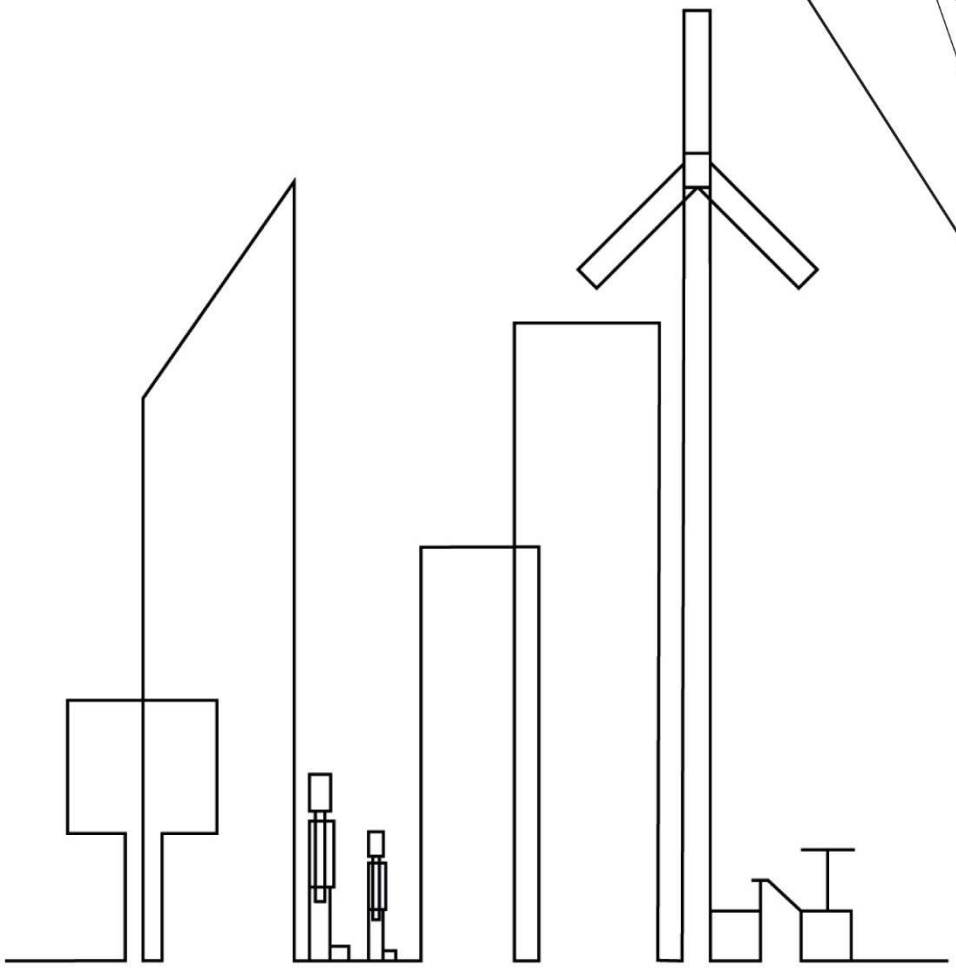

URBAN CO-CREATION DATA LAB

SERVICES AND USE CASES FULL DEFINITION FOR ALL THREE CITIES

December 2021

Co-financed by the Connecting Europe Facility of the European Union

Milestone Title	MS 4 – Service and use cases full definition for all 3 cities
Related Activity	Activity 1: Data preparation and open data infrastructure assessment
Related Task	Task 1.2: Services and Use Cases refinement
Author's	NOVA IMS
Due Submission Date	31.12.2020
Actual Submission	31.12.2021
Action Number	2018-EU-IA-0099
Agreement Number	INEA/CEF/ICT/A2018/1837945
Instrument	Connecting Europe Facility (CEF) – Telecommunications Sector
Starting Date of Action	01.10.2019
Completion Date of Action	28.02.2022
Main objective	Refining and analysing further the initial set of proposed services utilization scenarios (use cases) and identify those with highest priority that could serve as reference implementations in the co-creation labs and city validation

Table of Contents

1 Executive Summary 4

2 Objectives 4

2.1 Strategic Objective 4

2.2 Operational Objective 4

2.3 Tasks 4

2.4 Outputs: 5

2.5 Milestones and means of verification 5

3 Methodology 6

3.1 Use cases refinement 6

3.2 Data preparation, modelling and evaluation for use cases refinement 7

1 Executive Summary

This document is the mean of confirmation of **Milestone 4 – Service and use cases full definition for all 3 cities** of the Activity 1: Data Preparation and open data infrastructure assessment of the Action **Urban Co-creation Data Lab**, funded by the European Commission under the H2020ICT- 28-2017 Connecting Europe Facility (CEF) – Telecommunications Sector under the grant agreement n. INEA/CEF/ICT/A2018/1837945. In this report the services and use cases full definition are specified for the city of Lisbon along with the identification of the use cases that were implemented in the Provincial Council of Badajoz.

