

URBAN CO-CREATION DATA LAB

DATA SETS AND REQUIREMENTS FOR THREE CITIES

May 2020

MS 2 - Data sets and requirements defined for all 3 cities

Milestone Title	MS 2 - Data sets and requirements defined for all 3 cities
Related Activity	Activity 1: Data preparation and open data infrastructure assessment
Related Task	<ul style="list-style-type: none"> ▪ Data Definition and requirements definition; ▪ Services and Use Cases refinement
Author's	NOVA IMS
Due Submission Date	31.12.2019
Actual Submission	31.05.2020
Action Number	2018-EU-LA-0099
Agreement Number	INEA/CEF/ICT/A2018/1837945
Instrument	Connecting Europe Facility (CEF) - Telecommunications Sector
Starting Date of Action	01.10.2019
Completion Date	30.09.2021
Main objective	Prepare data and open data infrastructure for Smart Management Platform

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	6
3	Methodology	6
3.1	Understanding the municipality business needs	6
3.2	Identification and understanding of the datasets and their characteristics	6
3.3	Use cases refinement	7
3.4	Datasets pre-processing and feature engineering	7
4	Annexes	8

1 Executive Summary

This document is the **Milestone 2 - Data sets and requirements defined 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.

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 Lisboa. 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 Lisboa, 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 Lisboa 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/transport.

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 behaviours;

- **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 parking in the city of Lisbon to improve surveillance efficiency

To create new models either to predict or to generate viable alternatives for illegal parking in the city;

- **POLLUTION** - Elaboration of predictive models for the propagation of liquid and atmospheric pollutants in the city of Lisbon

To develop predictive models for the propagation of liquid and atmospheric pollutants;

- **CROWD MANAGEMENT** - Evaluation and elaboration of models for predicting the impact of major events on Lisbon city services

To build impact prediction models based on the mobility/people flows in large events.

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:

Data sets definition and requirements definition are presented in Annex 1 (with the indication of use for each use case) and service and use cases full definition are presented in Annex 2.

2.5 Milestones and means of verification

<i>Milestone number</i>	Milestone description	Indicative completion date	Completion date	Means of verification
2	Data sets and requirements defined for all 3 cities	31/12/2019	31/05/2020	Data sets definition and requirements definition completed and available for Lisbon city

3 Methodology

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

- I. understating the municipality business needs for each dimension of the city addressed in this Action (i.e. micro mobility, waste management, parking, pollution and crowd management);
- II. identification and understanding of the datasets and their characteristics available in the Lisbon Smart Management Platform (LSMP);
- III. use cases refinement; and
- IV. datasets pre-processing and feature engineering for the development of the refined case studies.

3.1 Understanding the municipality business needs

In this first stage of the Action, several meetings were made between the research team that is developing the technical work, the Lisbon Urban Intelligence Management Centre (CGIUL) (responsible for the Lisbon municipality data management), and departments of Lisbon municipality, namely the mobility department, urban hygiene department and the civil protection department. The main goal of these meetings was to understand several problems that these departments face in operational terms regarding their daily activity that are not being addressed and understand what are their mainly activities in the municipality operation. The intention was to provide a first view of possible use cases that these departments could have interest in developing. To notice that this was an iterative process along this phase of the Action until the final definition and refinement of the use cases that will be developed in this Action (see Annex 2).

3.2 Identification and understanding of the datasets and their characteristics

In this phase of the Action a literature review was carried out by the research team to identify the relevant datasets that are used as explanatory variables to address use cases regarding each one of the city dimensions addressed in the Action (i.e. micro mobility, waste management, parking, pollution and crowd management). Several meetings and requests were made between the research team and the CGIUL to assess the available datasets in the LSMP that have a potential interest for the development of the case studies. The identification of the available datasets and their characteristics allowed the beginning of the use cases refinement. Several information about the identified datasets was collected to increase the understating of them, namely:

- Group;
- Identification;
- Description;

- Source;
- Format;
- Date of the first record;
- Date of the last record;
- Spatial representation;
- Attributes;
- Type

Samples of the datasets were provided in an initial stage testing.

