wind_beaufort_number	categorisation using Beaufort's scale[16], based on wind speed
wind_category	categorisation based on wind speed
wind_description	description based on wind speed

For the computation of weather scenarios, information was combined about cloud coverage, precipitation, wind speed and apparent temperature. In particular, the apparent temperature was computed, which is equivalent of the temperature perceived by humans, based on the effects of air temperature, relative humidity and wind speed. The formula for apparent temperature introduced by Robert Steadman in 1984 (Steadman, 1964) was used, which takes into consideration four

[15] For further details, see Python's `strftime()` function documentation: https://docs.python.org/3/library/datetime.html#strftime-and-strptime-format-codes

[16] https://en.wikipedia.org/wiki/Beaufort_scale

environmental factors: wind, temperature, humidity and radiation from the sun. The formula for the apparent temperature (AT) is:

$$AT = T_a + 0.348 \cdot e - 0.70 \cdot ws + 0.70 \frac{Q}{ws + 10} - 4.25$$

where: T_a is the dry bulb temperature (in $°C$); e is the water vapor pressure (in hPa); ws is the wind speed at an elevation of 10 meters (in m/s); Q is the net radiation absorbed per unit area of body surface (in W/m^2). The water vapor pressure (e) is computed as:

$$e = \frac{rh}{100} \cdot 6.105 \cdot e^{\frac{17.27 - T_a}{237.7 + T_a}}$$

While temperature (T_a), relative humidity (rh), and wind speed data (ws) were available from the OpenWeatherMap API, to obtain net radiation data (Q) another data source was used. Net radiation data were downloaded from the NASA Earth Observatory website,[17] which publishes daily maps of the net radiation absorbed by Earth.

A User's Story

In the following section, we present a brief user story of a user applying the *Woorti* app to describe the data collected in context. We will call our user Luigi.

One day, while Luigi is having breakfast and watching the television, he sees an advertisement of a new App just released. It is called Woorti and is used to keep track of personal trips along with all the aspects that influenced the travel. Luigi decides to download the app. He registers to the app and inserts some demographic information (**user_details**). After that, he is asked how much he values three aspects of life while he travels: *enjoyment*, how much fun he has; *productivity*, if he is able to work or do personal tasks; and *fitness*, if he is able to exercise and stay healthy; both in general terms (**user_generic_worthwhileness_values**) and with respect to his favourite transport modes (**user_specific_worthwhileness_values**).

Luigi wants to try the app while going to work (**purposes**). He starts from his apartment, walks until the bus stop where he waits for 5 minutes the bus. He gets on the bus for 20 minutes, and in the meantime he reads a book and uses the smartphone to check something on the Internet (**activities**). He easily finds a seat and enjoys his time on the bus, since it is not crowded, and he likes to watch the landscape outside (**experience_factors**). The bus stop is just a few minutes from his office, and since it's a beautiful and sunny day (**weather**), he decides to walk until there. Before starting his working day, he stops at the bar to take a coffee and spends some minutes filling in the questionnaire on the app after his trip. He validates the trip checking that the start and end points were correct (**legs coordinates**): he did in total four legs during his trip (walking, waiting event, bus, walking) (**legs**). He decides to review his leg on the bus, and he completes all the questions giving also a rate for the whole trip and for the leg he reviewed (**trips**), including a score for the enjoyment, productivity and fitness (**worthwhileness_elements_from_trips**).

---

[17] NASA-NEO, https://neo.sci.gsfc.nasa.gov/. These data are available in several machine-readable formats, such as GeoTIFF or CSV. Daily net radiation data GeoTIFF are available for bulk download at: https://neo.sci.gsfc.nasa.gov/archive/geotiff.float/CERES_NETFLUX_D/.

Case Study: Experience Factors Impacting Negatively Use of Public Transport, Cycling and Walking versus Private Cars

In this section, we present a case study to showcase the MoTiV dataset. We leverage the experience factor evaluations to compare trips done by cars versus trips conducted with other modes of transport: public transport, cycling and other emerging micromobility, and walking. With this use case we want to answer the question: *"What are the negative experience factors of cyclists and users of public transport for the same trip legs performed by car?"*

We can answer this question step-by-step:

1. we select trip legs performed by car and the users that have performed at least on of such trips;
2. we restrict this set to the users that have chosen at least one preferred transport mode within the transport categories: *cycling and emerging micromobility*, *public transport (short distance)*, and *public transport (long distance)*. For the sake of the study, we call these the alternative transport category;
3. we identify users that have performed similar trips using different transport modes, we select users that have performed the same journey using a car - in the *private motorised* category - or using a bike or public transport;
4. For the same users, we look at negative experience factors for trip legs performed using modes of transports in the alternative categories.