2 Objectives

2.1 Strategic Objective

The main objective of this activity was to prepare data and open data infrastructure for Smart Management Platform.

2.2 Operational Objective

Data preparation and open data infrastructure assessment.

2.3 Tasks

The development of this activity included the implementation of following tasks:

Task 1.1: Data Definition and requirements definition

This task intended to identify, characterise, evaluate, and increase the overall quality of the data that will be used in the first testing of the UCD Lab services in Lisbon. It involved a set of activities that will improve the overall quality and consistency, namely:

- Identification of data sources e.g., from operational systems, mobile technologies, devices, open data infrastructures and platforms with emphasis in Lisbon, but which can be replicated in the other two cities;
- Assessment of the quality of the data available in terms of accuracy, usefulness, reliability, latency, etc. and perform Data/Metadata Harmonization and Validation for preparing the UCD Lab services experiments;
- Identification of further stakeholders and end-users' needs and requirements to define the necessary features and functions of the platform particularly regarding the proposed services;
- Final definition and harmonization of the data sets, for the 5 services, with stakeholders and end-users' needs and requirements definition.

In the case of Lisbon, the set of selected open data from Lisboa Aberta and Lisboa Smart Management Platform (LSMP) was supposed to be made for the 5 proposed analytical services:

1. Assessment of human flows in highly crowded amusement areas;
2. Analytics to predict patterns in the production of solid urban waste;

3. Identify patterns and impact of illegal parking;
4. Predictive analytics for propagation of pollution in cities;
5. Predictive analytics for impact of events in mobility/transports.

Nevertheless, and after the global assessment of the data available quality and the identification of the stakeholders and end-users' needs and requirements, the proposed analytical services were reviewed as follows:

- MOBILITY - Evaluation and prediction of patterns and behaviours of micro mobility in the city of Lisbon, to support new planning and management approaches altogether with new tools to evaluate impact and prediction of micro mobility user behaviour;
- WASTE MANAGEMENT - Identification of patterns/profiles and solid waste production prediction in the city of Lisbon, to identify patterns to support the prediction of the production of urban waste associated with a variety of context information (e.g. events, climate situation, etc.);
- PARKING - Identification of patterns and prediction of irregular parking in the city of Lisbon to improve surveillance efficiency on irregular parking in Lisbon;
- POLLUTION - Elaboration of predictive models for the propagation of liquid and atmospheric pollutants in the city of Lisbon, to simulate their propagation after accidents with hazard substances in Lisbon;
- EMERGENCY – Identification of patterns and predictive modelling of traffic accidents in the city, to optimize the response of firefighters and civil protection services to emergency due to traffic accidents in Lisbon.

Due to the pandemic the selection of the two additional cities for testing and validation of local use cases during the Action, had to be delayed. Nevertheless, the data selection and definition will be performed for each of the two additional cities (indicatively, Barcelona and Amsterdam).

Task 1.2: Services and Use Cases refinement

Under this task, the services and use cases full definition was produced after:

- Refining and analysing further the initial set of proposed services utilization scenarios (use cases) and identify those with highest priority that could serve as reference implementations in the co-creation labs and city validation;
- Providing more detailed specifications for the city services validations, clearly identifying the similarities and differences between them.

2.4 Outputs:

Services and use cases full definition for Lisbon.

2.5 Milestones and means of verification

Milestone number	Milestone description	Indicative completion date	Completion date	Means of verification
4	Services and use cases full definition for all 3 cities	31/12/2020	31/12/2021	Services and use cases full definition for Lisbon

3 Methodology

The methodology developed in this phase of the Action was implemented in two stages:

I. Definition of the strategy for the use cases refinement;
II. Interaction with municipality services and data tests.

3.1 Use cases refinement

The iterative process between the research team of NOVA Information Management School, the Lisbon Urban Intelligence Management Centre (CGIUL), several municipality departments (Mobility Department, Urban Hygiene Department, Civil Protection Service, Sanitation Department and the Municipality Police) and external entities (EMEL – a parking and mobility company that supports Lisbon Municipality) along the data definition and requirements definition described in report Data sets and requirements defined for all 3 cities, allowed to refine the use cases with the highest priority to be developed in the Action along with their detailed specifications. These refined use cases are the base to create the services that will be deployed in the LSMP. The methodology followed for the development of the analytical services is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM). The phases of this methodology are presented in Figure 1.