3.3 Use cases refinement

The iterative process between the research team, CGIUL and the municipality departments along the phases described in section 3.1 and section 3.2, allowed to develop the use cases with the highest priority to be developed in the Action along with their detailed specifications. These refined use cases will be the base to create the services that will be deployed in the LSMP.

3.4 Datasets pre-processing and feature engineering

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 to increase their quality, and to transformations that are necessary to the harmonization of the datasets to develop the use cases. The missing values are being imputed using several imputation methods and data is being aggregated in time and space.

The aggregation procedures are being developed in three stages: 1) contextual static data that is relevant for the development of each use case, is being aggregated to a spatial grid and aggregated by time unit (accordingly with the objective of each use case). This data will then be aggregated in time and space to the target variable of interest for each use case.

In Figure 1 are presented the four initial stages of the methodology developed at this phase of the Action, along with following ones namely: 5) Modelling and Evaluation and 6) Analytical services.

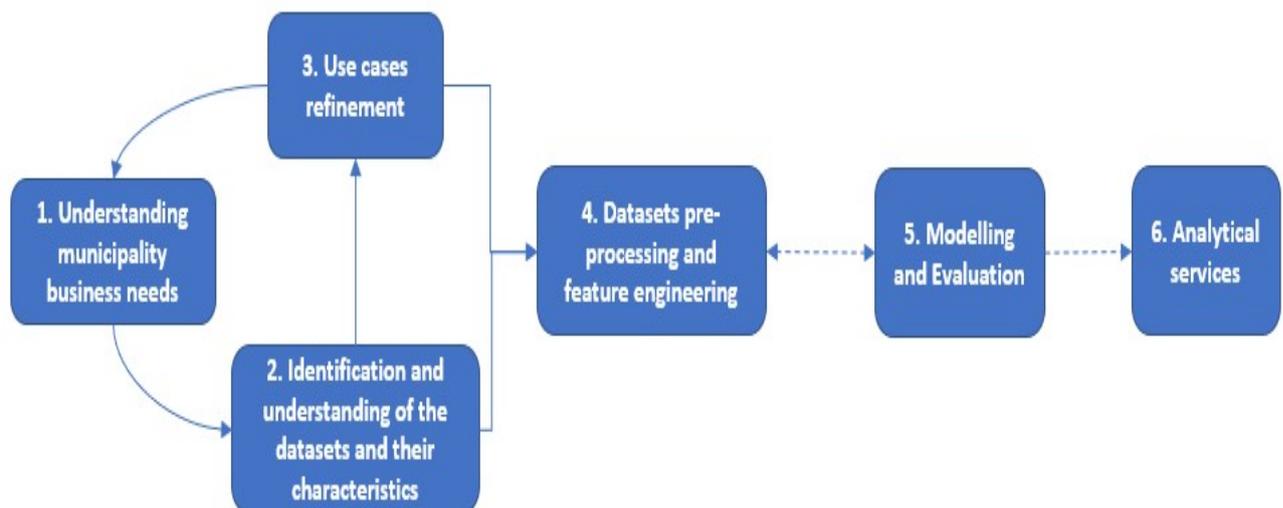


Figure 1. Flowchart of the methodology developed at this stage of the Action. Stage 5 and stage 6 will be developed next and the connection with previous stages is indicated by a dashed arrow.