The rationale of this process is to find users that have travelled a given route multiple times using both cars and alternative modes of transport and look at which are the factors that have impacted negatively the travel experience of users when using bikes or public transport. In this way, we want to get some insights on which are the experience factors that are hindering the use of modes of transports alternative to cars.

For the definition of similar trips we match trips using their starting and ending points: When using different modes of transport, it can lead to taking different paths, but for our analysis, we are interested in the fact that a user needs to travel from a given pair of locations. To identify points that are close in space we adopt the following procedure, illustrated in Figure1.: each point is transformed to a curved square, where each side is an arc of length 0.004 degrees. If the squares representing two points intersect, we consider them a matching pair. To simplify the computational complexity of the matching process, we only compare trip legs in the same country. To estimate the distance between two points, we use an approximate conversion between the precision of decimal degrees in the EPSG:7030/WGS 84,[18] 0.001 degrees correspond 78.71 m, so two points are matching if they are within a maximum distance of 445 m. In fact:

[18] http://epsg.io/7030-ellipsoid, taken as the E/W at 45 degrees N/S.

$$D = \sqrt{2} \cdot 0.004\, deg \cdot 78.71\, \frac{km}{deg} \approx 445\, m$$

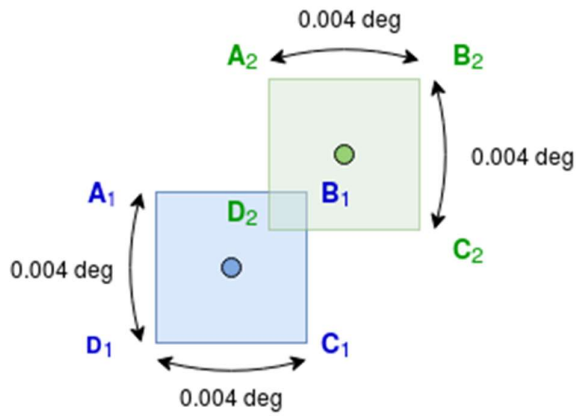

Figure 1: Procedure for identifying points that are close in space: each point is transformed to a curved square, where each side is an arc - with sides of length 0.004 degrees. If the squares representing two points intersect, we consider them a matching pair.

Table 7 presents some statistics related to each step of the process as presented above. For brevity, we will refer to trip legs performed with a mode of transport in the category *private motorised* as legs performed "by car" and trips in the categories *cycling and emerging micromobility*, *public transport (short distance)*, *public transport (long distance)* as legs performed "by alternative modes of transport."

Table 7: Statistics about the user considered by our case study.

Step	Step description	Trips	Legs	Users
1	All trips, legs, users	64,098	158,897	3,269
2	Users that have performed at least one trip leg by car	21,764	62,227	2,083
3	Users that have selected at least one preferred alternative transport mode	51,973	132,259	1,771
4	Users that have performed at least one trip leg by car and by alternative transport modes	20,032	27,921	1,376

Table 8 and Table 9 present the results of our case study: Table 8 contains the overall top-10 negative experience factors for all trip legs in the categories *cycling and emerging micro-mobility*, *public transport (short distance)*, and *public transport (short distance)*. Table 9 presents the top negative experience factors only for the trip legs selected by our case study for the same transport categories.

Table 8: overall top-10 negative experience factors for all trip legs the categories cycling and emerging micromobility, public transport (short distance), and public transport (short distance).

Cycling and emerging micromobility		Public transport (short distance)		Public transport (long distance)	
Factor	**#**	**Factor**	**#**	**Factor**	**#**
Cars Other Vehicles	1,751	Privacy	500	Internet Connectivity	53
Air Quality	1,277	Crowdedness Seating	499	Privacy	49
Road Path Availability And Safety	1,219	Other People	463	Seating Quality Personal Space	44
Noise Level	1,097	Seating Quality Personal Space	445	Reliability Of Travel Time	43
Road Path Quality	969	Noise Level	437	Noise Level	36
Traffic Signals Crossings	854	Internet Connectivity	418	Other People	35
Crowding Congestion	837	Charging Opportunity	406	Vehicle Ride Smoothness	34
Today's Weather	591	Air Quality	349	Today's Weather	34
Simplicity Difficulty Of The Route	417	Scenery	309	Crowdedness Seating	33
Facilities Shower Lockers	371	Reliability Of Travel Time	290	Food Drink Available	31

Table 9: top negative experience factors for the trip legs selected by our case study.