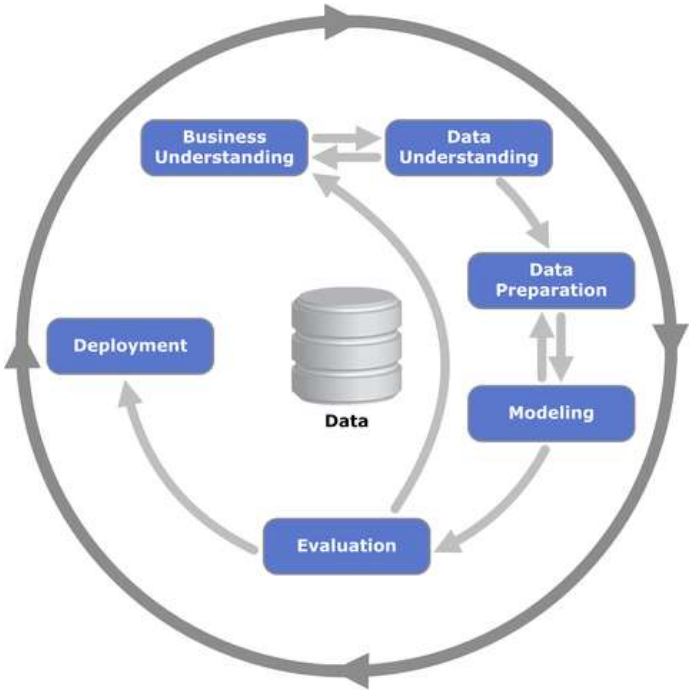

Figure 1. Flowing chart of the phases in the Cross-Industry Standard Process for Data Mining (CRISP-DM).

One of the advantages of this methodology is that the phases sequence is not strict and moving back and forth between different phases as it is always required. This aspect was of major importance for the refinement of the use cases, as they were defined based on the data available and their respective characteristics in the experiments made in the modelling phase, the evaluation results obtained and the interaction between the research team and the municipality services.

3.2 Data preparation, modelling, and evaluation for use cases refinement

In this phase of the Action is being made the necessary data pre-processing of the data to develop the use cases. For the development of the use cases the datasets will be subject to a series of operations and transformations that are necessary to the harmonization of the datasets to develop the use cases.

In the data preparation phase, several spatial and temporal aggregation procedures were developed to make experiments using several models, namely machine learning and time series models, that were evaluated using several metrics. In Table 1 are presented the spatial and time aggregation, the models and respective evaluation made that allowed the refinement of the use cases.

Use case	Spatial aggregation	Time aggregation	Models	Evaluation
#1 Micromobility	Hexagonal grid; bike station	9 months; 1 hour; 3 hours	OLS; GLM; SVM; KNN; RF	MAE; RMSE; MAPE
			ARIMA, ARIMAX, SARIMA, SARIMAX	Autocorrelation; Standardized residuals over time; AIC; MAE; RMSE; MAPE
#2 Waste Management	Waste collection circuit	Weekly	OLS; GLM; SVM; KNN; RF	MAE; RMSE; MAPE; MAE per capita
			ARIMA, ARIMAX, SARIMA, SARIMAX	Autocorrelation; Standardized residuals over time; AIC; MAE; RMSE; MAPE
#3 Parking	Street level	3 hours	Deep neural network, decision trees	AUC; MAPE
#4 Pollution	City area with 1 Km^2	1 to 15 minutes	Spatially filtered incompressible Navier-Stokes Equations; Vreman; Atmospheric Boundary layer with roughness	Shinjuku urban area ABL
#5 Emergency	Street level	3 hours	Deep neural network, decision trees	AUC; MAPE

Table 1. Spatial and time aggregation, models and respective evaluation tested for each use case. OLS stands for Ordinary Least Squares; GLM for Generalized Linear Model; SVM for Support Vector Machine; KNN for K Nearest Neighbours; RF for Random Forest; ARIMA for Auto Regressive Integrated Moving Average; SARIMA for Seasonal Auto Regressive Integrated Moving Average; ARIMAX for Autoregressive Integrated Moving Average with Explanatory Variable; SARIMAX for Seasonal Auto Regressive Integrated Moving Average with Exogenous Factors; MAE for mean absolute error; RMSE for root mean square error; MAPE for mean percentage absolute error; WMAPE for weighted mean absolute percentage error; AIC for Akaike information criterion; and AUC for Area Under Curve.

