DATA MARKET AUSTRIA

www.datamarket.at

Mobility Pilot - Service Description

Deliverable number	*D8.2*
Dissemination level	*Public*
Delivery date	*30.09.2018*
Status	*Version 1.0*
Author(s)	*Bernhard Niedermayer (Catalysts), Lukas Winkler (Catalysts)*

The Data Market Austria Project has received funding from the programme "ICT of the Future" of the Austrian Research Promotion Agency (FFG) and the Austrian Ministry for Transport, Innovation and Technology (Project 855404)

Executive Summary

This document contains the description of the Data Market Austria (DMA) WP8 pilot in mobility domain: the taxi demand heatmap on the basis of mobility data (moving data of crowds), weather data, and additional open data (such as traffic situations, jams, or events).

Involved DMA partners in this pilot are: SWC, Catalysts, T-Mobile (+ Taxi 40100), ZAMG, Joanneum Research, and Siemens.

The objective of the pilot at hand is to design and realise an application that supports taxi fleet managers to better plan and operate taxi fleets in respect to positioning of taxis, as well as in the course of 'where should a taxi go after a successful finalised taxi ride'.

This document describes the implemented service, including its API definition, data structures, metadata, and service level definitions. We also give a very brief insight into the technology stack which was used.

Table of Contents

1. Introduction

1.1 Problem Definition - Value Proposition

Taxis are an integral part of the transport systems world-wide. Mainly in cities, taxis are widely-used and the competition between taxi fleet companies and/or individual drivers is high. The margin (profit) per ride is relatively small and thereby the level of utilization is one crucial factor for competitive advantage and business success in this industry.

Individual taxi drivers develop an experience on the optimal strategy after finishing a ride. Options are to stay and wait, to drive to the nearest taxi stand, or to drive to the nearest “hotspot”, etc. Such hotspots are often well-known in the form of implicit knowledge of taxi drivers, but are only rarely-used by the coordinator of taxi fleets to optimise placement of taxis and/or to guide a taxi driver to the next and nearby hotspot after a managed ride.

Thus, operators of a taxi fleet (taxi fleet companies) have a strong demand to optimise the positioning (and guidance) of taxis of the fleet around the region the company is operating in (e.g., City of Graz, City of Vienna).

This can be supported by (i) predicting the demand for taxi rides for the next 15 minutes in a given grid, (ii) thereby identifying relevant hotspots in a given grid (along different event types, people crowds, moving data of people (crowds), etc.), and (iii) by giving the fleet operator and taxi driver information about the best way to serve such demand (route, position, etc.) and in consequence reach the highest possible taxi fleet utilization.

The Taxi Demand Heatmap is a service offering predictions about future demand for taxi rides in the form of demand level per grid cell. The time range of the prediction is the next 15 minutes.

2 Description of the Service

2.1 API

Figure 1 and Figure 2 show the endpoints for data retrieval and the service’s health check.

Request	
Method	**URL**
GET	`/api/v1/data`
Parameters/Request Body	
Type	**Params**
QUERY	fb

QUERY	oebb
QUERY	viennaGv
QUERY	topLocation
QUERY	viennaTicket
QUERY	viennaHospital
QUERY	viennaBuilding
QUERY	weather
QUERY	districts
Return taxi demand heatmap including inputs according to the parameters given.	
Response	
Status	**Response**
200	**demand + geo reference**

Figure 1: API - Heatmap data

Request	
Method	**URL**
GET	/api/v1/alive
Parameters/Request Body	
Type	**Params**
Live probe of the service	
Response	
Status	**Response**
200	- **empty -**

Figure 2: API - Health

2.2 Data structure

The result is returned as a JSON String. The actual data structure is composed of the objects and primitives as depicted in Figure 3.