4 Annexes

Annexe 1 - Lisbon Datasets Catalogue;

Annexe 2 - Refined use cases

Annexe 1 - Lisbon Datasets Catalogue

GROUP	DATASET	DESCRIPTION	#1 - Micromobility	#2 - Waste management	#3 - Parking	#4 - Pollution	#5 - Crowd management
Culture	Cultural associations	General information and location of cultural associations in the city of Lisbon	X	X	X		X
Culture	Auditoriums/amphitheatres	General information and location of auditoriums and amphitheatres in the city of Lisbon	X	X	X		X
Culture	Cultural centre's	General information and location of cultural canterers in the city of Lisbon	X	X	X		X
Culture	Cinemas	General information and location of cinemas in the city of Lisbon	X	X	X		X
Culture	Temporary occupation of public space	This dataset provides information about the events that took place in public space		X	X		X
Education	Public schools - 1st cycle	General information and location of 1st cycle schools (public network) in the city of Lisbon	X	X	X		X
Education	Private schools - 1st cycle	General information and location of private 1st cycle schools in Lisbon	X	X	X		X
Education	Public schools - 2nd cycle	Generic information and location of schools 2nd cycle (public network) in the city of Lisbon	X	X	X		X
Education	Private schools - 2nd and 3rd cycle	General information and location of 2nd and 3rd cycle private schools in the city of Lisbon	X	X	X		X
Education	Public schools - 3rd cycle	General information and location of 3rd cycle schools (public network) in the city of Lisbon	X	X	X		X
Education	Public schools - pre-scholar	General information and location of pre-schools (public schools) in the city of Lisbon	X	X	X		X
Education	Private schools - pre-scholar	General information and location of private kindergartens in Lisbon	X	X	X		X
Education	Public schools - professional education	General information and location of professional schools in the city of Lisbon	X	X	X		X
Education	Public schools - Secondary	Generic information and location of public secondary schools in the city of Lisbon	X	X	X		X
Education	Private schools - Secondary	General information and location of private secondary schools in the city of Lisbon	X	X	X		X
Education	Higher education institutions	General information on "Higher Education Institutions" strategic actors located in the city and region according to clusters, selected by the Direção Municipal de Economia e Inovação.	X	X	X		X
Environment	Weather data	Station readings for the following variables: • average air temperature (° C) • average relative humidity (%) • mean wind direction (°) • average wind intensity (m/s) • total amount of precipitation (mm) • total global solar radiation (Kj/m2)	X	X	X	X	X
Environment	Weather stations location	Location of the three weather stations in Lisbon	X	X	X	X	X

MS 2 - Data sets and requirements defined for all 3 cities

GROUP	DATASET	DESCRIPTION	#1 - Micromobility	#2 - Waste management	#3 - Parking	#4 - Pollution	#5 - Crowd management
Environment	COS2018	Land use/land cover map for 2018	X	X	X		X
Health	Health centres	General information and location of health canterers in the city of Lisbon	X	X	X		X
Health	Private hospitals	General information and location of private hospitals in the city of Lisbon	X	X	X		X
Health	Public hospitals	General information and location of public hospitals in the city of Lisbon	X	X	X		X
Infrastructures	City buildings 3D model	This dataset provides information about buildings occupation area and their height				X	
Infrastructures	Road network longitudinal slope	Road network slope computed through the road's longitudinal inclination	X				
Infrastructures	Bike lanes	Bike lanes network	X				
Infrastructures	Open Street Maps points of interest	POIs collected by volunteers in Open Street Maps project	X	X	X		X
Mobile operators	Mobile phone data	Count of people connected to the Altice mobile network					X
Mobility	Micro mobility parking	This dataset provides information about the location and status of micromobility vehicles	X				
Mobility	Micro mobility trips	This data set provides information about the micro mobility vehicles trips	X				
Mobility	GIRA trips	This data set provides information about the micro mobility vehicles trips regarding GIRA service	X				
Mobility	GIRA stations	Information about the number of available docks and ratio in a specific station in a given period	X				
Mobility	GIRA stations location	Location of GIRA stations	X				
Mobility	Waze Jams	This dataset provides information about traffic jams	X		X		X
Mobility	Traffic constrains	This dataset provides information of programmed traffic constrains (e.g. work on streets, events)					X
Mobility	Train stations	General information and location of train stations in the city of Lisbon	X	X	X		X
Mobility	Metro stations	General information and location of metro stations in Lisbon	X	X	X		X
Mobility	River's piers	Generic information and location of river interfaces in the city of Lisbon	X	X	X		X
Mobility	Bus ticketing						X
Mobility	Bus stops	This data set provides information about the location of the bus stops in Lisbon					X
Mobility	Metro ticketing						X
Mobility	Car tows by the municipality police	Irregularities reported by the municipality police regarding abusive parking			X		
Mobility	Parking spaces	Information about the parking spaces available			X		
Mobility	Car parks	Information about the car parks available			X		
Mobility	Parking meters	Information about the parking meters in operation			X		