After the several interactions with the municipality services of the Lisbon City Council and the initial experiments made with the data available the consortium defined the use cases that are presented in Table 2.

Main sector	Designation	Problem to be addressed	Expected outcome
#1 Micromobility	Evaluation and prediction of patterns and behaviours of micro mobility	Micro mobility poses great challenges in the city environment, as nowadays micro mobility is changing how citizens commute in cities. In this sense there is the need to understand and anticipate which are the spatial and temporal patterns of micro mobility commute in the city, along with parking, storage and operations of micro mobility vehicles.	Predictive model of bike docks occupation for a bike sharing service, considering weather and business days and weekends. The model will allow to predict bike docks occupation in periods of 3 hours during a week, providing relevant information for bike rebalancing operations. The outputs provided by the model will also be relevant for the planning of the bike stations network expansion.
#2 Waste management	Identification of patterns/profiles and solid waste production prediction	Solid waste production and collection is nowadays a huge challenge for the municipalities. Indeed, waste collection costs range between 40 to 60% of waste management and is responsible for the production of 4,2 to 12 kg of CO_2 per tonne of waste. Predicting and understanding the relations between the socio-demographic characteristics and the waste production, will lead to an improvement in the operations efficiency of waste collection and transportation by the municipalities.	Predictive model of mixed waste production considering weeks with holidays. The predictive model will allow to predict mixed waste production for a set of mixed waste collection circuits weekly during a month. The model will allow to optimize the number of trucks needed to collect waste in each circuit, as it will be able to provide information about the expected mixed waste production for a specific mixed waste collection circuit
#3 Parking	Identification of patterns, explanatory factors and prediction of abusive parking	As population that lives, works and visits cities are increasing, parking capability is under pressure, namely due to unattractive or insufficient public transportation, inadequate drivers' education and insufficient regulation. Predicting abusive parking can aid the municipality services to optimize parking inspection and dissuade possible drivers' irregular behaviour.	Identification of patterns and risk of illegal parking in the city of Lisbon, at street level and time of day. The model will be based on the illegal parking occurrences registered by the Municipality Police of Lisbon. The proximity to services will also be included in the model, namely the proximity to schools, and health facilities. This service will be used to optimize the dispatch of police officers for areas with more risk of parking illegalities.
#4 Pollution	Elaboration of propagation models for the prediction of atmospheric and liquid pollutants behaviour	As there is an increase in people living in cities, is growing an increase concern regarding atmospheric and liquid pollution. Indeed, there is lack of information about propagation of liquid and atmospheric pollutants so civil	Models of atmospheric (natural gas) and liquid pollutants (nitrobenzene) propagation at city micro-scale, with a 1 to 15 minutes temporal resolution will be developed. The models will be developed using buildings 3D geometry, weather data, terrain and the physical characteristics of the

Main sector	Designation	Problem to be addressed	Expected outcome
		protection and sanitation services could understand pollutants propagation and optimize their services in case of an environmental accident. There is the need to model atmospheric and liquid pollutants propagation in the city, to assess pollution impacts in the city environment	pollutants. This service will allow to the civil protection services, isolate the areas that are more vulnerable to the pollution clouds, allowing to manage the city areas that must be isolated, for protection of people and goods.
#5 Emergency	Identification of patterns and predictive modelling of traffic accidents	The recent and future increase in cities population will input a big pressure in cities infrastructures, namely in roads, increasing the probability of the occurrence of traffic accidents, carrying significant challenges in cities mobility, transportation systems and the more important in human safety. In this sense is of extreme importance understand the infrastructural and environmental characteristics when traffic accidents occur along predicting them, to allow for city emergency services an optimized response to emergency, and for city managers plan road traffic, considering the risk of traffic accidents.	Identification of traffic accidents risk at street level. This prediction model will be deployed in PGIL, where it will provide, the streets and periods of the day in which there's a higher risk of traffic accidents. This service will be used by the emergency services to optimize the location of emergency vehicles in the proximity of streets with more risk of traffic accidents. This will allow the reduction of the response time to traffic accidents.

Table 2. Use cases refinement and expected outcomes.

All these use cases were presented to the Provincial Council of Badajoz, that showed interest in the development of use case #2 Waste management considering the same objectives and expected outcomes as the use case developed for Lisbon. To notice that the Provincial Council of Badajoz don't have information about waste collection in the city of Badajoz but from the several villages of the region of Badajoz that is composed by 164 municipalities. In the Report on completed use cases implementation in each city the differences in implementation are presented and discussed.