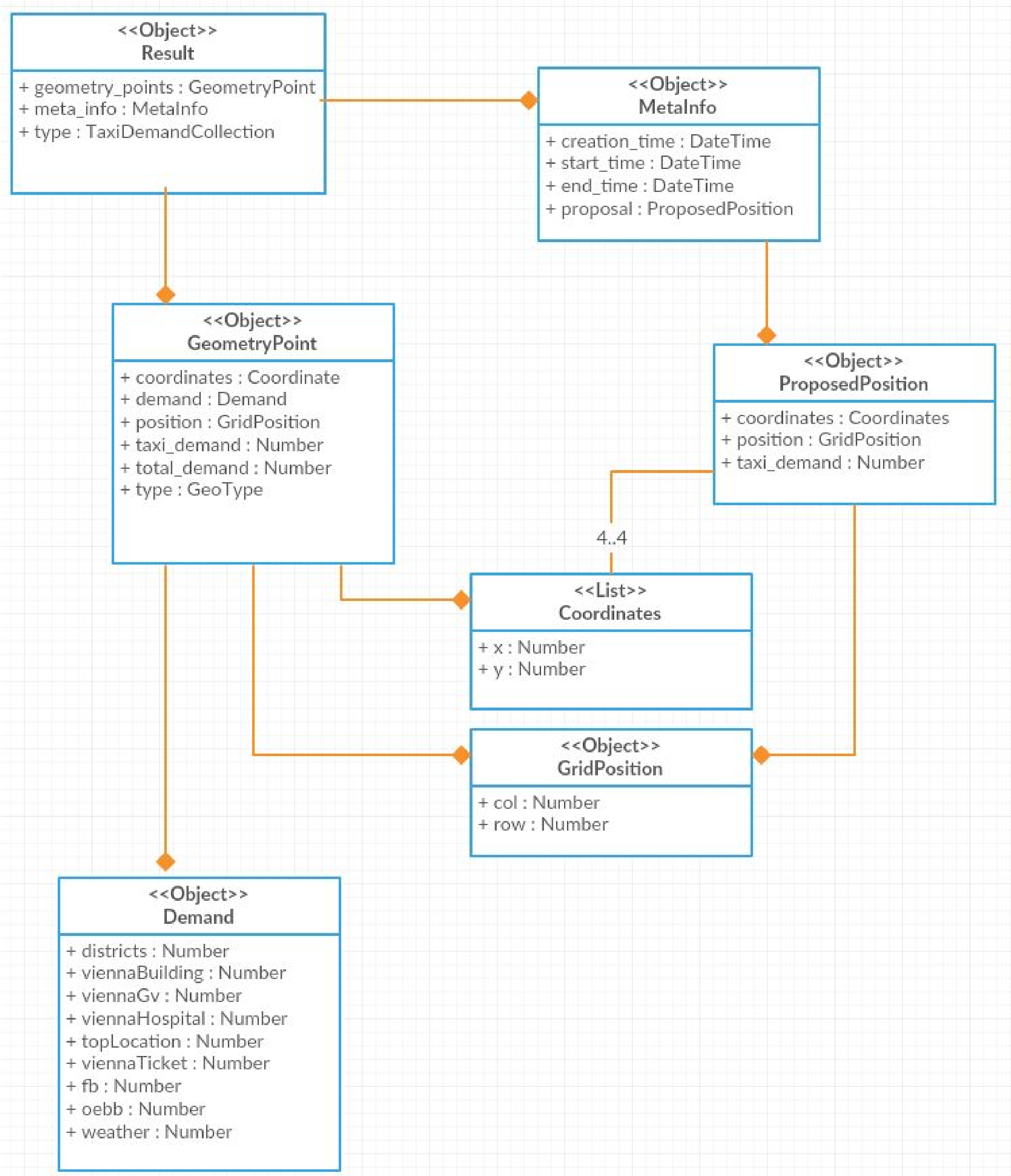

Figure 3: Heatmap data structure

2.3 Primary Content / Data Involved

- Past taxi rides in the Vienna area (Taxi 40100):
  - GPS track
  - distance
  - driving time
  - starting/end location
  - starting/end time
- Weather conditions (ZAMG):
  - temperature
  - rainfall
  - see parameter: weather
- Train delays (public):
  - minutes of delay
  - see parameter: oebb
- Events (public):
  - type of event
  - location
  - start/end time
  - expected number of visitors
  - see parameters: viennaTicket, viennaGv fb
- Static points of interest and base geo distribution of people
  - inhabitants per district, district area
  - building sizes
  - well known event locations
  - hospitals
  - see parameters: topLocation, viennaHospitals, districts, viennaBuilding
- Distribution of people over the area (T-Mobile):
  - number of persons per grid cell
- Traffic jams and traffic / road conditions as for instance construction places (public):
  - expected delay in minutes
- Time of day, day of week, season (public)

2.4 Meta Data

The service metadata description follows the DMA metadata schema as given in Table 1 - Table 3.