Cycling and emerging micromobility		Public transport (short distance)		Public transport (long distance)	
Factor	**#**	**Factor**	**#**	**Factor**	**#**
Road Path Availability And Safety	315	Privacy	1,33	Today's Weather	2
Road Path Quality	307	Other People	1,11	Reliability Of Travel Time	2
Cars Other Vehicles	267	Air Quality	1,03	Vehicle Ride Smoothness	1
Air Quality	173	Noise Level	1,00	Ability to Do What I Wanted	1

Noise Level	132	Crowdedness Seating	99	Privacy	1
Road Path Directness	131	Seating Quality Personal Space	99	Other People	1
Today's Weather	123	Today's Weather	84	Cleanliness	1
Traffic Signals Crossings	122	Internet Connectivity	77		
Crowding Congestion	111	Charging Opportunity	76		
Lighting Visibility	88	Scenery	67		

As we see by comparing the results between the two tables, when cycling, we find two main areas of concern: safety (availability of bicycle paths, safety from other cars, visibility and traffic signals), and quality (noise level, air quality). Road path directness has a somewhat more important role when cycling is used as an alternative to traveling by car w.r.t. the general negative experience factors. Weather is ranked among the top-10 negative experience factors in both bases. For short-distance, public transport main obstacles are lack of privacy and crowdedness in many forms (including noise level and air quality), while reliability of travel time does not appear in the top-10.

Research Opportunities MoTiV's data collection process was the largest initiative to apply an innovative smartphone-based approach to collect individuals' mobility behaviour and travel patterns, with a focus on VTT in Europe. Its findings will have important implications for urban and transport planners, policymakers, and authorities to implement more user-centric designs in urban mobility plans and to prepare inclusive transport policies, which tailored to individuals' needs and travel preferences. Crucially, in addition to more analysis of the project results, an open-source version of the European-wide dataset created by MoTiV will be made publicly available as a part of the MoTiV Data Management Plan (DMP)(described in Deliverable 4.1), for wider use by the scientific community, policymakers and other relevant stakeholders to stimulate further research on VTT and will also serve as a reference for analysis and assessment of the measures connected to SUMP and other EU key policy indicators on citizens' quality of life. The dataset will be accessible via the project website and relevant Open Data repositories in July 2020.

The MoTiV dataset offers a wealth of potential research activities that largely contribute to a European transport system, where the end-user takes center stage. To this respect, it is relevant to identify and to distinguish at least three different target groups that could benefit from the dataset and that are in a good position to take up new opportunities for innovation.

1. The scientific community
2. Public transport operators
3. Transport authorities and decision-makers

Mobile applications can provide superior travel data both in resolution and up-to-datedness. The dataset from MoTiV is a fine-grained portrait of multi-modal end-user mobility experiences for a continuous period of time, and thereby help to shape future mobility policies. The comparison with data sorted out from conventional travel surveys is sticking, as conventional mobility planning tools

heavily rely on data coming from household travel surveys and travel diaries, which are quite expensive, and thus are being updated with a low frequency.

The ambition of the project is, therefore, to contribute with sound analytical elements to the above-mentioned general stakeholder, which entails a great potential to exploit the knowledge generated with the MoTiV project. Some concrete areas of exploitation are further described.

Scientific Community

This group might be considered as a first level beneficiary of the MoTiV dataset, as it is important to stress that the specific objectives of the project entailed the need to expand the capabilities of already existing datasets by adding-on new insights and data concepts. To this respect, the dataset is of paramount importance to revise and challenge "conventional" cost-benefit narratives and paradigms, **stimulating further scientific research on VTT** or related areas and to support informed policy decisions.

The database offers an opportunity for the scientific community to challenge the assumptions that underpin such mindsets and to gradually nurture a more comprehensive and overarching assessment of the value of travel time, especially when addressing significant infrastructure investments in the transport realm.