MS 2 - Data sets and requirements defined for all 3 cities

GROUP	DATASET	DESCRIPTION	#1 - Micromobility	#2 - Waste management	#3 - Parking	#4 - Pollution	#5 - Crowd management
Mobility	Parking fares	Information about the parking fares areas			X		
Mobility	Parking for bikes	Location of parking areas for bikes for short, medium and long term	X				
Mobility	Parking for bikes and scooters (hotspot format)	Location of parking areas for bikes and scooters in hotspot format	X				
Other services	Na Minha Rua	Intervention requests (internal and external) recorded in the platform "Na Minha Rua"			X		
Social economic	Census 2011 GEOMETRY	This dataset provides information about the geometry of Lisbon census blocks		X	X		
Social economic	Census 2011	This dataset provides information about the sociodemographic profile of each Lisbon census block in 2011	X	X	X		
Social economic	Preference rights	Information about the preference rights owned by the municipality when a house is sold		X			
Tourism	Local accommodation	Information about the registered local accommodation		X	X		
Tourism	Tourist accommodation	Location of the registered tourist accommodation		X	X		
Urban hygiene	Filling containers	This data set provides information about the filling of waste containers, along with information about temperature, time of last collection, type of waste		X			
Urban hygiene	Garbage collection circuits	This dataset provides information about the location of garbage collection circuits		X			
Urban hygiene	Garbage collection freights	This dataset provides information about the total amount of garbage collected for each collecting circuit in a specific day		X			

 Still to be worked on

Annexe 2 - Refined use cases

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 micro mobility commute by spatial unit according with the weather predicted in the day before, proximity to schools, public services, public transportation network. The model will provide a probability of starting and ending a trip in each spatial unit. Besides the predicted commute pattern, the model results will be also useful for micro mobility vehicles operation (e.g. reinforce the number of available vehicles in spatial units, that in a certain hour have a higher probability of being registered the start of a trip)
#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	Identify patterns to support the prediction of the production of solid waste. The predictive model of waste production will be based, on the socio-demographic profile of the spatial units, the presence of population and services (e.g. schools, local accommodation, restaurants). The predictive model will be deployed in PGIL, allowing the creation of a service to predict solid waste production a day before, to optimize waste collection
#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 prediction of illegal and abusive parking in the city of Lisbon, by spatial unit and time of day. The model will be based on the complaints about abusive parking registered in "Na Minha Rua" (similar to Fix My Street platform) and in the car tows registered by the Municipality Police of Lisbon. The proximity to services will also be included in the model, namely the proximity to schools, health services or the proximity to cultural events
#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 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	Models of atmospheric (natural gas) and liquid pollutants (nitrobenzene) propagation at city micro-scale, with a 15 minutes temporal, resolution will be developed. The models will be developed using buildings 3D geometry, weather data and the physical characteristics of the pollutants.
#5 - Crowd management	Identification of patterns and predictive modelling of the impact that the realization of marathons has on public transportation demand	The recent and future increase in cities population will input a big pressure in cities infrastructures, namely crowd and traffic management. More specifically, this pressure increases with the realization of special events namely marathons, carrying significant challenges in cities mobility and transportation systems. In this sense is of extreme importance to assess the impacts that crowds have in the transports system and how municipalities and transport companies can optimize their resources to face the increase of population during the realization of marathons	Prediction by hour of public transportation demand before, during and after the realization of marathons. This prediction model will be deployed in PGIL, where it will provide a day earlier, the expected public transportation demand. This service will be used by the municipality and public transportation companies to optimize their services provided to citizens