Data Catalogue

datasets	*(reference dataset description as of Table 2)*
main description	Prediction of the demand for taxi rides in the next 15 minutes across an urban area.
publisher	DMA operator[1]
title	Taxi Demand Heatmap

[1] The DMA operator will act as a publisher of non-commercial demonstrators. Another entity might later publish this service in a commercial setting.

catalog_unique_identifier	*(generated)*
language	@en
user_generated_tags	#mobility, #taxi, #taxi-demand
tags	#mobility
access_rights	:public
price_model	:PayPeriodically

Table 1: Meta Data - Catalogue

Dataset

title	Vienna Taxi Demand Heatmap
description	Prediction of the demand for taxi rides in the next 15 minutes across Vienna.
user_generated_tags	#mobility, #taxi, #taxi-demand
tags	#mobility
contact_point	<http://datamarket.at>
dataset_distribution	*(reference distribution below)*
theme	@Mobility
publisher	DMA operator[2]
version	v1.0
unique_identifier	*(generated)*
language	@en
access_rights	:public

Table 2: Meta Data - Dataset

Dataset Distribution

access_URL	<https://web-taxi-demand.cloud-apps.catalysts.cc/>
format	<http://www.iana.org/assignments/media-types/text/csv>
license	CC BY

[2] The DMA operator will act as a publisher of non-commercial demonstrators. Another entity might later publish this service in a commercial setting.

service_level_definition	*(reference to SLAs as of Table 4)*
service_level_agreement	*(reference to SLAs as of Table 4)*
price_model	:PayPeriodically
description	*(inherited)*
unique_identifier	*(generated)*

Table 3: Meta Data - Dataset Description

2.5 IT Prerequisites

In a prediction task, two phases can be distinguished: the first phase is the model training phase, which is expensive in terms of computational resources. The second phase is the production phase, where the model of the first phase is applied to generate predictions based upon new data.

For the model training phase, the Catalysts Machine Learning Cluster is used, providing access to storage in the petabyte dimension, as well as multiple high-end GPUs. In terms of tooling, the Python stack is used for exploratory data analytics.

The production phase requires less resources. The trained model is embedded into a web-service, which publishes the API as described in Figure 1. The service itself is dockerized and prepared for deployment on the OpenShift environment. For better management of deployables and to comply with DMA best practices a health check endpoint is provided as well (see Figure 2).

3 Legal Considerations / GDPR

With regard to data privacy and protection, we distinguish two phases: the first phase is the model training phase, where Data Providers allow access to certain datasets to members of the DMA consortium in their role as Provider of the Taxi Demand Heatmap Service. In a second phase, the DMA consortium members in their role as Service Providers allow DMA Customers access to the heatmap data.

Considering the model training, the DMA consortium members have signed bilateral agreements, satisfying the GDPR, as well as the current state of the art in data privacy and protection. Privacy sensitive data sources that are used in this phase are:

- T-Mobile phone position data
- Taxi 40100 data of taxi rides

Both of these data sources are only used for validating assumptions about influences on the demand for taxi rides and will not be used for later prediction.

In the prediction, the area under consideration (i.e., the Vienna urban area) is divided into a grid with a cell size of 100x100 meters. The data delivered to the DMA consumers is a number per cell, indicating the predicted demand within this cell area within the next 15 minutes. This information is not person-related and can thus be considered uncritical.

4 Non-Functional Requirements - SLAs

The Service Level Agreements follow the DMA default schema as given in Table 4 - Table 6.

General Properties

Identifier	*(generated)*
Name	Taxi Demand Heatmap SLDs
Description	The service level definitions for the taxi demand heatmap
Measurement period	day

Table 4: SLAs - General Properties

General Indicators

response time [h]	best effort
resolution time [h]	best effort
complaint count [#]	0
support hours [h/weekdays]	8/5

Table 5: SLAs - General Indicators

Service Indicators

availability [%]	best effort
incident count [#]	0
response time [ms]	best effort
latency [ms]	2000
data transfer rate [mbit/s]	1000
percentage of successful requests [%]	99
performance	*(to be discussed)*[3]

[3] At the time of writing the specification of SLA definitions is still ongoing:
There are numerous performance metrics which are currently evaluated wrt suitability in this context.

informability	*(to be discussed)*[4]

Table 6: SLAs - Service Indicators

4.1 Functionality

Security

- The API providing the taxi demand heatmap is protected by the token-based DMA security mechanism.

Extensibility

- The heatmap will be provided for the Vienna area.
- Prediction models for further cities can be trained given the necessary input data.
- Prediction models including additional data sources can be trained, but require an additional data analysis phase.