Public Transport Operators

From a public transport operator point of view, the dataset informs: i) about the combination of the public transport services with other modes; ii) allows to derive O/D matrixes, which is particularly important in open access transport systems (most notably the ones that don't have gateways for entering and leaving the vehicles, like the metro usually has); and allows to understand the value proposition of the transport mode of each particular citizen, which helps to define communication and marketing campaigns and to improve the overall service. Related to these matters, one can consider that benchmarking the worthiness elements, which people with the same trip motivations hold, can be particularly valuable to the strategical planning of a public transport company and can offer them a **competitive edge**.

Transport Authorities

Urban planning agencies and governance actors are often unable to effectively take stock of the opportunities offered by new mobility technologies, because they are not equipped to address the mobility experience as a relevant and operable aspect of urban transport practices. The data production and planning models used by governance actors to qualify their decisions and to measure the effects of their interventions rarely account sufficiently for the holistic and self-perceived end-user experience. This concrete knowledge gap entails that governance actors are in control of very few operable strategies by which to **strategically modulate urban policies in the frame of Sustainable Urban Mobility Plans** and other urban planning policies.

The MoTiV dataset and any other data gathered under similar conditions and methodologies expand the knowledge about end-users and their mobility patterns. By doing this, it offers governance increased smart and data-driven capabilities to strategically use disruptive technologies to shape personal mobility systems of European cities.

What is Left to Do

It would be important to marketeer the MoTiV database and any other data-collection practice that is built on the MoTiV experience and one opportunity to do so effectively could require the development of a dashboard of visually attractive KPI's tailored to each specific stakeholder. This important development would allow one to quickly grasp the most important mobility trends and control any meaningful changes in the mobility ecosystem. This enhancement of visualization capacities is an exploitation element worth exploring in further EU projects and/or by entrepreneurs wishing to step in the big data ecosystem applied to mobility and to **build advanced visualization toolkits**.

Future Applications

Other prominent opportunities, and how they will impact both users and transport operators provided:

- **User profiling and clustering:** Crossing user mobility with their experience factors can be directly used to profile the users and cluster them (Basile et al., 2020). The identification of users with similar behaviour and a similar value of travel time might directly impact the shaping of journey planners, with solutions that target that cluster being presented first - e.g., a user belongs to a cluster associated with low emissions, hence sorting by emissions might be the default option. This has clear benefit for both the platform, which can tailor itself around the users and create trust, and for the users themselves, who would have a more personalized interface.
- **Recommender systems:** Current recommender systems associate mobility with Point-of-Interest recommendation (Liu et al., 2017). While in this domain classic notions of collaborative filtering can be enriched with geographical information, our dataset can offer much richer notions of peer users. With the MoTiV dataset, a peer user is not only someone who visited PoI similarly to another but can be someone who gives values to the same experience factors. In addition to this, the MoTiV dataset can enable novel forms of recommendation, based on the previous observations, such as a recommendation of the activities to perform a given type of a trip – e.g., reading in train trip.
- **Ad targeting:** most platforms nowadays survive thanks to advertisements, which are usually personalized being based on the experience of the users in a platform (Saia qt al., 2016). In the context of value of travel time, ad targeting should go beyond this, to consider the factors that positively or negatively impact user mobility, and tailor ads around them. These ads would be much more effective, thus benefiting the platform, and by presenting the users with ``complementary'' services, based on their preferences.

Conclusions and Future Work

Extracting actionable knowledge from user behaviour in traveling can help learn them more and provide them with better services (e.g., personalized rankings). For this reason, it is important not only to monitor the users, but to *understand* them and the choices they make, according to the *value of their travel time*. Plus, trips are complex entities, made up of several legs, which might disclose information about user behaviour at a finer granularity.

To accomplish the goal of getting to know users, their travel preferences, and their value of travel time, in this deliverable we presented the *MoTiV (Mobility and Time Value)* dataset. The dataset collects information about user mobility, capturing both raw information about trips and their legs (e.g., coordinates, time, and weather), plus information about the *worthwhileness* of a trip leg.

To assess the impact that our dataset can have in the real-world, we presented also a use-case to analyse what are experience factors that can negatively impact the use of public transport, by comparing cycling and walking versus private cars. As our use case has shown, the MoTiV dataset can provide valuable information for transport operators and service providers, so that their services can be tailored on the needs of their users. This deliverable also presents opportunities in other domains, such as personalized recommendation or customer clustering.