4.2 Usability

- The API is documented using the Swagger format and UI.
- The taxi demand heatmap service implement the DMA Service API, such that it can be operated and managed by the DMA platform.
- The API will follow the DMA style and naming conventions.
- During the development phase, there is a GUI visualizing the heatmap as OpenStreetMap overlay.

5 Technical Description

The taxi demand heatmap service is introduced as an independent component. Its purpose is to combine several data sources to predict the demand for taxi rides in the Vienna area. The computation of this demand will be based upon a model, which is trained in a preceding data analysis step.

[4] At the time of writing the specification of SLA definitions is still ongoing:
The Informability describes primitive information for managing a web service (e.g. protocol version number, encryption algorithm, messaging pattern, etc).

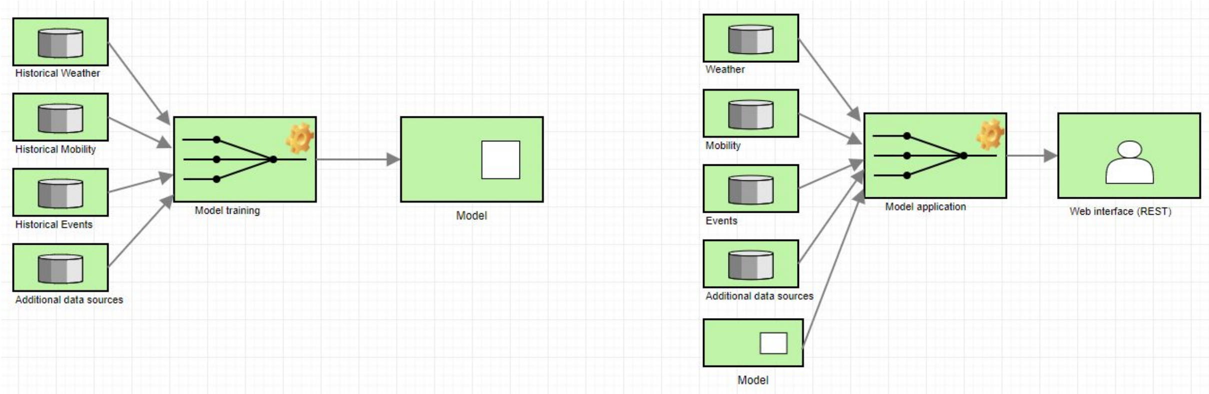

Figure 4: Model training and application (prediction)

The data analysis is done using the Python stack (including libraries such as numpy[5], matplotlib[6], scikit learn[7], pandas[8], etc.). Geo-referencing is done using gdal[9].

According to this, the prediction component is realized using the Python stack as well. For the implementation of the web service Flask[10] has been used.

Figure 4 shows a exemplary visualization of the heatmap as an OpenStreetMap overlay and the possibility to include and exclude different influencing factors into/from the prediction.

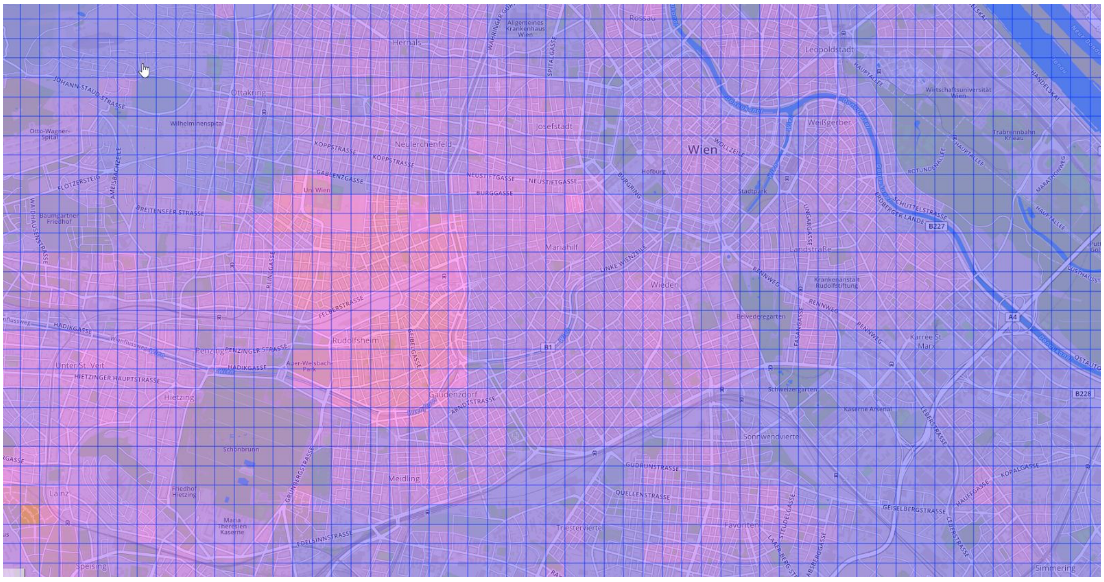

Figure 4 - Taxi demand heatmap of the Vienna area (2018-11-02 12:45)

Throughout the development and continuous improvement of the taxi demand heatmap we allow to show event markers on the overlay for better traceability of computed results. Figure 5 shows a detail of the same heatmap as in Figure 3 including these markers.

[5] http://www.numpy.org
[6] https://matplotlib.org
[7] http://scikit-learn.org
[8] https://pandas.pydata.org
[9] https://www.gdal.org
[10] http://flask.pocoo.org

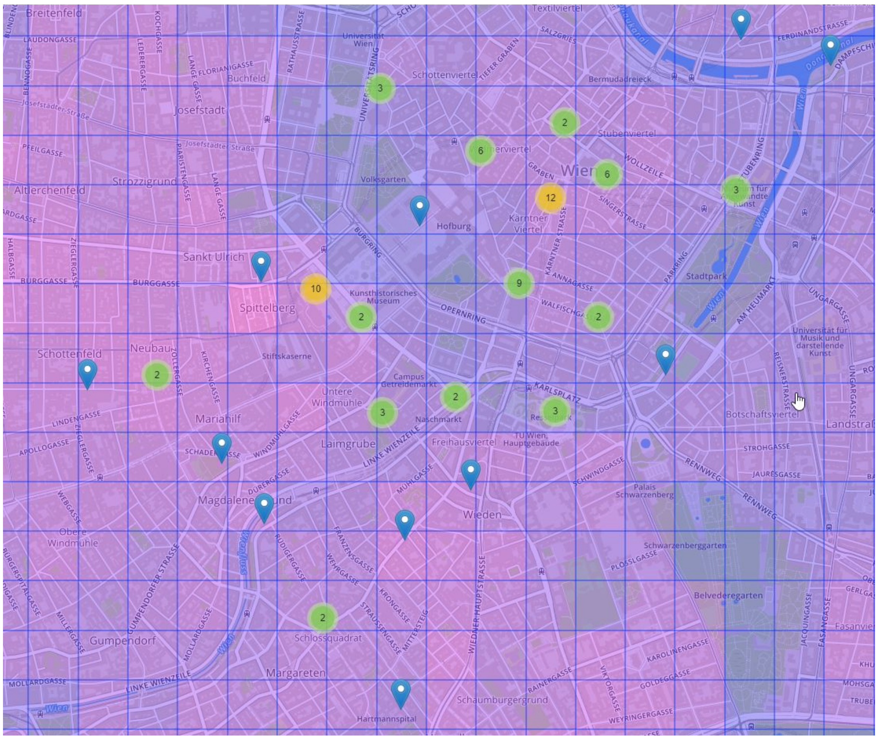

Figure 5 - Taxi demand heatmap with event markers

Figure 6 depicts an excerpt of the development stage GUI allowing to easily include or exclude individual input data sources.

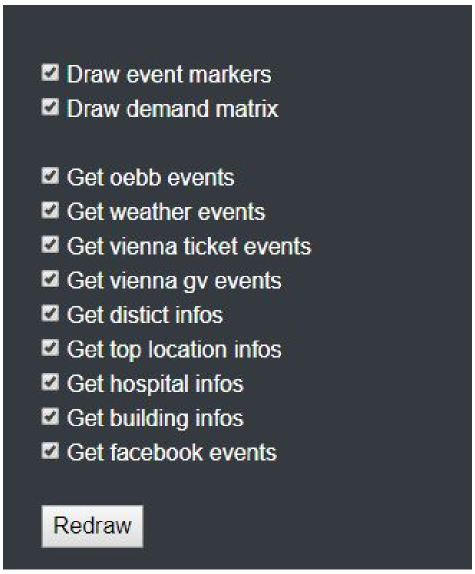

Figure 6 - Configuration menu of the taxi demand heatmap allowing to include different sets of input data