Given the valuable information contained in our dataset and the multiple opportunities for its exploitation, it is our goal to keep the MoTiV consortium and community engaged, to make the dataset grow even more. In addition to this, we plan to use the dataset to take action in concrete real-world scenarios, such as the design of cost-benefit analyses; to see how infrastructures and services can be improved, thanks to the needs and preferences of the users.

References

Basile, S., Consonni, C., Manca, M., & Boratto, L. (2020, July). Matching User Preferences and Behavior for Mobility. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media* (pp. 141-150).

Boratto, L., Carta, S., Kaltenbrunner, A., & Manca, M. (2018). Guest editorial: Behavioral-data mining in information systems and the big data era. *Information Systems Frontiers*, *20*(6), 1153-1156.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, *26*, 301-313.

Devarasetty, P. C., Burris, M., & Shaw, W. D. (2012). The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation research part A: policy and practice*, *46*(8), 1227-1240. https://doi.org/10.1016/j.tra.2012.05.002

Esztergár-Kiss, D. (2016). Optimization of multimodal travel chains. http://hdl.handle.net/10890/1534

Silvia Gabrielli, Paula Forbes, Antti Jylhä, Simon Wells, Miika Sirén, Samuli Hemminki, Petteri Nurmi, Rosa Maimone, Judith Masthoff, Giulio Jacucci Design challenges in motivating change for sustainable urban mobility, (2014). Design challenges in motivating change for sustainable urban mobility. *Computers in Human Behavior*, *41*, 416-423.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *nature*, *453*(7196), 779-782.

Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363-381.

Jakob, M., Hrncir, J., Oliva Felipe, L. J., Ronzano, F., & Finnegan, J. (2014). Personalized fully multimodal journey planner. In *ECAI 2014: 21st European Conference on Artificial Intelligence: 18-22 august 2014, Prague, Czech Republic: proceedings* (pp. 1225-1226). IOS Press.

Liu, Y., Pham, T. A. N., Cong, G., & Yuan, Q. (2017). An experimental evaluation of point-of-interest recommendation in location-based social networks.

Manca, M., Boratto, L., & Carta, S. (2018). Behavioral data mining to produce novel and serendipitous friend recommendations in a social bookmarking system. *Information Systems Frontiers*, *20*(4), 825-839. https://doi.org/10.1007/s10796-015-9600-3

Manca, M., Boratto, L., Roman, V. M., i Gallissà, O. M., & Kaltenbrunner, A. (2017). Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Social Networks and Media*, *1*, 56-69.

Saia, R., Boratto, L., Carta, S., & Fenu, G. (2016). Binary sieves: Toward a semantic approach to user segmentation for behavioral targeting. *Future Generation Computer Systems*, *64*, 186-197. https://doi.org/10.1016/j.future.2016.04.006

Schrammel, J., Prost, S., Mattheiss, E., Bothos, E., & Tscheligi, M. (2015, June). Using individual and collaborative challenges in behavior change support systems: findings from a two-month field trial of a trip planner application. In *International Conference on Persuasive Technology* (pp. 160-171). Springer, Cham.

Sierpiński, G., & Staniek, M. (2017). Heuristic approach in a multimodal travel planner to support local authorities in urban traffic management. *Transportation Research Procedia*, *27*, 640-647.

Sourlas, D., & Nathanail, E. (2018, May). A Thorough Review and Analysis of Journey Planners. In *The 4th Conference on Sustainable Urban Mobility* (pp. 690-697). Springer, Cham.

Steadman, R. G. (1984). A universal scale of apparent temperature. *Journal of Climate and Applied Meteorology*, *23*(12), 1674-1687.

Vargas, A., Weffers, H., & da Rocha, H. V. (2011, October). Analyzing user interaction logs to evaluate the usability of web applications. In *2011 3rd Symposium on Web Society* (pp. 61-67). IEEE.

Zarwi, F. E., Vij, A., & Walker, J. (2017). Modeling and forecasting the evolution of preferences over time: A hidden Markov model of travel behavior. *arXiv preprint arXiv:1707.09133*. https://arxiv.org/pdf/1707.09133.pdf

Zografos, K. G., Androutsopoulos, K. N., & Apospori, E. (2012). User acceptance and willingness to pay for the use of multimodal trip planning systems. *Procedia-Social and Behavioral Sciences*, *48*, 2405-2414. https://doi.org/10.1016/j.sbspro.2012.06.1211